Introduction to Philosophy

Analytic Philosophy

Western traditon: beginning in Europe; Late 19th century–present

Ideals: clarity of thought & objectivity of results

Strategies: precision of expression & rigour of argument

Method: analysis >> analysis of language/the world

Tool: Formal logic

Reducing mathematics to logic:

Russell; Moore; Frege; Wittgenstein; Kripke; Timothy Williamson

Philosophical Paradoxes

Paradox= *para* (against/beyond) + *doxa* (opinion/belief)

VS contradiction: paradox requires belief

Interesting: against common sense, belief deeply held(=intuitively true)

Examples of Philosophical Paradox

- **Paradox of Analysis**

1) For any 'A=B', either 'A' and 'B' have the same meaning or not.
2) So, 'A=B' is either trivial or incorrect.

- **Moore's Paradox**

"It is possible that it is raining and I do not believe it is raining."

↑despite its apparent contradiction, the statement seems available to be understood

>> the nature of 'self'?

- **Russell's Paradox**

1) For every predicate, there is a set which consists of all and only those objects that satisfy the predicate. (Naive Comprehension)
2) No set "is a set which contains all and only those sets that do not contain themselves."
3) So, Naive Comprehension is false.

- **Barber Paradox**

"There can exist no barber who shaves all and only those who do not shave themselves."

- **Liar Paradox**

"This statement is a lie."

- **Paradox of Omnipotence**

"Can God create a stone too heavy for him to lift?"

- **Paradox of Inference**

1) "Given A and B, then Z. If A and B, why Z?"
2) "Because of C (if A and B, then Z)."
3) "If A, B and C, why Z?"
4) "Because of D (if A, B and C, then Z)."
5) …infinitum

- **The Sorites Paradox (AKA the Heap, the Bald Man)**

Problems of anti-epistemicists: lack of motivation to be precise and discover exact truths

The nature of Paradox

1. Quine's Paradox

Definition: "any conclusion that at first sounds absurd but has an argument to sustain it."

Lycan's criticism:

# Criticism 1: "False identification of a paradox on the basis of one's own (subjective) preference for a way of resolving it."

>> it's not a conclusion, Too SUBJECTIVE!

# Criticism 2: "False identification of a paradox on the basis of one's own (subjective) epistemic attitude, specifically on the basis of how plausible one finds individual propositions."

>> emphasis on the epistemic attitude on the final conclusion

# Substitutive Proposal: "inconsistent set of individually plausible propositions"

>> inconsistency=right+right+…+right→absurd

Possible defence:

1) Correct resolution of a given paradox also reveals its true nature.
2) Heliocentrism v.s geocentrism, before and after Copernican Revolution

Philosophical vs non-philosophical paradoxes

Non-Philosophical Paradoxes

- **Scientific paradoxes**

Black Hole Information Paradox, Easterlin Paradox, Einstein–Podolsky–Rosen Paradox, French Paradox, Helianthus Paradoxus (paradoxical sunflower), Hispanic Paradox, Lek Paradox, Olbers' Paradox, Paradox of the Plankton, Paradoxical Embolism, Pseudis Paradoxa (paradoxical frog), Pulsus Paradoxus (paradoxical pulse), Twin Paradox

- **Children's paradoxes**

Philosophical paradoxes

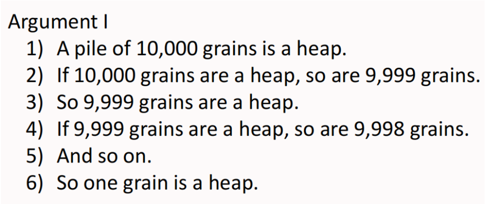
1) Object vs subject (compare psychology)

Interested in cases that tend to be found paradoxical by even the most intelligent subjects

2) Generality

Only require minimal knowledge to be found paradoxical (i.e. max. intuitive)

The Sorites Paradox



Argument I

1) A pile of 10,000 grains is a heap.
2) If 10,000 grains are a heap, so are 9,999 grains.
3) So 9,999 grains are a heap.
4) If 9,999 grains are a heap, so are 9,998 grains.
5) And so on.
6) So one grain is a heap.

