



---

**Scientific Reasoning  
Lecture-Seminar 4  
'Thought-Experiments'**

Dr. Ioannis Votsis  
NCH, Philosophy Faculty

[ioannis.votsis@nchlondon.ac.uk](mailto:ioannis.votsis@nchlondon.ac.uk)



---

# Introduction

# From the pre-Socratics to the modern day

- Often in narrative form and sometimes involve diagrams.
- They go back to the Pre-Socratics but gain prominence in the 17<sup>th</sup> century (Galileo, Newton and Leibniz).
- In 1811, Hans Christian Ørsted introduces ‘Tankeexperiment’ and in 1883 Ernst Mach introduces ‘Gadankenexperimente’.
- Systematic analyses have to wait until the 20th century.
- Brown & Fehige (2014): They are roughly: “devices of the imagination used to investigate the nature of things” (p. 1).

# A universal phenomenon?

- Such experiments can be found in a number of disciplines:
  - \* economics
  - \* history
  - \* literature
  - \* mathematics
  - \* philosophy
  - \* physics
  - \* psychology
  - \* sociology
  - \* ...



# Some central questions

- Can thought experiments teach us something about the world? If so, under what conditions and how?
- What is the relationship between the results of thought experiments and the results of 'regular' experiments?
- How, if at all possible, can we distinguish between good and bad thought experiments?



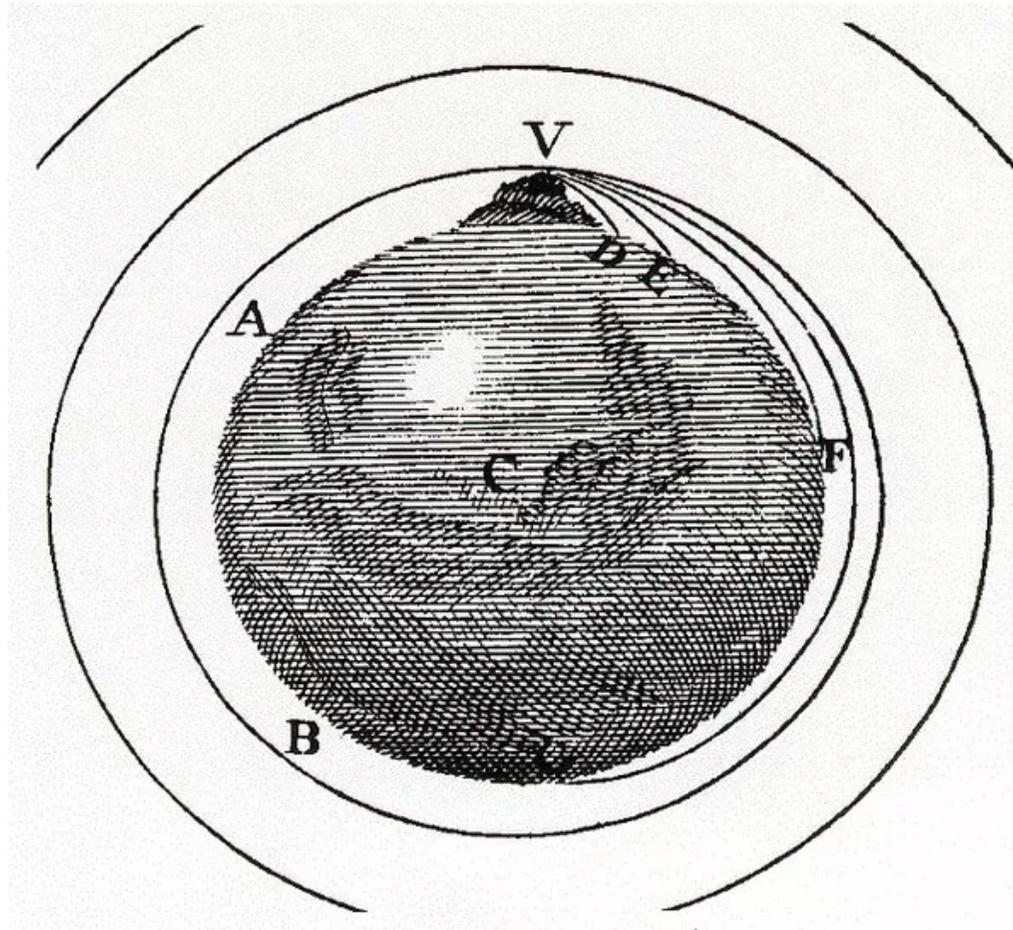
---

# Thought Experiments Classified

## What are they good for?

- **Clarification:** To help throw light on difficult/abstract ideas.

