

the
knowledge
argument

the
conceivability
argument

swapping
arguments

the
interaction
argument

Are there
immaterial
souls, and
could you
be one?

Our topic today is the dualist view that we are (either wholly or in part) immaterial things.

You may recall that we came up with two dualist answers to the identity question:

Simple dualism: I am an immaterial soul

Fusion dualism: I am a combination of a soul and a body

Corresponding to these were two answers to the survival question:

Soul survival:
X is me just in case X is the same immaterial soul as me

Fusion survival:
X is me just in case X has the same immaterial soul as me and is the same material thing as me

the knowledge argument

the conceivability argument

swapping arguments

the interaction argument

Simple dualism: I am an immaterial soul

Soul survival:
X is me just in case X is the same immaterial soul as me

Fusion dualism: I am a combination of a soul and a body

Fusion survival:
X is me just in case X has the same immaterial soul as me and is the same material thing as me

For right now we're going to set aside the question of which of these views is better.

Instead, we're going to ask a more fundamental question: is there any good reason to believe in the existence of immaterial souls?

Most of us don't believe in fairies, there being no good evidence for their existence. Why should souls be any different?

the knowledge argument

the conceivability argument

swapping arguments

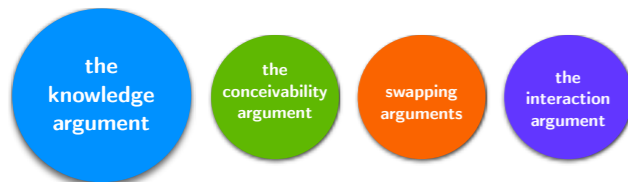
the interaction argument

Most of us don't believe in fairies, there being no good evidence for their existence. Why should souls be any different?

Imagine that aliens came to earth, and that these aliens were convinced materialists. Suppose that they endorse the organism view of human beings. Now suppose that you had to explain to them why you take yourself to be an immaterial thing rather than an organism. How would you do it?

We're going to look at two different arguments for the existence of souls.

The first begins with the idea that souls are required to explain the existence of **consciousness**.



The first begins with the idea that souls are required to explain the existence of **consciousness**.

Properties related to conscious experience include the property of feeling an itch or a pain, or the sensation of seeing red or hearing a loud noise. These properties are sometimes called **phenomenal properties**.

What do materialist theories of persons imply about phenomenal properties? Here is what Peter van Inwagen says:

“If a human person is a physical thing, any change whatever in a human person must be a purely physical change. If, for example, Tim becomes elated because of some news contained in a letter he has just received, this change in Tim, his becoming elated, must be the very same thing as some purely physical change.”

the
knowledge
argument

the
conceivability
argument

swapping
arguments

the
interaction
argument

“If a human person is a physical thing, any change whatever in a human person must be a purely physical change. If, for example, Tim becomes elated because of some news contained in a letter he has just received, this change in Tim, his becoming elated, must be the very same thing as some purely physical change.”

This seems to be a statement of the general principle that if human beings are wholly physical things, then all of their properties must be physical properties.

Conversely, this principle implies that if some of our properties are **not** physical properties, then we must not be wholly physical things.

So let's ask: are phenomenal properties physical properties?

the
knowledge
argument

the
conceivability
argument

swapping
arguments

the
interaction
argument

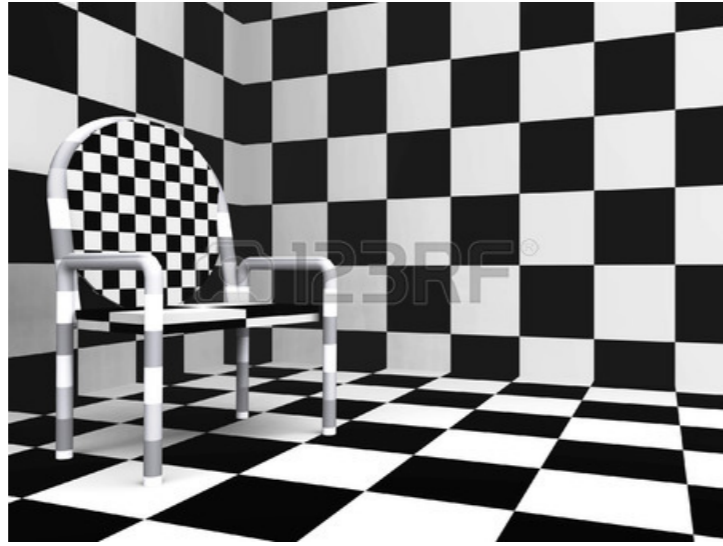
Conversely, this principle implies that if some of our properties are **not** physical properties, then we must not be wholly physical things.

So let's ask: are phenomenal properties physical properties?

This is the view that the philosopher Frank Jackson tried to refute with his example of Mary and the black-and-white room.



Mary is a brilliant scientist who has been confined her entire life to an environment in which everything is colored white or black.



Mary is a brilliant scientist who has been confined her entire life to an environment in which everything is colored white or black.

She is so brilliant, in fact, that she has learned every fact that there is to learn about the physical world. In particular, she has learned all of the facts about the neurophysiology of color vision, and has studied extensively everything that happens to the brain when subjects are experiencing color.

So she knows all of the physical properties that brains have when the person whose brain it is is experiencing color. It seems like this should be possible; people who are color blind can still learn physics.

