Philosophy and AI
Lecture 2: Epistemology and Philosophy of Science

Marco Degano

5 February 2025

Readings

Required:

- Glymour 2015: chapter 7, pp. 159–160; 168–179:
  - Introduction, Ancient Inductive Skepticism; Hume's Inductive Skepticism; Metaphysical Skepticism; Conclusion.

Optional:

- Ichikawa, Jonathan Jenkins and Matthias Steup, "The Analysis of Knowledge", The Stanford Encyclopedia of Philosophy `https://plato.stanford.edu/archives/fall2024/entries/knowledge-analysis`.
- Peter Adamson on Pyrrho and skepticism `https://www.historyofphilosophy.net/pyrrho`

Outline

Outline

What is knowledge?

- **Epistemology's Core Question:** What is knowledge, and how can we acquire it?
- **Practical Importance:** Knowing what constitutes knowledge helps us:
  - Differentiate reliable beliefs from guesses or misconceptions.
  - Make informed decisions based on justified beliefs.
- **Forms of Knowledge:**
  - *Knowledge-how:* Knowing how to perform actions (e.g., ride a bike).
  - *Knowledge-why:* Understanding explanations (e.g., why antibiotics work).
  - *Knowledge-that:* Propositional knowledge (e.g., knowing that the speed of light is approximately 300,000 km/s).
- **Focus:** Propositional knowledge (*knowledge-that*) is central because it serves as a foundation for other forms of knowledge.

Justified True Belief (JTB) Analysis

- According to the JTB analysis, knowledge is *justified true belief.*
- **Formalization:** An agent a knows that p iff (if and only if):
  - p is true. (Truth condition)
  - a believes that p. (Belief condition)
  - a is justified in believing p. (Justification condition)
- **Origins:** While often associated with Plato (episteme/knowledge vs doxa/belief), this explicit formulation emerged in the 20th century.
- **Motivation:** The goal is to clarify 'knowledge' by analyzing it into simpler, intuitive concepts: truth, belief, and justification.
- **Challenges:** These components themselves are not effectively specified, but they are assumed to be more accessible than the concept of knowledge.

Gettier Examples

- **Challenge to JTB:** Edmund Gettier (1963) introduced cases demonstrating that JTB conditions can be met without constituting knowledge.
- **Example (Russell's Stopped Clock):**

  If you look at a clock which you believe to be working, but it has actually stopped, and you happen to look at it at the correct time, you will form a true and justified belief about the time. However, this does not constitute knowledge because it is mere coincidence.

- **Key Insight:** Justification can be undermined by reliance on faulty evidence or processes, even when the belief turns out to be true.
- **Impact:** Gettier cases show that the "justified true belief" analysis is insufficient to fully capture what we mean by knowledge.

Responses I

Recall the formalization of the JTB analysis: knowledge iff justified true belief.

- **Accepted:** The $\Rightarrow$ ("if") direction of JTB is widely accepted: knowledge must involve truth, belief, and justification.
- **Contested:** The $\Leftarrow$ ("only if") direction is problematic: justified true belief does not always result in knowledge.
- **Focus of Debate:** Identifying the additional condition(s) needed to account for knowledge.

Responses II

Question: Can the right side of JTB be strengthened (while still being meaningful and achievable) to ensure knowledge?

- **No:** skepticism. Skeptics argue that robust knowledge is unattainable or extremely limited.
- **Yes:** solutions to skepticism. There are meaningful understandings of knowledge and justification on which we can achieve knowledge.

Outline

Pyrrhonian Skepticism

- The school of Pyrrho of Elis (ca. 360–275 BC) developed many skeptical doubts, known through the writings of Sextus Empiricus (ca. 200 AD).
- They formulated what is now known as the Münchhausen (or Agrippan) trilemma, along with other significant skeptical challenges such as the problem of induction (which will be discussed later).
- **Münchhausen Trilemma:** Named after Baron Münchhausen, a fictional character who attempted to pull himself out of a swamp by his own hair. Agrippa, an ancient philosopher of the Pyrrhonian school, is credited with identifying this trilemma in epistemology.

Baron Münchhausen

The Münchhausen Trilemma

- To justify or prove something in order to claim knowledge, we face three options:
  1. **Dogmatic Argument:** We stop at some foundational claim and accept it without further justification, which is arbitrary.
  2. **Circular Argument:** A claim is justified by referring back to itself.
  3. **Regressive Argument:** Every justification needs another justification, leading to an infinite regress.
- **Key Claim:** None of these options yields a convincing justification. Therefore, it is argued that we can never achieve a justification strong enough to yield knowledge.

