the rules of belief

doubt and belief

Moore's proof

What should I believe?

Today we begin a new topic. For the next few weeks, we will be investigating the question, What should I believe?

This is different than all of the questions we've asked so far. It is not a question about how the world is — like the questions about whether God exists, about whether we have free will, and about what we are. Instead, it is a practical question — a question about what we ought to do.

Moreover, I think that this is a kind of question about which most of us have lots of opinions. Let's look at some examples.

One kind of interesting case to think about is belief in conspiracy theories.

Some of these are silly and, perhaps, harmless. Examples might include the belief that moon landing was faked, or that the earth is flat. Believers in these kinds of conspiracy theories typically discount evidence which seems to count against their theories. (For example, the fact that ~400,000 people worked toward the Apollo moon landing, and none have admitted that it was faked, or photographs of the earth from space which seem to show it to be spherical.)

It seems pretty clear that people who believe conspiracy theories of this kind are making a mistake of a certain kind; they are believing something that they should not believe. Let's call these cases of bad belief.

And this is not just because the conspiracy theories are false. Intuitively, sometimes you can have very good reason to form a belief which turns out to be false (say, if you have misleading evidence). The mistake that conspiracy theorists are making is a different kind of mistake.

What are some other examples of people making mistakes of this kind — people believing things that they should not believe?

What are some other examples of people making mistakes of this kind — people believing things that they should not believe?

One very common kind of example comes from cases of wishful thinking. Example: me, every August, thinking about the upcoming Notre Dame football season.

Another kind of example: people who form beliefs about their future on the basis of the horoscopes published in the *Observer*.

These are all examples of people believing things they should not believe. What are some examples of the opposite phenomenon — people forming beliefs as they should form them?

These are all examples of people believing things they should not believe. What are some examples of the opposite phenomenon — people forming beliefs as they should form them?

The easiest examples are people who seem to weigh, and respond appropriately to, their evidence. Examples: Sherlock Holmes; responsible scientists; careful jurors.

Notice that none of these people are infallible; even responsible scientists make mistakes. But intuitively they are going about belief formation in the right way.

There are also plenty of simple and everyday examples of this kind of thing. Suppose that you see people walking around outside with umbrellas open over their heads, and form the belief that it is raining. Could your belief be false? Of course. But intuitively, given your experience of the world, you are forming the belief that you should form.

Let's call these cases of good belief.

These are all easy cases. But there are plenty of hard cases too — and, in fact, you might think, philosophy is a kind of machine for generating hard cases! Haven't we already seen lots of cases in which there are arguments on both sides of an issue, and where it is hard to tell which argument is better?

Here is a hard case of interest:

The Believer

I've always believed that there is a God. I never really thought about what my evidence is for this claim. But now I wonder whether I have good reason for my beliefs. Some of the arguments for God's existence sound good, but all face objections that I am not sure how to answer. Still, I continue to believe that God exists.

Is The Believer forming the beliefs he or she should form, or not? Is it a case of good belief, or a case of bad belief? What do you think?

Is The Believer forming the beliefs he or she should form, or not? Is it a case of good belief, or a case of bad belief? What do you think?

It is worth emphasizing that **all** of us are like The Believer on some issues. Most of us have moral beliefs, or political beliefs, which we hold strongly but which we might find it difficult to argue for in a persuasive way.

Here's a way in which we might try to answer our question about whether The Believer should believe as s/he does. When we think about examples of good belief, and bad belief, the following thought seems very plausible:

> It isn't just an inexplicable fact that horoscope beliefs are bad beliefs, and that responsible scientist beliefs are good beliefs. Instead, there are general principles which determine whether someone should, in a certain circumstance, form a certain belief, or not.

Let's call these general principles the rules of belief.

It isn't just an inexplicable fact that horoscope beliefs are bad beliefs, and that responsible scientist beliefs are good beliefs. Instead, there are general principles which determine whether someone should, in a certain circumstance, form a certain belief, or not.

Let's call these general principles the rules of belief.

These rules might come in two flavors. One kind of rule might be a rule which tells you that in certain circumstances you should form a belief. Let's call these **positive rules**. A second kind of rule might tell you that in certain circumstances you should not form a belief. Let's call these **negative rules**.

It seems very plausible that there must be rules of this kind which explain the difference between cases of good belief and cases of bad belief.