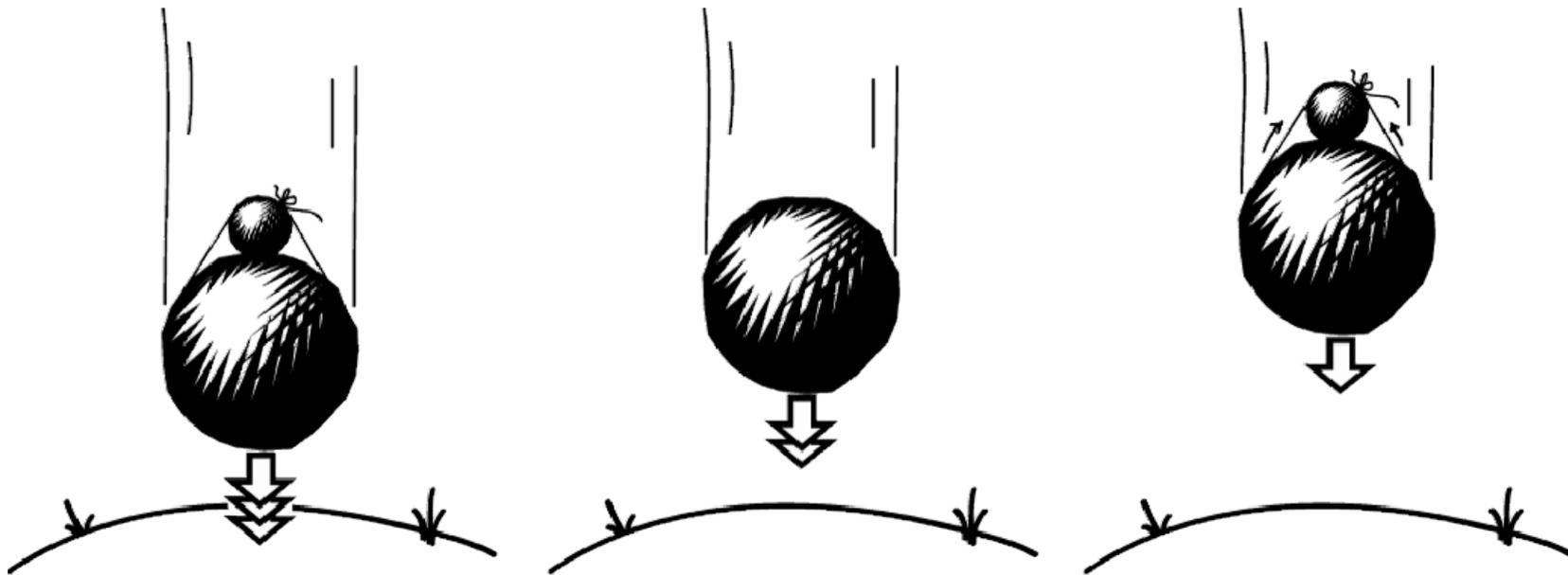
*Example:* Newton's explanation of how Moon stays in orbit.



## What are they good for? (2)

- **Testing:** To provide reasons for or against a hypothesis.

*Example:* Galileo's compound object thought-experiment.



# Why not just consult actual experiments?

- A number of reasons may be given:

physically

technologically

financially



impossible or currently unfeasible  
to conduct an actual experiment.

ethically impermissible

# Classification

- Different classification systems proposed over the years.
- One such system distinguishes between:

**Destructive:** Providing reasons to reject a hypothesis, e.g. by showing that it is internally or externally inconsistent.

vs.