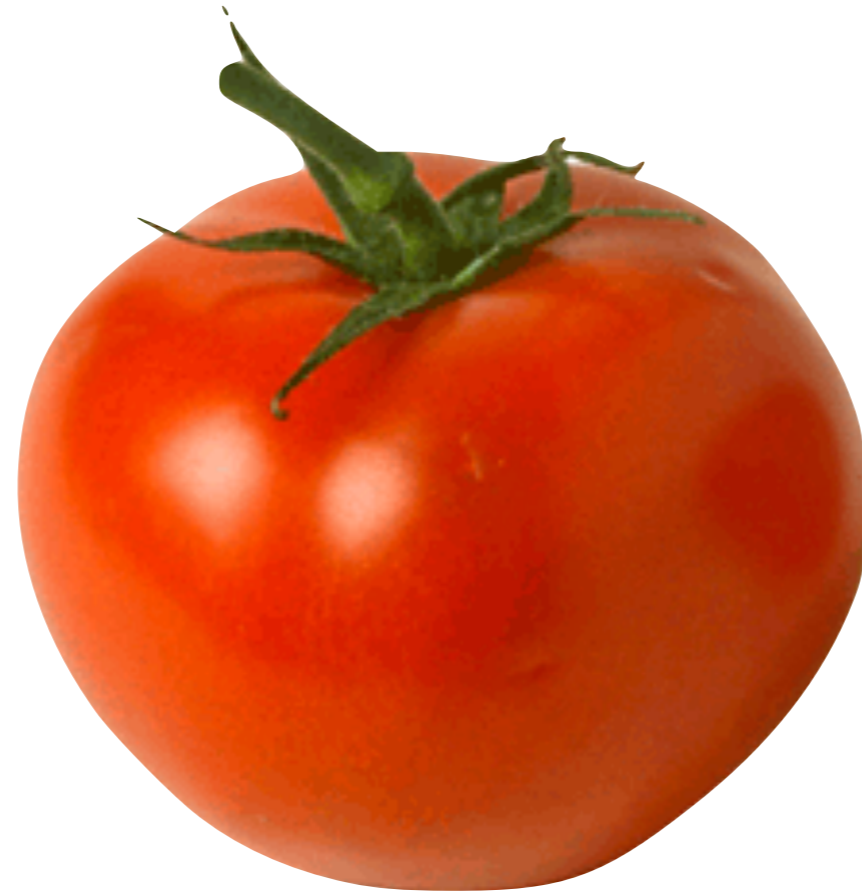
the
knowledge
argument

the
conceivability
argument

swapping
arguments

the
interaction
argument

One day, someone brings a new object into Mary's room:



Does Mary learn anything new when she sees the tomato?

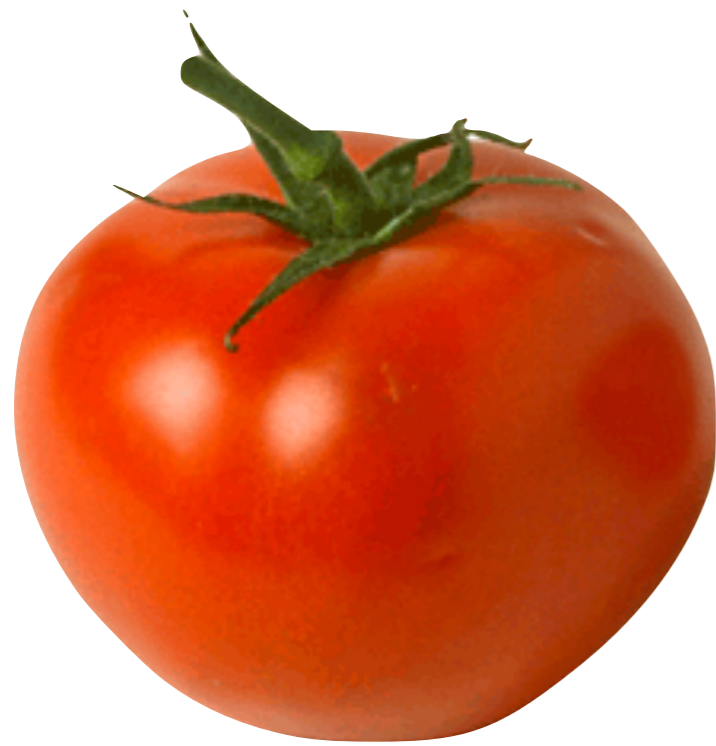
Intuitively, yes: she learns **what it is like to experience red things**. It seems that she learns something about the phenomenal property of sensing redness.