And it also seems plausible that, if we can figure out what these rules are, we'll be able to figure out whether The Believer should believe what s/he does.

Our first attempt to formulate a rule of beliefs comes from a text familiar from our discussion of the nature of the self: Descartes' *Meditations*.

> Some years ago I was struck by the large number of falsehoods that I had accepted as true in my childhood, and by the highly doubtful nature of the whole edifice that I had subsequently based on them. I realized that it was necessary, once in the course of my life, to demolish everything completely and start again right from the foundations if I wanted to establish anything at all in the sciences that was stable and likely to last.

Descartes was aware that, by accepting common sense and the opinions of others, he could come to believe falsehoods. He therefore undertook to try to find a method — a way of forming beliefs — which would provide a secure foundation for belief.

Descartes was aware that, by accepting common sense and the opinions of others, he could come to believe falsehoods. He therefore undertook to try to find a method — a way of forming beliefs — which would provide a secure foundation for belief.

> Reason now leads me to think that I should hold back my assent from opinions which are not completely certain and indubitable just as carefully as I do from those which are patently false. So, for the purpose of rejecting all my opinions, it will be enough if I find in each of them at least some reason for doubt.

He hit upon the method of doubt: for any belief which he could coherently doubt to be true, he would give up that belief. Only then could he be certain not to believe any falsehoods.

He then applies this method to one of the most fundamental kinds of belief we form: beliefs about our environment formed on the basis of sensory experience.

He then applies this method to one of the most fundamental kinds of belief we form: beliefs about our environment formed on the basis of sensory experience.

He raises the question: how can I tell whether a given sense experience of mine is accurate?

After noting that only 'madmen' doubt the reliability of their sense experiences, Descartes notices something about his own experiences:

> As I think about this more carefully, I see plainly that there are never any sure signs by means of which being awake can be distinguished from being asleep.

As you know, Descartes goes on to argue that, although he must doubt the reliability of his sense experiences, he cannot doubt that he exists. But rather than going on to think about that aspect of his views (as we did in our discussion of dualism), today I want to focus on Descartes' central point about his sensory experiences of the world: namely, that "there are never any sure signs by means of which being awake can be distinguished from being asleep."

The key point is the following claim:

The matching hallucination assumption
For any sense experience, I can imagine a situation which is indistinguishable from that sense experience but in which my environment is not as the experience says it is.

The matching hallucination assumption

For any sense experience, I can imagine a situation which is indistinguishable from that sense experience but in which my environment is not as the experience says it is.

We are all familiar with experiences which seem not to represent our environment accurately.

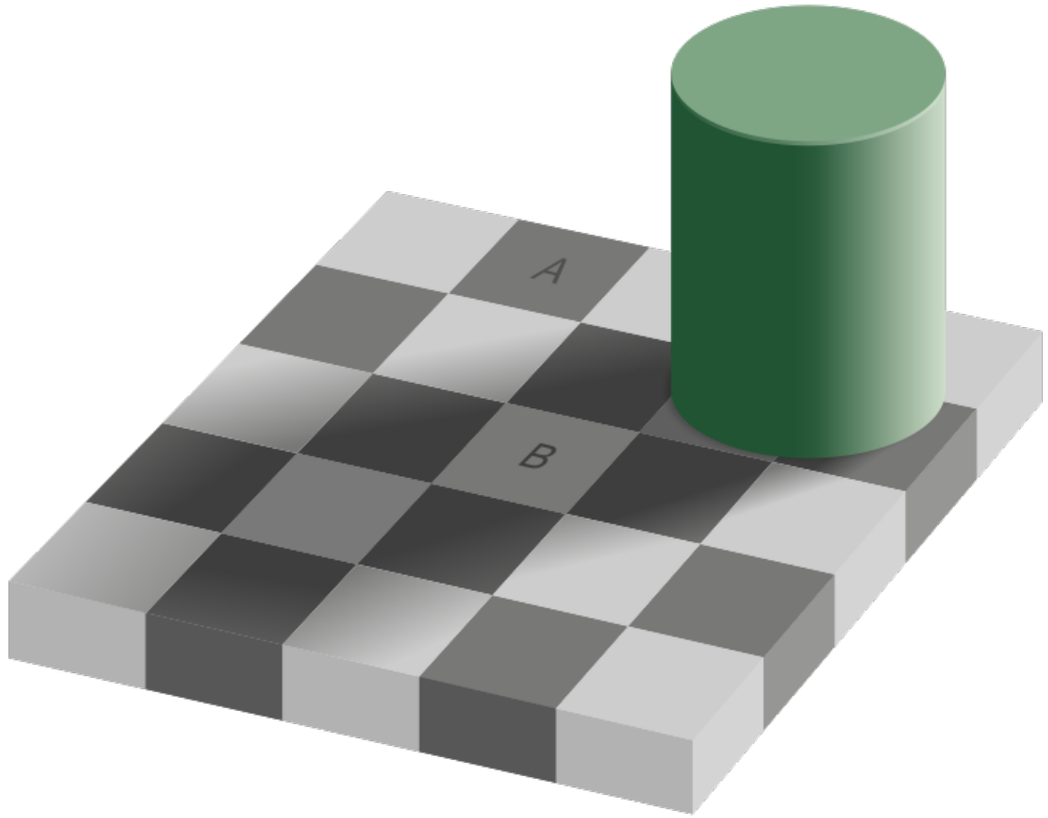
Some are every day experiences.