Problem: no sharp cut-off point

Proposed Solutions:

1) Epistemicism: "there exist sharp cut-off points; vagueness is ignorance"

2) Degrees of truth: "truth comes in degrees"

>> problems: hard to quantify; conjunctive statements[1]

3) Supervaluationism: "vagueness requires higher-order truth values."

---

[1] The conjunctive statement is given the same value as that of its lowest conjunct. For example, if 'Amanda is tall' is assigned 0.8 and 'Sam is tall' is assigned 0.95, then the conjunction 'Amanda is tall and Sam is tall' gets the value 0.8. This may seem reasonable enough. But if Amanda is borderline tall so that 'Amanda is tall' is assigned 0.5, the necessary falsehood 'Amanda is tall and not tall' gets the value 0.5.

>> but still cannot answer the boundary between Supertrue, Borderline cases and Superfalse

Science and Philosophy

I fully agree with you about the significance and educational value of methodology as well as history and philosophy of science. So many people today and even professional scientists seem to me like somebody who has seen thousands of trees but hasnever seen a forest. A knowledge of the historic and philosophical background gives that kind of independence from prejudices ofhis generation from which most scientists are suffering. This independence created by philosophical insight is in my opinion the mark of distinction between a mere artisan or specialist and a real seeker after truth. (Einstein)

Models to answer what is philosophy

Socrates: Philosophy is a science of the human (≈psychology)

↑=Naturalism VS platonism/non-naturalism

Overview: Williamson 2022, "What Philosophy Is"

Central claims

"Philosophy is a non-natural science."

"seamlessly...fits into the intellectual landscape of total science"

"Science and philosophy are products of the same natural pursuit of knowledge."

"Those who offer the asking-without-answering model of philosophy have not thought hard enough about what it is to ask a question"

"Philosophy has made much more progress than the [critic] realises."

"Much scientific [and philos.] progress consists in the development of better models."

"The word 'philosophy', like 'mathematics', has no generally accepted definition."

"Both...to draw a neat line around philosophy and the denial that [it] is part ofscience stand out as futile and ill-motivated."

>> Q: Why is philosophy not a natural science (too)?

Argument 1

"Those who offer the asking-without-answering model of philosophy have not thought hard enough about what it is to ask a question."

P1 Normally, when asking questions, we want answers.
P2 If we have no chance of answering phil. questions, asking them must be idle.
C The value of philosophy does not consist in asking questions.

Argument 2

"Philosophy has made much more progress than the [critic] realises."

P1.1 Logic is a sub-discipline of philosophy.
P1.2 The sub-discipline of logic has made significant progress.
C1 Philosophy has made significant progress.
P2.1 The critic does not recognise the development of better models as progress.
P2.2 The development of better models is progress.
P2.3 Many areas of philosophy have developed better models to a significant degree.
C2 Philosophy has made significant progress.
P3.1 The critic does not realise how much philosophy overlaps other disciplines.
P3.2 There has been significant progress in these overlap areas.
P3 Philosophy has made significant progress.

AI Ethics

Artificial Intelligence could spell the end of human race. ——Steven Hawking

What is AI (artificial intelligence)?

The science and engineering of making intelligent machines!(McCarthy 1955)

For our purpose, we can broadly understand Al as getting machines to do things that require cognitive functions such as thinking learning, and problem-solving when done in intelligent beings such as humans.(Liao 2020)

Philosophical issues include:

- What, if anything, are intelligent machines?
- What is intelligence?
- Can machines have subjectivity, consciousness, etc.?
- Etc. etc.

The Classification of Philosophy

Practical: Ethics & Aesthetics

Theoretical: Metaphysics & Epistemology (Kant-grounding discipline)

Metaphysics = "what is"

Epistemology = "what is knowledge"

History: About the history of all aforementioned subjects

The Classification of Ethics

Meta-ethics: what is morality?

Metaphysics: What is the nature of morality? Are there moral facts?

Epistemology: Is there moral knowledge? If so, do we have any? If so, how?

Normative ethics: what is moral (morally good/bad)?

Virtue ethics (Aristotle)

Duty theories/deontology (Kant)

Consequentialism (e.g, utilitarianism)

Non-consequentialism (e.g, Moore's intuitionism)

Applied/Practical Ethics: is that moral (morally good/bad)?