the
knowledge
argument

the
conceivability
argument

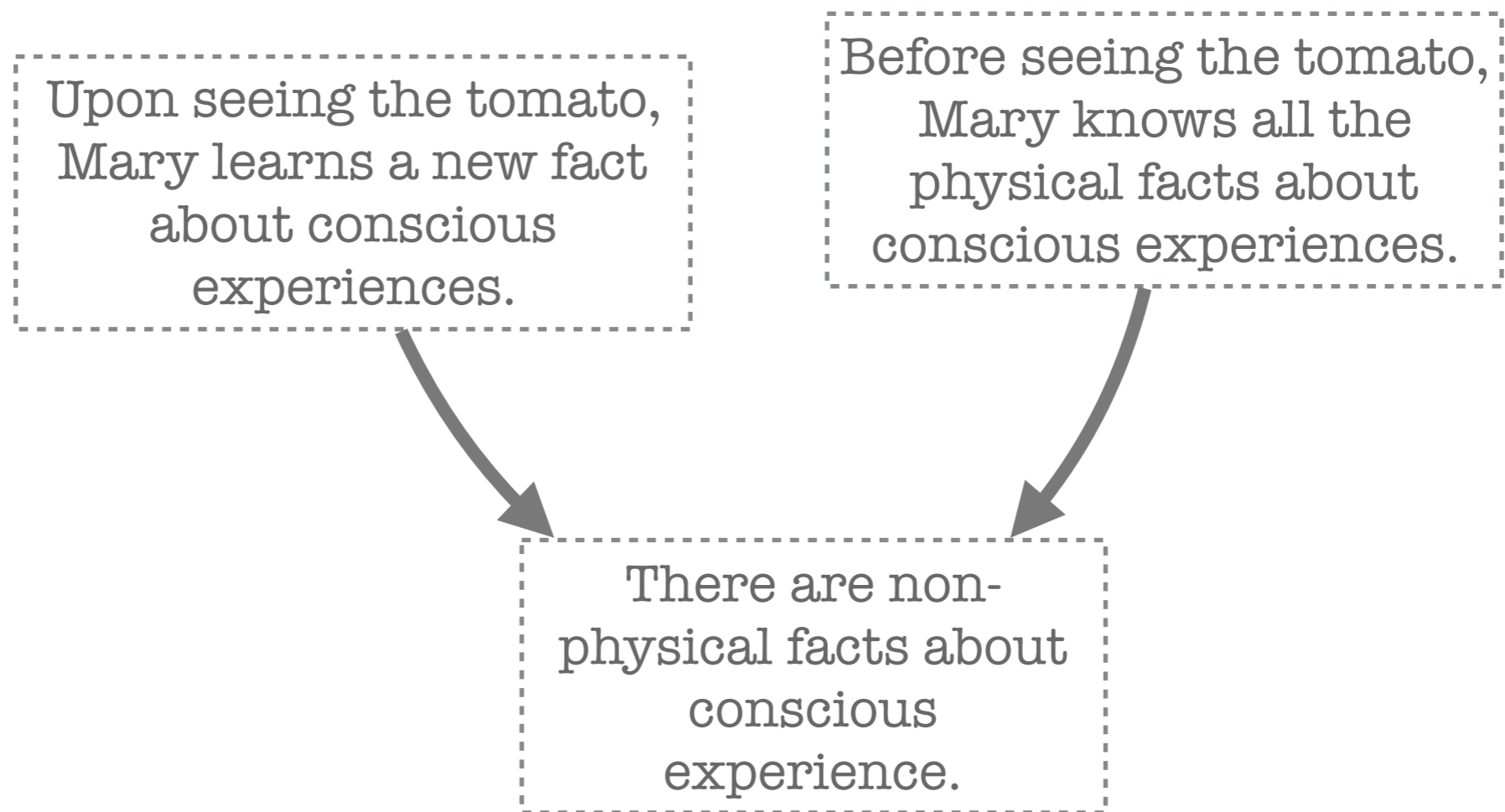
swapping
arguments

the
interaction
argument



But this seems to show that phenomenal properties are not physical properties.

Here is one way in which the argument, which is sometimes called the **knowledge argument**, can be laid out.



Upon seeing the tomato,
Mary learns a new fact
about conscious
experiences.

Before seeing the tomato,
Mary knows all the
physical facts about
conscious experiences.

If phenomenal properties are
physical properties, then all
of the facts about conscious
experiences are physical
facts.

There are non-
physical facts about
conscious
experience.

Phenomenal
properties are not
physical properties.

If people are wholly
physical, then all of
their properties are
physical properties.

People are not wholly
physical things.

the
knowledge
argument

the
conceivability
argument

swapping
arguments

the
interaction
argument

The knowledge argument

1. If phenomenal properties are physical properties, then all of the facts about conscious experiences are physical facts.
2. Before seeing the tomato, Mary knows all the physical facts about conscious experiences.
3. Upon seeing the tomato, Mary learns a new fact about conscious experiences.
4. There are non-physical facts about conscious experience. (2,3)
5. Phenomenal properties are not physical properties. (1,4)
6. If people are wholly physical, then all of their properties are physical properties.

C. People are not wholly physical things. (5,6)

Is you were a materialist, how would you reply?

the
knowledge
argument

the
conceivability
argument

swapping
arguments

the
interaction
argument

The knowledge argument

1. If phenomenal properties are physical properties, then all of the facts about conscious experiences are physical facts.
2. Before seeing the tomato, Mary knows all the physical facts about conscious experiences.
3. Upon seeing the tomato, Mary learns a new fact about conscious experiences.
4. There are non-physical facts about conscious experience. (2,3)
5. Phenomenal properties are not physical properties. (1,4)
6. If people are wholly physical, then all of their properties are physical properties.

C. People are not wholly physical things. (5,6)

One of the most popular responses to the knowledge argument from materialists involves denying that (3) is true.

Here's one way to make this seem plausible. Consider Lois Lane. She knows that Superman can fly; and she knows who Clark Kent is. But she does not know that Clark Kent is Superman.

Here's one way to make this seem plausible. Consider Lois Lane. She knows that Superman can fly; and she knows who Clark Kent is. But she does not know that Clark Kent is Superman.