The matching hallucination assumption

For any sense experience, I can imagine a situation which is indistinguishable from that sense experience but in which my environment is not as the experience says it is.

We are all familiar with experiences which seem not to represent our environment accurately.

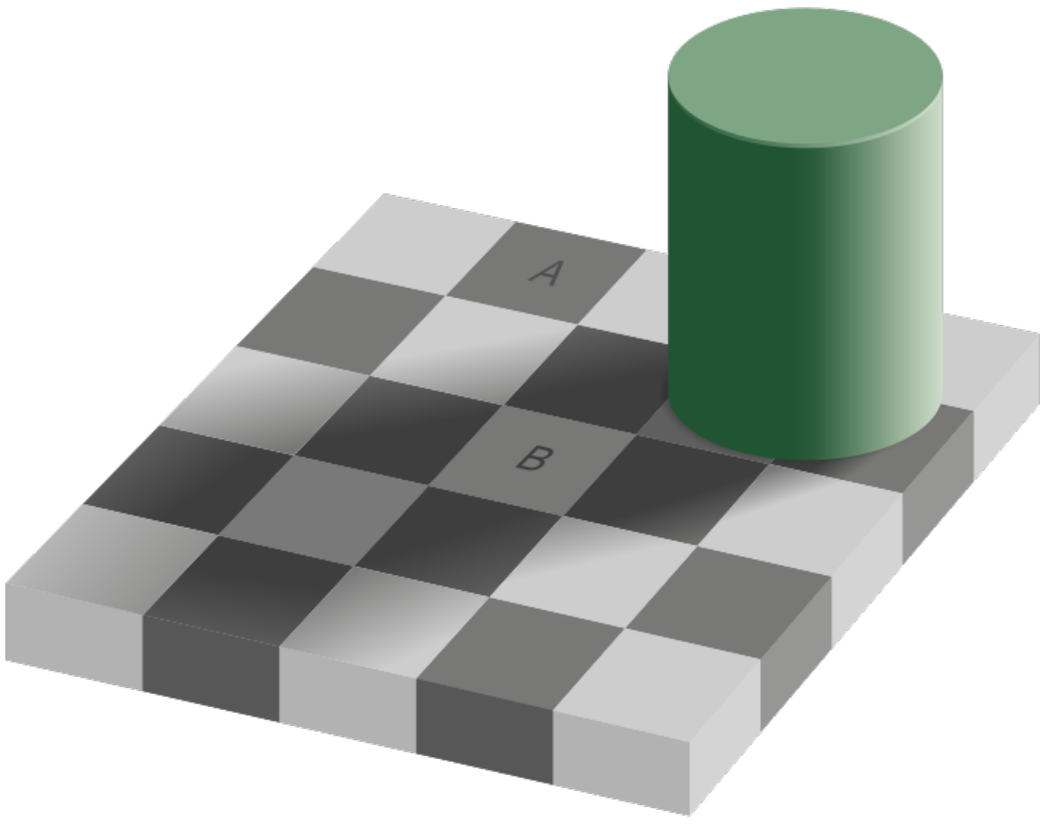
Others are intentionally constructed illusions which are used in vision science to study our mechanisms for representing the world around us.