**Constructive:** Providing reasons to accept a hypothesis, e.g. by illustrating its claims & thereby providing understanding.

NB: In rejecting, we provide reasons to accept the contrary hypothesis & vice-versa.

# The a-priori view

- Brown (1986): thought experiments help us acquire worldly truths a-priori. They thus transcend empirical knowledge.

*NB:* Also called the 'intuition view' because those considering such thought experiments are meant to share intuitions.

- **Objection:** Intuitions need not be shared as various experimental studies have shown and can lead us astray.

# The a-posteriori view

- Norton (1991): A thought experiment is an (explicit or implicit) argument with premises grounded in experience.

*NB:* Sometimes called the ‘argument view’.

- On this view, all sorts of narrative features of experiment may be psychologically helpful but not necessary.

- Objections:

(1) Presented as arguments they may lose their force.

(2) How exactly can we claim that the conclusions of such arguments provide *new empirical data*?

## The a-posteriori view: Example

1. A heavier mass falls faster than a lighter one.
  2.  $X$  is heavier in mass than  $Y$ .
  3.  $X$  and  $Y$  are tied together (via a rope) into  $Z$ .
  4.  $Z$  is heavier in mass than  $X$ .
- 
5. Thus,  $Z$  will fall faster than  $X$  (by 1 and 4).

6. If two tied masses differ in their initial speeds, then the faster will slow down because of drag from the slower one.
  7.  $X$  and  $Y$  differ in initial speeds (by 1 and 2).
  8.  $X$  (in  $Z$ ) will slow down because of drag from  $Y$ .
  9.  $Z$ 's speed is determined by the speed of its part  $X$ .\*
- 
10. Thus,  $Z$  will fall slower than  $X$  on its own (by 5).

**Contradiction:** Between 5 and 9. By *reductio ad absurdum*, a heavier mass does not fall faster than a lighter mass.

# How to counter thought-experiments

- Among other things, one can:
  - \* question the plausibility of the initial conditions.
  - \* cast doubt on the justifiability of the conclusion.
  - \* put forth thought experiments with contrary conclusions.
  - \* criticise the details as inadequate.
  - \* adopt a more general sceptical attitude towards them.



---

# Some Well-Known Thought Experiments

# Famous thought-experiments

- An (obviously) incomplete list:

Galileo's cannon and musket ball

Newton's bucket

Maxwell's demon

Einstein's elevator

Schrödinger's cat

Poincaré's sphere world

The sleeping beauty problem

The lottery paradox

Putnam's twin-earth

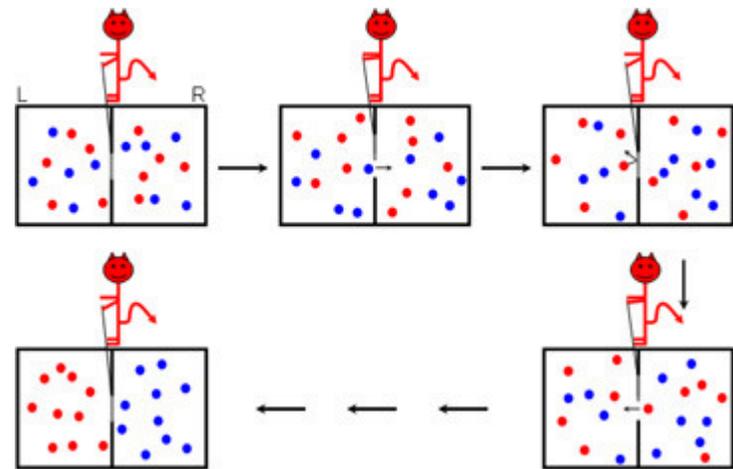
Searle's Chinese room argument

Thomson's violinist

Jackson's Mary in the black-and-white room

# Maxwell's demon

- Entropy is a measure of disorder in a closed system. On the classical 2<sup>nd</sup> law of thermodynamics, it always increases.
- Practically speaking, this means that heat *only* flows from hot to cold bodies until a thermal equilibrium is reached.
- Heat in this context is understood in terms of how energetically molecules move.
- This thought-experiment is meant to show, contra the classical law, that entropy *can* sometimes decrease.