Suppose that one day she sees Clark don the Superman costume and fly away. She is surprised — just like Mary is surprised.

If you asked her why she is surprised, she might say, "I did not know that Clark could fly!"

But of course in a way she did. She knew that Superman could fly. And Clark = Superman. So isn't the fact that Superman can fly just the same as the fact that Clark can fly?

It looks like Lois is surprised, not because there is some new fact that she learns, but because (in some sense) she learns a new way of thinking about a fact she already knew.

Could the materialist say that, similarly, Mary does not learn a new fact, but instead learns a new way of thinking about a physical fact she already knew?

Let's turn to the second main argument for dualism. This argument is due to René Descartes. Descartes was one of the most important philosophers who ever lived — a distinction which is especially impressive given that he devoted most of his energies to mathematics (in which he developed what is now analytic geometry) and natural science.

In 1649 Descartes moved to Sweden to join the court of Queen Christina of Sweden. After complaining that "men's thoughts are frozen here, like the water," Descartes died in February of 1650, during his first winter in Sweden.





Descartes' argument begins with his thought that all of our beliefs about the existence of material things can be called into doubt:

“Every sensory experience I have ever thought I was having while awake I can also think of myself as sometimes having while asleep. Since I do not believe that what I seem to perceive in sleep comes from things located outside me, I did not see why I should be any more inclined to believe this of what I think I perceive while awake.”

Descartes is saying that we can imagine any sensory experience we have occurring in sleep rather than waking life. But in sleep our seeming sensory experiences do not reflect the reality of the material world around us; so, we can imagine all of the sensory experiences we have failing to reflect the world around us. That is, we can coherently imagine a scenario in which there are no tissue boxes, cats, planets, or other material things, even though in our experience it seems to us that there are such things.

Now let's ask another question: when we conceive of the possibility that there are no material things, are we conceiving a situation in which **nothing at all** exists?



Now let's ask another question: when we conceive of the possibility that there are no material things, are we conceiving a situation in which **nothing at all** exists?

“I have convinced myself that there is absolutely nothing in the world, no sky, no earth, no minds, no bodies. Does it now follow that I too do not exist?

No: if I convinced myself of something then I certainly existed. ... This proposition - *I am, I exist* - is necessarily true whenever it is put forward by me or conceived in my mind.”



Descartes here seems to be saying that, when I imagine a world in which there are no material things, I am still imagining that I exist. This suggests the following claim:

I can clearly imagine a scenario in which I exist, but no material things exist.

Suppose that this claim about imagination is true. What could this have to do with the question of what I am? We aren't, after all, interested in what we can imagine about ourselves; we are interested in the question of what sorts of things we really are.



The answer to this question comes in the following passage:

“I know that everything which I clearly and distinctly understand is capable of being created by God so as to correspond exactly with my understanding of it. Hence the fact that I can clearly and distinctly understand one thing apart from another is enough to make me certain that the two things are distinct.”

Each of the two sentences in this passage makes a claim which is central to Descartes' argument. Let's focus on the first one first.



“I know that everything which I clearly and distinctly understand is capable of being created by God so as to correspond exactly with my understanding of it. Hence the fact that I can clearly and distinctly understand one thing apart from another is enough to make me certain that the two things are distinct.”

Descartes seems to be saying that if I can clearly imagine something to be the case, then God could make it the case: God could bring it about. It seems to follow from this that Descartes would endorse the following principle:

If I can clearly imagine something being the case, then it is possible for it to be the case.

Is there any reason to think that this is true?



I can clearly imagine a scenario in which I exist, but no material things exist.

If I can clearly imagine something being the case, then it is possible for it to be the case.

It is possible that I exist and no material things exist.

It is possible that I \neq any material thing.

What premise could we insert to get us to the intended conclusion?

I \neq any material thing.



I can clearly imagine a scenario in which I exist, but no material things exist.

If I can clearly imagine something being the case, then it is possible for it to be the case.

It is possible that I exist and no material things exist.

It is possible that I ≠ any material thing.

If it is possible that P, then P is true.

I ≠ any material thing.



Let's go back to Descartes' text to see what the missing premise could be.

“I know that everything which I clearly and distinctly understand is capable of being created by God so as to correspond exactly with my understanding of it. Hence the fact that I can clearly and distinctly understand one thing apart from another is enough to make me certain that the two things are distinct.”

Here Descartes does not seem to be asserting the unrestricted (and absurd) claim that anything possible is true; rather, he's asserting the following more restricted principle:

⋮If it is possible that $x \neq y$, then $x \neq y$.⋮

Is this principle true?



If it is possible that $x \neq y$, then $x \neq y$.

A good case can be made that it is. Consider first the following principle:

The principle of the necessity of self-identity

For any object x , necessarily, $x = x$.

In ordinary English, one might state the principle of the necessity of self-identity as the claim that it is impossible for a thing to be distinct from itself. This principle seems true: it does not seem possible that you could have existed without being yourself - in that case, one wants to say, it would not have been you that existed!

But if this is true, then the following principle also seems true:

If $x = y$, then necessarily, $x = y$.

After all, if x and y are literally the same thing, then the same reasoning which seems to show that the principle of the necessity of self-identity is true also seems to show that this principle is true.



If $x=y$, then necessarily, $x=y$.

This follows because 'if p then q ' is true, so is 'if not- q then not- p .'