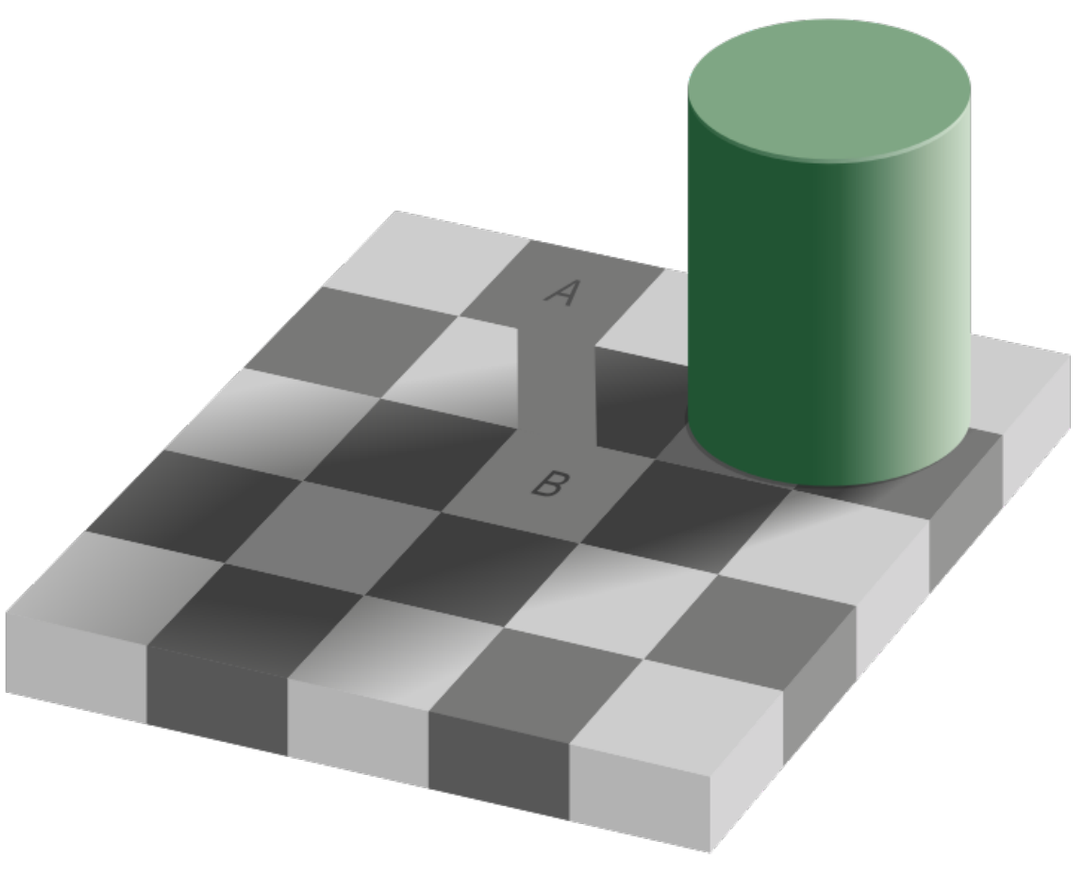


The matching hallucination assumption

For any sense experience, I can imagine a situation which is indistinguishable from that sense experience but in which my environment is not as the experience says it is.

We are all familiar with experiences which seem not to represent our environment accurately.





The matching hallucination assumption

For any sense experience, I can imagine a situation which is indistinguishable from that sense experience but in which my environment is not as the experience says it is.

But these are just specific examples. Is it really true that for any experience, we can imagine a matching — i.e., indiscriminable — illusion?

A number of different thought experiments suggest that we can.

One, which Descartes mentions, is the possibility that we are simply dreaming.

The matching hallucination assumption

For any sense experience, I can imagine a situation which is indistinguishable from that sense experience but in which my environment is not as the experience says it is.

But these are just specific examples. Is it really true that for any experience, we can imagine a matching — i.e., indiscriminable — illusion?

A number of different thought experiments suggest that we can.

Another, which Descartes also discusses, is the possibility that we are being deceived by an evil demon.

The matching hallucination assumption

For any sense experience, I can imagine a situation which is indistinguishable from that sense experience but in which my environment is not as the experience says it is.

But these are just specific examples. Is it really true that for any experience, we can imagine a matching — i.e., indiscriminable — illusion?

A number of different thought experiments suggest that we can.