Modern Perspectives on the Trilemma

- **Mathematics and Justification:**
  - In modern mathematics, the situation is less dire than the trilemma suggests.
  - **Axiomatic Method:** Identify simple axioms (self-evident statements accepted without proof) and derive theorems logically from them. This approach embraces the dogmatic horn of the trilemma.
- **Regress in Practice:**
  - Not all regressive arguments are problematic. If we know we can always extend justification when needed, this might suffice for practical purposes.

Mathematical Induction and Skepticism

- **Accepted Mathematical Induction:** For a property P of natural numbers, if:
  (a) P holds for 0 **[P(0)]**, and
  (b) Whenever P holds for n, it also holds for $n+1$ **[**$\forall n : P(n) \implies P(n+1)$**]**,
  then P holds for all natural numbers **[**$\forall n : P(n)$**]**.
- **A Regressive Argument:** Induction involves a regress argument in principle:
  - Justify $P(0)$ using (a).
  - Justify $P(1)$ using (b) and $P(0)$.
  - Justify $P(2)$ using (b) and $P(1)$, and so on.
- **Skeptical Challenge:** Skeptics might argue that even assumptions (a) and (b) require justification, leading to further doubts.

Interlude: The life of a skeptic

Why and how to adopt a skeptic stance?

1. **Investigation (Zetesis)**
   - Search reveals conflicting views
   - Leads to philosophical aporia (state of uncertainty)
2. **Suspension (Epochē)**
   - Withholding judgment
   - Release from dogmatic anxiety
3. **Tranquility (Ataraxia)**
   - Mental peace follows naturally
   - Freedom from disturbance

But:

- If skeptics claim we can know nothing, how can they know this?
- How to act without justified beliefs?

(You will not be tested on the Greek terminology, of course)

Outline

What is Metaphysical Skepticism?

- **Pyrrhonian skepticism**: Doubts any knowledge because we can never find reasonable justification.
- **Metaphysical skepticism**: Doubts any knowledge because the world may always differ from what we believe.
- This skepticism concerns the structure of the world (hence the term **metaphysical**).
- A key example: Descartes' Evil Demon thought experiment.

The Evil Demon Thought Experiment

Imagine a **malevolent omnipotent demon** trying to deceive you.

- Every experience you have is an illusion created by the demon.
- You believe you have hands, that other people exist, and that you can move, but none of it is true.
- 20th century analogy: **brain-in-a-vat** or the movie **Matrix**.
- Modern AI analogy: **simulation hypothesis**.

The General Metaphysical Skeptical Argument

- To justify my belief in p, I need a reliable **procedure** that can determine, based solely on experience, that p is true.
- But there are **two logically possible worlds** where we have exactly the same experiences (in this respect these worlds are indistinguishable), and yet p is true in only one of them.
- Any experiential procedure will have to choose: either p is true in both worlds, or p is false in both worlds. This procedure is therefore **unreliable**.
- **Conclusion**: My belief in p is not justified, and thus I do not know that p.

Examples of Metaphysical Skepticism (I)

The argument is general and applies to many statements p by identifying a corresponding possible world w.

The essence is: p is **underdetermined** by experience.

- p: The world as I experience it exists.
  - w: Descartes' demon world.
  - $\Rightarrow$: Skepticism about the world (idealism).
- p: Objects and people exist even when I am not observing them.
  - w: A world where everything is created whenever I look at it.
  - $\Rightarrow$: Skepticism about the continuous existence of things (Berkeleyan idealism).

Examples of Metaphysical Skepticism (II)

- p: Some events in the past actually happened.
  - w: A world where only the present exists.
  - $\Rightarrow$: Skepticism about the past (presentism).
- p: Other people have consciousness.
  - w: A world where all other people are zombies.
  - $\Rightarrow$: Skepticism about other people (solipsism).

Note: The skeptic does not claim p is false, only that we do not have knowledge of p.

Bostrom's Simulation Argument

- **Key claim**: Consciousness is like a computer program and can be simulated. (this raises a plethora of issues, but we take this for granted now)
- If computational power advances enough, we could create simulations of consciousness.
- Future technology will likely allow us to simulate entire conscious beings.
- Once we can do this, many simulations will be created, making simulated beings far more numerous than real ones.
- **Conclusion**: It is likely that we are in a simulation.