If it is not necessary that $x=y$, then $x \neq y$.

This follows by the definition of necessity and possibility.

If it is possible that $x \neq y$, then $x \neq y$.

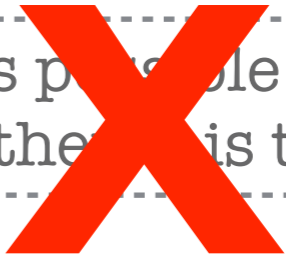
I can clearly imagine a scenario in which I exist, but no material things exist.

If I can clearly imagine something being the case, then it is possible for it to be the case.

It is possible that I exist and no material things exist.

It is possible that I ≠ any material thing.

If it is possible that P, then P is true.



I ≠ any material thing.

the knowledge argument the conceivability argument the swamping argument the interaction argument

If it is possible that $x \neq y$, then $x \neq y$.

The conceivability argument

1. I can clearly imagine a scenario in which I exist, but no material things exist.
2. If I can clearly imagine something being the case, then it is possible for it to be the case.
3. It is possible that I exist and no material things exist. (1,2)
4. It is possible that $I \neq$ any material thing. (3)
5. If it is possible that $x \neq y$, then $x \neq y$.

C. $I \neq$ any material thing. (4,5)

This is sometimes called the **conceivability argument** for dualism, since it rests heavily on a claim about what we can conceive of, or imagine.

Suppose that you are a fusion dualist, who accepts fusion survival. Could you endorse this argument?

It seems not; it looks like you would have to reject (3), and hence also either (1) or (2).

This looks like one advantage of simple dualism over fusion dualism: the simple dualist, but not the fusion dualist, can defend her position with the conceivability argument.

The conceivability argument

1. I can clearly imagine a scenario in which I exist, but no material things exist.
2. If I can clearly imagine something being the case, then it is possible for it to be the case.
3. It is possible that I exist and no material things exist. (1,2)
4. It is possible that $I \neq$ any material thing. (3)
5. If it is possible that $x \neq y$, then $x \neq y$.

C. $I \neq$ any material thing. (4,5)

If you were either a fusion dualist or a materialist of some sort, how would you reply?

Let's consider an objection to the second premise of Descartes' argument:

2. If I can clearly imagine something being the case, then it is possible for it to be the case.

Here is a possible counterexample to this premise:

The barber

Deep in the Indiana countryside, there's a small town; and in this town there's a barber. Some of the men in this small town - the industrious ones - shave themselves every morning. But others (the lazy ones) don't; and the barber shaves all of them. (There's no one else around who will do it.) Moreover, he (the barber's a man) never shaves any of the industrious ones - he never shaves any of the men that shave themselves.

The barber

Deep in the Indiana countryside, there's a small town; and in this town there's a barber. Some of the men in this small town - the industrious ones - shave themselves every morning. But others (the lazy ones) don't; and the barber shaves all of them. (There's no one else around who will do it.) Moreover, he (the barber's a man) never shaves any of the industrious ones - he never shaves any of the men that shave themselves.

Can you clearly imagine this small town? Is it possible for there to be a town of this sort?



The barber

Deep in the Indiana countryside, there's a small town; and in this town there's a barber. Some of the men in this small town - the industrious ones - shave themselves every morning. But others (the lazy ones) don't; and the barber shaves all of them. (There's no one else around who will do it.) Moreover, he (the barber's a man) never shaves any of the industrious ones - he never shaves any of the men that shave themselves.

Does the town's barber shave himself, or not?

No.

Yes.

But then he does shave himself because he shaves every man that does not shave himself.

But then he doesn't, because he doesn't shave any man that shaves himself.

So if he shaves himself, then he doesn't, and if he doesn't, he does.

the
knowledge
argument

the
conceivability
argument

swapping
arguments

the
interaction
argument

This is thus an example of a scenario which, at first glance, seems possible, but then turns out, on closer inspection to be impossible, because it contains a hidden contradiction. Might the materialist plausibly say the same thing about Descartes' scenario - the imagined scenario in which I exist, but there are no material things?

This is an appealing thought, if you are a materialist. But you should ask yourself: what contradiction could this be? What could be impossible about you existing in the absence of any material things?



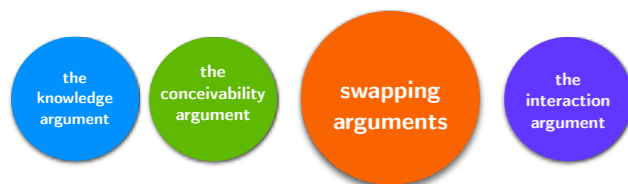
We've now encountered the main arguments for the view that you are, either entirely or in part, an immaterial thing.

We have already encountered two kinds of arguments against that view.

The first are the "easy arguments" for materialism. Those seem to be more of a problem for the simple dualist than for the fusion dualist.

The second is the argument that we should not believe in souls for the same reason we don't believe in fairies: there is no good evidence for their existence. That might be countered with the knowledge argument or the conceivability argument.

Let's look at some others.



Let's look at some others.

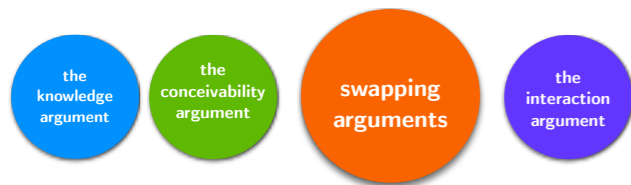
I'll call the first class of arguments against dualism **swapping arguments**.