We might also imagine that we are simply brains in vats which are being stimulated to cause illusory sense experiences as part of some nefarious scientific experiment.

The matching hallucination assumption

For any sense experience, I can imagine a situation which is indistinguishable from that sense experience but in which my environment is not as the experience says it is.

Once we notice this, though, this point can be used to generate a powerful argument for the conclusion that we cannot know anything around us on the basis of sense experience.

This is because the following principle seems very plausible:

If I cannot distinguish between two situations, then I cannot know which of them is real.

If I cannot distinguish between two situations, then I cannot know which of them is real.

Suppose that I tell you that, behind the lectern, I have an object. It is either a triangle or a circle.

Given that it is behind the lectern, the two different possibilities are indistinguishable to you.

Does it follow that you cannot know (from your present vantage point) whether there is a triangle or a circle behind the lectern?

The matching hallucination assumption

For any sense experience, I can imagine a situation which is indistinguishable from that sense experience but in which my environment is not as the experience says it is.

If I cannot distinguish between two situations, then I cannot know which of them is real.

I can never know whether any sense experience of mine is accurate.

Sense experience is my only way of knowing whether there is an external world.

I do not know whether there is an external world.

1. For any sense experience, I can imagine a situation which is indistinguishable from that sense experience but in which my environment is not as the experience says it is.
2. If I cannot distinguish between two situations, then I cannot know which of them is real.
3. I can never know whether any sense experience of mine is accurate. (1,2)
4. If I cannot know whether any of my sense experiences are accurate, I cannot know whether there is an external world.

———————————————-

C. I do not know whether there is an external world. (3,4)

Skepticism about some domain is the claim that one cannot have knowledge about that domain. This is an argument for skepticism about our knowledge of the external world.

We've already seen that there are strong reasons for accepting premises (1) and (2), and premise (4) seems quite plausible.

1. For any sense experience, I can imagine a situation which is indistinguishable from that sense experience but in which my environment is not as the experience says it is.
2. If I cannot distinguish between two situations, then I cannot know which of them is real.
3. I can never know whether any sense experience of mine is accurate. (1,2)
4. If I cannot know whether any of my sense experiences are accurate, I cannot know whether there is an external world.

———————————————-

C. I do not know whether there is an external world. (3,4)

You might be tempted to reply like this: 'OK, this shows that I can't know that there is an external world. But I should still believe that there is one.'

The problem is that a parallel argument seems to rule even this out.

1. For any sense experience, I can imagine a situation which is indistinguishable from that sense experience but in which my environment is not as the experience says it is.
2. If I cannot distinguish between two situations, then I cannot know which of them is real.
3. I can never know whether any sense experience of mine is accurate. (1,2)
4. If I cannot know whether any of my sense experiences are accurate, I cannot know whether there is an external world.

———————————————-

C. I do not know whether there is an external world. (3,4)

1. For any sense experience, I can imagine a situation which is indistinguishable from that sense experience but in which my environment is not as the experience says it is.
2*. If I cannot distinguish between two situations, then I should not believe that one but not the other is real.
3*. I should never believe that any sense experience of mine is accurate. (1,2*)
4*. If I should never believe that any sense experience of mine is accurate, I should never form beliefs about the external world.

———————————————-

C*. I should never form beliefs about the external world. (3*,4*)

Is the second argument as strong as the first?

It is tempting to think that we should be able to respond to Descartes by finding some way to show that certain experiences are not illusions.

For example, one might argue that, since our sense experiences are usually accurate, it is reasonable to form beliefs about the external world on their basis.

But how do we know that our sense experiences are usually accurate? Presumably on the basis of past sense experiences. And those experiences can be doubted just as much as our present experiences. Any attempt to respond to Descartes seems to assume the very thing we are trying to show.

If you think about it, it seems like any attempt to reply to Descartes' argument is going to face this kind of problem.

The seeming impossibility of replying to Descartes' argument — and hence the seeming impossibility of providing a satisfactory proof of an external world — was seen by the great German philosopher Immanuel Kant as a "scandal to philosophy."

Descartes' reasoning relies on the following rule of belief:

> **Doubt → No Belief**
> If you cannot distinguish between a situation in which P and a situation in which not-P, do not believe P.

By the line of reasoning just laid out, this rule seems to lead directly to the (very) surprising conclusion that you should never form beliefs about the objects you (seem to) perceive.