Critiques of the Simulation Argument

What are possible critiques of the simulation argument?

- **Consciousness is like a program**:
  - This is uncertain. What if consciousness depends on biology or other unknown factors?
- **Computational power makes simulation possible**:
  - What if consciousness requires non-computable functions?
- **Simulating consciousness in a universe**:
  - Creating a simulated consciousness in a realistic universe is much harder than simulating a mind alone.
- **Multiple simulations**:
  - Many simulations may be flawed or "buggy" unlike our consistent world.
  - Perfect simulations could create nested simulations, but this vastly increases computational demands, making it unlikely.

Outline

What is Inductive Skepticism?

- Inductive skepticism questions our ability to justify general conclusions from specific observations.
- Example: Observing 500 sunrises in the East-how can we conclude the sun will *always* rise in the East?
- Focuses on empirical laws we rely on, such as: *Water boils at 100°C under standard pressure.*
- Unlike metaphysical skepticism, inductive skepticism is practical: it challenges science and prediction in everyday life.

Induction

- **David Hume** popularized the problem of induction, though it appears in Pyrrhonian skepticism and Indian philosophy.
- **The problem of induction**:
  - Induction $\neq$ deduction: It generalizes from experience, unlike reasoning logically from premises.
  - Example: Observing repeated patterns (e.g., fire producing heat) and assuming they will continue.

Three Key Distinctions

- Deduction: Certainty through **logical necessity**.
  Example: Since all triangles have three sides and this shape is a triangle, this shape must have three sides.
- Induction: **Generalizations** based on **experience**.
  Example: Every time I've dropped a ball, it has fallen to the ground, so I conclude that all balls fall when dropped.
- Abduction: Inferring **causes from observed effects**.
  Example: The ground is wet, so it probably rained recently.

The Problem of Induction

- **Historical shift in science:**
  - Pre-17th century: Deductive systems like Euclid's axiomatic method dominated.
  - Post-17th century: Observation-based generalization became central (e.g., Newton's laws).
- **Core issue:**
  - Inductive inferences are unreliable:
    - No deductive/demonstrative justification: past regularities do not guarantee future behavior.
    - No inductive/probable justification: justifying induction by induction is circular reasoning.

Hume's Challenge to Inductive Justification

No deductive/demonstrative justification:

- Deductive reasoning requires that the conclusion follows necessarily from the premises.
- Observed regularities (e.g., the sun rising in the East) stem from our repeated experiences.
- Our minds form associations through habit or custom, not through a demonstration of logical necessity.
- Hence, no deductive argument can guarantee that these patterns will continue in the future.

Hume's Challenge to Inductive Justification

No inductive/probable justification:

- Inductive reasoning presupposes that the future will resemble the past (the uniformity of nature).
- However, using past experiences to justify this assumption is circular:
- Example: The "old future" behaved like the past, but this does not logically ensure that the "new future" will follow the same pattern.
- Induction presupposes the very principle it seeks to justify: that patterns observed in the past will hold in the future.

Conclusion: Induction lacks a non-circular foundation.

Implications for Knowledge

- If induction cannot provide certainty:
  - Scientific "knowledge" becomes probabilistic, not absolute.
  - Predictions based on induction are inherently uncertain.
- Modern applications:
  - Statistics, machine learning, and AI rely on patterns from past data, acknowledging uncertainty.

Conclusion

- Both inductive and metaphysical skepticism suggest we lack a reliable method for acquiring knowledge about the world.
- Two possible solutions:
  1. **Deny the skeptical description** (e.g., directly access truths, Plato).
  2. **Lower the standard of knowledge:**
     - Reliability need only apply to worlds we care about, not all logically possible worlds.
- In the next lecture, we will explore responses to skepticism in line with point 2 focusing on **probability theory** and **Bayesianism**.

What's Next?

- Bayesianism provides a framework for handling uncertainty:
  - By expressing uncertainty through probability distributions.
  - By deriving reliable methods for updating beliefs based on evidence.
- These methods influence contemporary fields like economics, statistics, and AI.

Exercises

1. Can you find more 'Gettier type' counterexamples to the JTB analysis of knowledge?
2. What is the difference between Phyrronian and metaphysical skepticism?
3. Do you find the indistinguishable experience argument scheme convincing? Reflect on the philosophical methodology of inferring something about the concept of knowledge in our world from the possible worlds considered in the various skeptical thought experiments.