Let's suppose that dualism is true. Then your roommate is (either wholly or in part) an immaterial soul.

Here is something that seems possible: last night, the soul connected to your roommate's body could have been swapped out for a different one which has exactly the same apparent memories and personality.

You would, it seems, have no way of telling whether this happened. So you have no evidence that it did not happen. So, you do not know whether you have the same roommate as you did yesterday.

But of course you **do** know this. So dualism must be false.

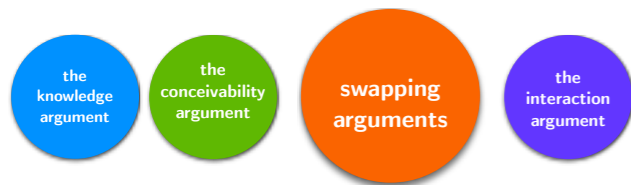


But of course you **do** know this. So dualism must be false.

More dramatically, the soul attached to your own body could have been swapped out last night. It seems that you would have no way of knowing whether this happened. After all, the new soul (which you are) could have been given the same personality and apparent memories as the old soul.

So you do not know whether you have been attached to this body for more than a day. But of course you do know that you have been attached to this body for more than a day. So, dualism must be false.

Call this the **argument from soul-swapping**. How should the dualist respond?



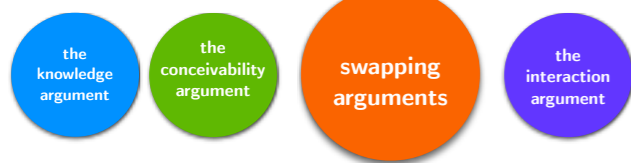
Call this the **argument from soul-swapping**. How should the dualist respond?

Here is a somewhat related argument.

Psychology-swapping

Ferdinand is a combination of a soul and a body. Suppose that Ferdinand suffers an accident which entirely changes his psychology. He has none of his former memories, and none of his former personality traits. He used to be very selfish; he's now very generous. He used to be best friends with Jane; now he finds Jane annoying, and prefers to spend time with Maria.

Let's call Ferdinand before the accident "Ferdinand-" and Ferdinand after the accident "Ferdinand+".



Psychology-swapping

Ferdinand is a combination of a soul and a body. Suppose that Ferdinand suffers an accident which entirely changes his psychology. He has none of his former memories, and none of his former personality traits. He used to be very selfish; he's now very generous. He used to be best friends with Jane; now he finds Jane annoying, and prefers to spend time with Maria.

Let's call Ferdinand before the accident "Ferdinand-" and Ferdinand after the accident "Ferdinand+".

Suppose that Ferdinand- used to occasionally take some cash from your wallet. Ferdinand+ would never do this. Would you be right to blame Ferdinand+ for the actions of Ferdinand-?

If not, then it looks like you are treating Ferdinand+ and Ferdinand- as different people.

Suppose that Ferdinand- used to occasionally take some cash from your wallet. Ferdinand+ would never do this. Would you be right to blame Ferdinand+ for the actions of Ferdinand-?

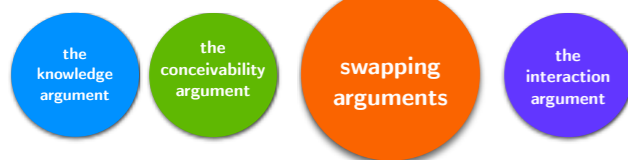
If not, then it looks like you are treating Ferdinand+ and Ferdinand- as different people.

But if they are different people, it looks like our dualist answers to the survival question are incorrect.

For there is no obvious reason why the numerically same soul could not be attached to Ferdinand's body throughout this process.

This points toward something more like psychological survival or small changes survival.

Call this the **psychology-swapping argument** against dualism.



Let's set the swapping arguments to the side. The historically most influential argument against dualism is one originally raised by Princess Elisabeth of Bohemia.

Elisabeth was one of the leading intellectuals of the 17th century. She worked in mathematics and physics as well as philosophy, and was active in German politics. She was known by her siblings as 'The Greek' because she mastered ancient Greek at such a young age.



Today Elisabeth is best known for her correspondence with Descartes, in which she raised the question of how an immaterial soul and a material body could interact.

Today Elisabeth is best known for her correspondence with Descartes, in which she raised the question of how an immaterial soul and a material body could interact.



Elisabeth pointed out that, if dualism is true, then it must happen quite often that the body causes effects in the soul, and that the soul causes effects in the body.

What might be some examples of your body causing effects in your soul?

How about examples of your soul causing effects in your body?

Consideration of examples show that, if dualism is true, then interactions between soul and body must happen all of the time. But Elisabeth argued that these kinds of causal interactions were entirely mysterious.

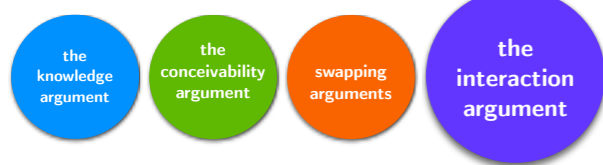
Consideration of examples show that, if dualism is true, then interactions between soul and body must happen all of the time. But Elisabeth argued that these kinds of causal interactions were entirely mysterious.