Indeed, to show that this rule leads to these kinds of surprising consequences, we don't even need to consider scenarios as extreme as Descartes' 'evil demon' scenario. You are not now in a position to distinguish between a situation in which your dorm room was robbed five minutes ago and one in which it wasn't. It then seems to follow from Doubt → No Belief that you shouldn't believe that the possessions you left in your dorm room are safe.

You also should not believe that the person sitting next to you is not currently secretly plotting against you.

Doubt → No Belief
If you cannot distinguish between a situation in which P and a situation in which not-P, do not believe P.

Our second reading for today is from someone who has a very different perspective on our beliefs about the external world than Descartes did.

According to G.E. Moore, it is no "scandal to philosophy" that we cannot prove the existence of the external world — for in fact, he thought, proofs of this kind are extremely easy to give.

We can think of Moore as endorsing the following positive rule of belief:

Proof → Belief
If you can prove P, believe P.

At first glance, our two rules look perfectly consistent. Moore tries to show that they are not.

the rules of belief

doubt and belief

Moore's proof

He presents his proof of an external world in the following passage:

I can now give a large number of different proofs, each of which is a perfectly rigorous proof; and that at many other times I have been in a position to give many others. I can prove now, for instance, that two human hands exist. How? By holding up my two hands, and saying, as I make a certain gesture with the right hand, 'Here is one hand', and adding, as I make a certain gesture with the left, 'and here is another'. And if, by doing this, I have proved *ipso facto* the existence of external things, you will all see that I can also do it now in numbers of other ways: there is no need to multiply examples.

the rules of belief

doubt and belief

Moore's proof

I can now give a large number of different proofs, each of which is a perfectly rigorous proof; and that at many other times I have been in a position to give many others. I can prove now, for instance, that two human hands exist. How? By holding up my two hands, and saying, as I make a certain gesture with the right hand, 'Here is one hand', and adding, as I make a certain gesture with the left, 'and here is another'. And if, by doing this, I have proved *ipso facto* the existence of external things, you will all see that I can also do it now in numbers of other ways: there is no need to multiply examples.

Moore's proof can be laid out as follows:

1. Here is one hand.
2. Here is another hand.

———————————

C. There are two hands. (1,2)

1. Here is one hand.
2. Here is another hand.

———————————————

C. There are two hands. (1,2)

It is, I think, safe to say that this is not the sort of proof that Moore's audience was expecting. We might ask: what does Moore mean when he says that this is a proof?

Moore tells us. He says that an argument is a proof if it satisfies three conditions:

Moore's definition of a proof

(1) Its premises are distinct from its conclusion.
(2) Its premises are known to be true.
(3) Its conclusion follows from its premises.

It is natural to find Moore's proof a little bit puzzling — and unsatisfying. But let's separate out two different questions which we can ask about his proof.

the rules of belief

doubt and belief

Moore's proof

1. Here is one hand.
2. Here is another hand.

———————————————

C. There are two hands. (1,2)

Moore's definition of a proof

(1) Its premises are distinct from its conclusion.
(2) Its premises are known to be true.
(3) Its conclusion follows from its premises.

It is natural to find Moore's proof a little bit puzzling — and unsatisfying. But let's separate out two different questions which we can ask about his proof.

If an argument meets Moore's definition of a proof, does it provide knowledge of its conclusion?

Does Moore's argument meet his definition of a proof?

If an argument meets Moore's definition of a proof, does it provide knowledge of its conclusion?

Moore's definition of a proof

(1) Its premises are distinct from its conclusion.
(2) Its premises are known to be true.
(3) Its conclusion follows from its premises.

You might think, at first, yes: if we know the premises, and the conclusion follows from the premises, doesn't this give us knowledge of the conclusion?

Well, not quite. It might be the case that the conclusion actually follows from the premises, but that we don't know that it does. Here's an example:

1. There are infinitely many numbers.
2. A prime number is one whose only divisors are 1 and itself.

———————————————

C. There are infinitely many prime numbers. (1,2)

If an argument meets Moore's definition of a proof, does it provide knowledge of its conclusion?