# Searle's Chinese room argument

- This thought-experiment appears in Searle's (1980) 'Minds, Brains and Programs', *Behavioural and Brain Sciences*.
- It's incredibly influential, sparking a debate among computer scientists philosophers, cognitive scientists & AI researchers.  
  
"By 1991 computer scientist Pat Hayes had defined Cognitive Science as the ongoing research project of refuting Searle's argument" (Cole 2014).
- It seeks to show that mere manipulation of symbols, even if successful at language tasks, does not entail understanding.
- As a result, it directly challenges the Turing test.

# Searle's Chinese room argument



- More broadly, the argument is meant to show that human minds cannot be merely information processing systems.

# Thomson's violinist: The anti-abortion argument

- Judith Thomson's (1971) thought experiment is aimed against the following anti-abortion argument:

1. All innocent persons possess a right to life.
2. The fetus is (an) innocent (person).
3. Abortion results in the death of a fetus.
4. Such deaths violate the right to life.
5. Violating rights is morally wrong.

-----  
Therefore, abortion is morally wrong.

## Thomson's violinist: The thought experiment

- “You wake up in the morning and find yourself back to back in bed with an unconscious violinist. A famous unconscious violinist. He has been found to have a fatal kidney ailment, and the Society of Music Lovers has canvassed all the available records and found that you alone [can] help. They have therefore kidnapped you, and last night the violinist's circulatory system was plugged into yours, so that your kidneys can be used to extract poisons from his blood as well as your own. The director of the hospital now tells you ‘... To unplug you would be to kill him. But never mind, it's only for nine months. By then he will have recovered from his ailment, and can safely be unplugged from you’. Is it morally incumbent on you to accede to this situation?” (pp. 48-49).

# Thomson's violinist: The underlying argument

- Formulated as an argument it would look as follows:
  1. All innocent persons possess a right to life.
  2. The violinist is an innocent person.
  3. Unplugging the violinist results in his death.
  4. Such deaths violate the right to life.
  5. Violating rights is morally wrong.

-----  
Therefore, unplugging the violinist is morally wrong.

- **Hoped-for intuition:** Unplugging is not morally wrong.
- The two arguments are structurally identical. If we deny the conclusion of the one, we should deny that of the other.

# Jackson's Mary and the black-and-white room

- Another influential paper, 'What Mary Didn't Know', presents a thought experiment that seeks to disprove physicalism.
- Physicalism, roughly, is the view that that only physical things exist and facts about them can be fully captured by physics.

**NB:** Where physics here means the *final* theory of physics.

- The experiment is meant to demonstrate the existence and knowability of non-physical categories like qualia.
- What is a quale (singular of qualia)? It is a quality of what it feels like to experience something.

## Jackson's Mary: The thought experiment

- “Mary is confined to a black-and-white room... she learns everything there is to know about the physical nature of the world. She knows all the physical facts about us and our environment, in a wide sense of ‘physical’ which includes everything in *completed* physics, chemistry, and neurophysiology, and all there is to know about the causal and relational facts consequent upon all this... If physicalism is true, she knows all there is to know... It seems, however, that Mary does not know all there is to know. For when she is let out of the black-and-white room or given a color television, she will learn what it is like to see something red, say. This is rightly described as learning – she will not say ‘ho, hum’. Hence, physicalism is false” (Jackson 1986, p. 291) [original emphasis].



---

# An Exercise or Two

# Two questions

(1) Is there a general recipe for constructing thought experiments? What ingredients would such a recipe have?

(2) Can you come up with your own thought experiment to disprove a certain hypothesis?

# Reading

- Brown, J.R. and Y. Fehige (2014) 'Thought Experiments', *The Stanford Encyclopedia of Philosophy* (Fall 2014 Edition), Edward N. Zalta (ed.), URL =<http://plato.stanford.edu/archives/fall2014/entries/thought-experiment/>.



---

The End