This gives rise to the **interaction argument against dualism**.

“it seems every determination of movement happens from the impulsion of a thing moved, according to the manner in which it is pushed by that which moves it ... Contact is required for [this]. ... contact seems to me incompatible with the idea of an immaterial thing.”



What's her argument here?



“it seems every determination of movement happens from the impulsion of a thing moved, according to the manner in which it is pushed by that which moves it ... Contact is required for [this]. ... contact seems to me incompatible with the idea of an immaterial thing.”

What’s her argument here?

Elisabeth here is considering a special case of causation: putting something in motion. For something to push something else, it seems that the two things must be in contact; and for two things to be in contact, both must occupy space (since being in contact is just a matter of occupying adjacent spaces). Since immaterial minds don’t occupy space, it seems that they can’t set things in motion — so, for example, my mind’s desire for coffee can’t be what sends my body down the hallway in search of some.

Elisabeth here is considering a special case of causation: putting something in motion. For something to push something else, it seems that the two things must be in contact; and for two things to be in contact, both must occupy space (since being in contact is just a matter of occupying adjacent spaces). Since immaterial minds don't occupy space, it seems that they can't set things in motion — so, for example, my mind's desire for coffee can't be what sends my body down the hallway in search of some.

We might represent this argument simply like this:

The interaction argument

- 1. Minds and bodies causally interact.
 - 2. The only things that causally interact with physical things are other physical things.
 - 3. Minds are physical things. (1,2)
-
- C. Minds are not immaterial souls. (3)

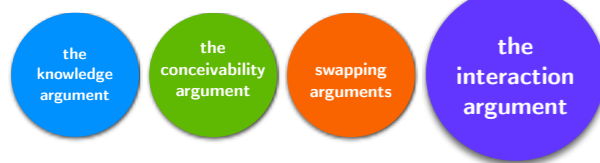
The interaction argument

1. Minds and bodies causally interact.
 2. The only things that causally interact with physical things are other physical things.
 3. Minds are physical things. (1,2)
-
- C. Minds are not immaterial souls. (3)

How might the dualist reply? On natural line of thought, which Descartes pursues, is to reject premise (2). Descartes argues that while some causation does work through contact, not all does. He gives the example of gravity.

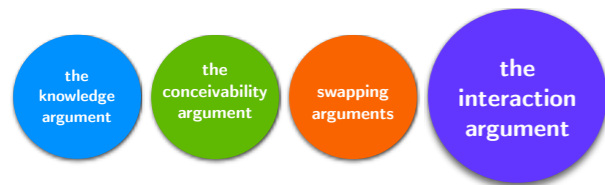
Elisabeth was skeptical about the idea that the example of gravity could provide a model for the interaction between mind and body. And many philosophers since have been on Elisabeth's side here: the idea of an immaterial thing, which is not located in space, interacting with a material thing does seem a bit puzzling.

Let's consider two ways of defending premise (2) of Elisabeth's argument.