Moore's definition of a proof

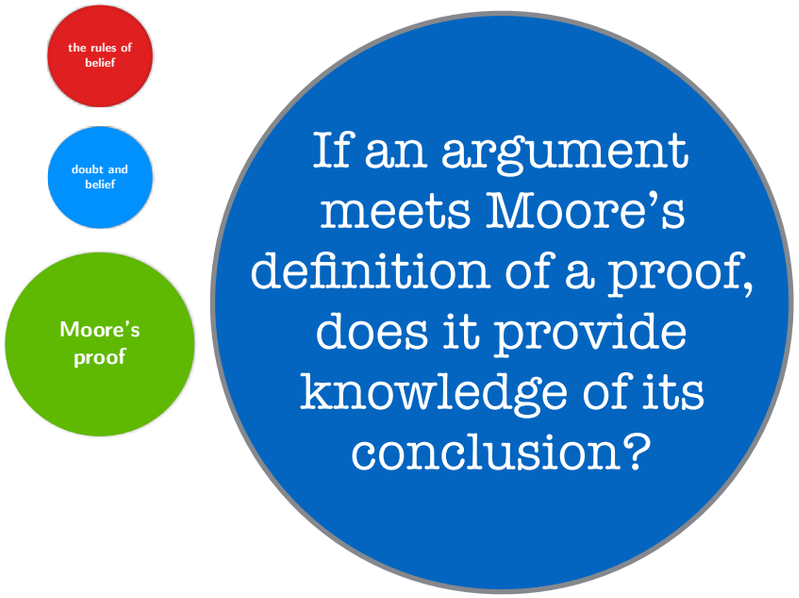
(1) Its premises are distinct from its conclusion.
(2) Its premises are known to be true.
(3) Its conclusion follows from its premises.

1. There are infinitely many numbers.
2. A prime number is one whose only divisors are 1 and itself.

C. There are infinitely many prime numbers. (1,2)

This meets Moore's definition of a proof, and yet might not provide knowledge of its conclusion to someone who does not know that it is valid.

This suggests a slight modification of Moore's definition.



Moore's definition of a proof

(1) Its premises are distinct from its conclusion.
(2) Its premises are known to be true.
(3*) Its conclusion is known to follow from its premises.

With this modification in hand, can we conclude that if an argument meets Moore's definition of a proof, then it provides knowledge of its conclusion?

Does Moore's argument meet his definition of a proof?

Moore's definition of a proof

(1) Its premises are distinct from its conclusion.
(2) Its premises are known to be true.
(3*) Its conclusion is known to follow from its premises.

1. Here is one hand.
2. Here is another hand.

———————————————

C. There are two hands. (1,2)

Let's turn now to the question of whether Moore's argument does in fact meet his definition.

It obviously meets condition (1); and also pretty obviously meets condition (3*). So our question boils down to this one: does Moore really know the premises of his argument?

Does Moore's argument meet his definition of a proof?

1. Here is one hand.
2. Here is another hand.

———————————————

C. There are two hands. (1,2)

It is pretty easy to adapt our earlier argument for skepticism about the external world to make an argument that Moore does not know the premises of his argument.

1. I can imagine a situation which is indistinguishable from a visual experience of my hands but in which I have no hands.
2. If I cannot distinguish between two situations, then I cannot know which of them is real.
3. I can never know whether a visual experience of my hands is accurate. (1,2)
4. If I cannot know whether any of my sense experiences are accurate, I cannot know whether there is an external world.

———————————————————————————-

C. I do not know whether I have hands. (3,4)

Moore anticipates the objection that he does not know the premises of his argument, and responds as follows:

(2) I certainly did at the moment *know* that which I expressed by the combination of certain gestures with saying the words 'There is one hand and here is another'. I *knew* that there was one hand in the place indicated by combining a certain gesture with my first utterance of 'here' and that there was another in the different place indicated by combining a certain gesture with my second utterance of 'here'. How absurd it would be to suggest that I did not know it, but only believed it, and that perhaps it was not the case! You might as well suggest that I do not know that I am now standing up and talking—that perhaps after all I'm not, and that it's not quite certain that I am!

Moore is emphasizing the fact that, in ordinary life, we do take ourselves to know claims like the premises of his argument. So why should we now, once we start doing philosophy, discard these beliefs?

Here is a different way to put the same point. We have, it seems, a conflict between the following two claims:

2. If I cannot distinguish between two situations, then I cannot know which of them is real.

I know that I have hands.

One can think of Moore as asking the proponent of our skeptical argument: which of these do you feel more sure of? Which, if you had to, would you bet your life on?