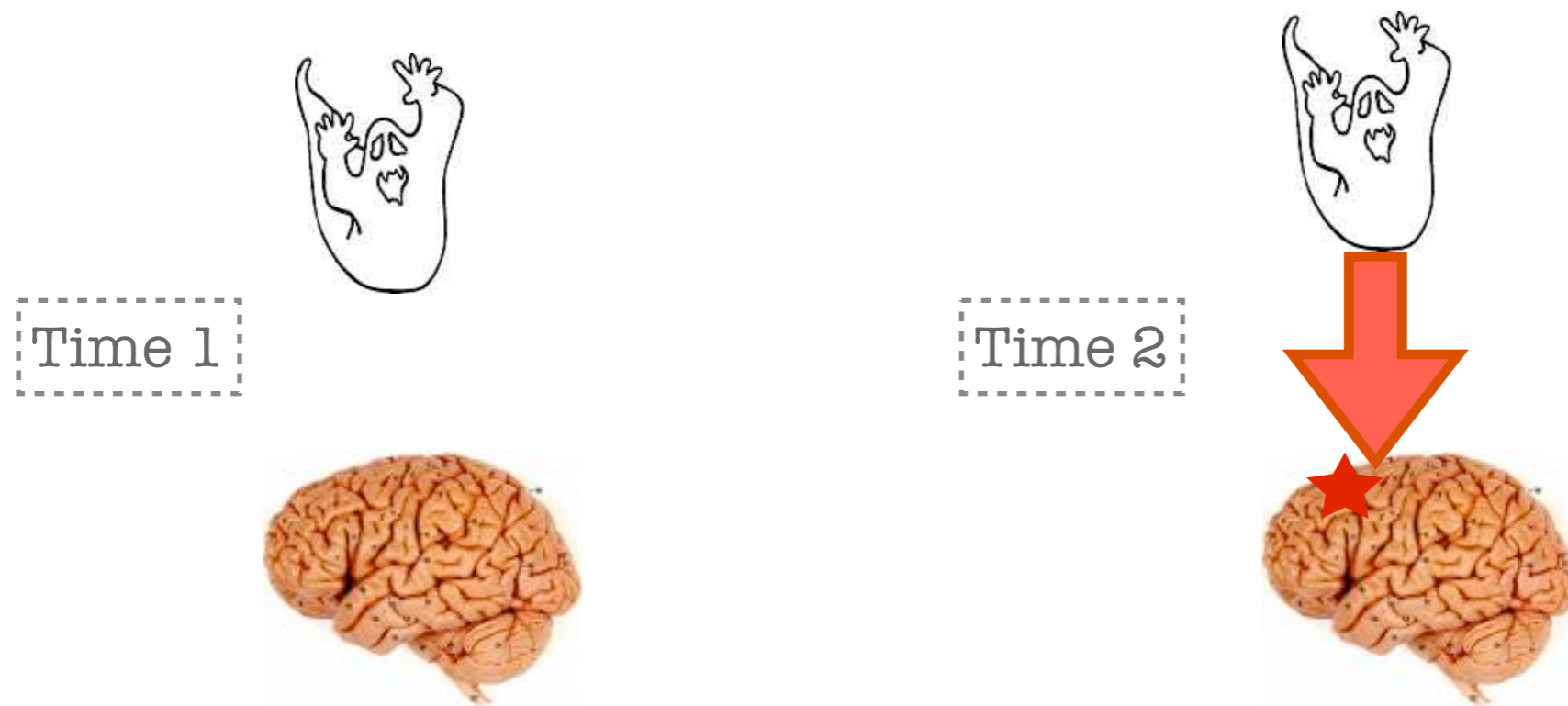


A first attempt is to argue that the dualist is committed to the violation of certain fundamental physical laws, such as the [law of the conservation of energy](#). This law says that the total energy of a closed physical system is constant; that the total energy of such a system may be neither increased nor decreased, but only transformed.

It seems, at first glance, that the interactionist dualist should be committed to denying this fundamental principle of physics. For consider a case in which an immaterial soul causes a change in the physical world - say, a case in which an immaterial soul causes a neuron to fire in the brain.



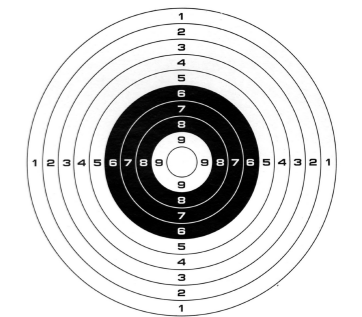
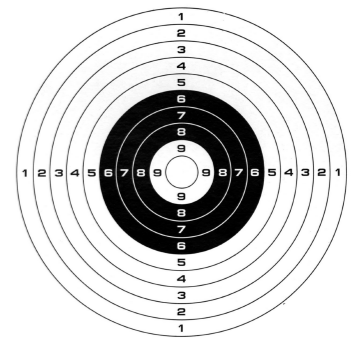
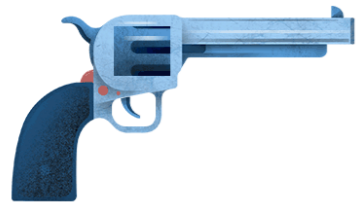
It seems, at first glance, that the interactionist dualist should be committed to denying this fundamental principle of physics. For consider a case in which an immaterial soul causes a change in the physical world - say, a case in which an immaterial soul causes a neuron to fire in the brain.



Now consider the physical system of which the brain is a part, at time 1 and then at time 2. Won't those two physical systems differ in their total energy? After all, everything is the same in those physical systems other than the activity of this neuron; and if it fires at one time but not the other, mustn't this involve a change in energy?

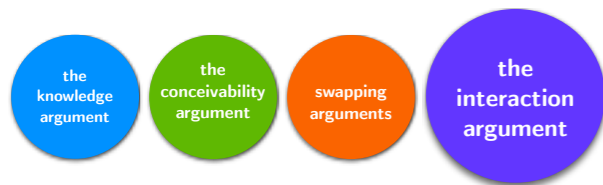
A second way to further Elisabeth's argument relies not on the idea that dualism violates certain physical laws, but on a certain kind of thought experiment.

Imagine that we have two guns aimed at distinct targets.



Presumably one gun caused one of the bullets to hit one of the targets, and the other gun caused the other bullet to hit the other. But what connects one firing to one of the targets, and the other to the other?

Easy answer: we trace the path of the bullet through space, from gun to target. This series of spatial connections is what connects the cause to the effect.



But now imagine that we have two immaterial souls, and two bodies.

Soul 1



Soul 2

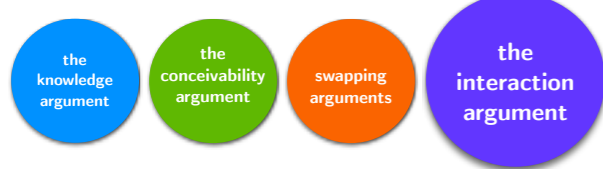


Now imagine that, at the same time, Soul 1 and Soul 2 decide to go for a walk.

Presumably one of the souls caused one of the bodies to walk, and the other soul caused the other body to walk. But which caused which?

Note that we can't answer this question in the same way that we answered the corresponding question in the case of the guns and bullets, for there is no path through space from the souls to the bodies.

This is sometimes called the **pairing problem**: it is the problem of explaining what pairs causes with effects, when either the cause or the effect is something immaterial.



This is sometimes called the **pairing problem**: it is the problem of explaining what pairs causes with effects, when either the cause or the effect is something immaterial.

Both the pairing problem and the problem with conservation laws are ways of defending the key premise of the interaction argument against dualism. The dualist seems committed to the existence of causal relations between immaterial souls and material things which are quite different than the causal interactions we perceive in the world and are, for that reason, mysterious.