The critical philosophy of race

AI ethics

- The study of morality
- A subfield of philosophy

So, Al ethics is the study of morality with regard to Al.

Central Issues in AI Ethics

Main issues in reading (Liao 2020, 1-13)

- Machine learning is data hungry >> where's the boundary?
- Garbage in / garbage out
- "Faulty algorithms"
- "Deep learning is a black box"
- "Machine learning is weak AI" = lack of self-awareness vague elements

"In addition to being narrow Al, current machine learning systems are also weak AI in that they do not have self-awareness or consciousness and they cannot think for themselves.” (p.9)

Some other issues

Manipulation of behaviour, surveillance, human-robot interaction, moral status (patients, agents), automation and employment, military applications, virtual reality, singularity, existential risk

“Child” Case

What should a self-driving car do in this case? Should the passenger have absolute priority? If not, how else might we solve the problem? Notice vague elements (child, low chance)!

Can we wait for a theoretical solution? Do we have to wait with philosophy?

No, philosophy can be even more practically oriented.

> **Child:** A runaway self-driving car is headed toward a child who will be killed. The self-driving car can swerve slightly to avoid hitting the child. Swerving the car slightly to avoid hitting the child has a low (but not zero) chance of harming the passenger in the car.

In general, what is to be done now?

Collectively: “Philosophy with a deadline?”(Bostrom, *Superintelligence*, 2014)

practically oriented

Individually: “Affirmation of life” (Nietzsche, *A Critique of Morality*)

(cultural) relativism

“God is dead”: death of religious standard of right and wrong >>hollowness

Nietzsche v.s Socrates

Plato’s *Euthyphro*:

- **The Theoretical Part**

-What is piety (‘he pious’)?

-(def.) Necessary and sufficient conditions

• Euthyphro’s answers	• Socrates’ objections
1) The pious = what is dear to the gods	Against (1): The same things are dear and not-dear to the gods.
2) The pious = what is dear to all gods	Against (2): The pious and the god-loved cannot be equivalent.
3) The pious = the part of the just that is concerned with the care of the gods	Against (3): This reduces to (1) or (2).

- **The Practical Part**

What is to be done?

1 The case of Euthyphro

• Traditional reading

Euthyphro begs off on excuse.

• Alternative reading

Euthyphro may be hurrying to retract the charges he brought against his father.

2 The case of Socrates

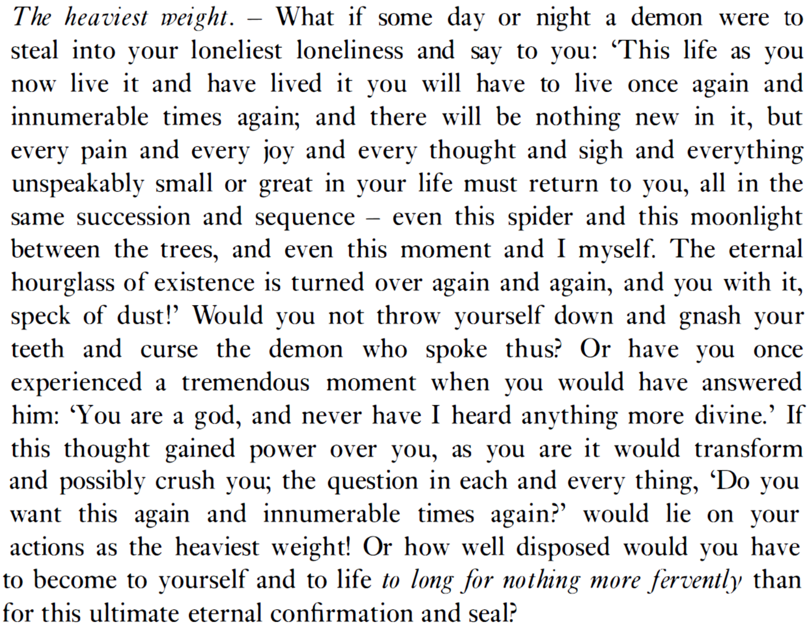
• “Philosophy with a deadline.”

• The death of Socrates

•Should Socrates’ve accepted the sentencing?

Nietzsche

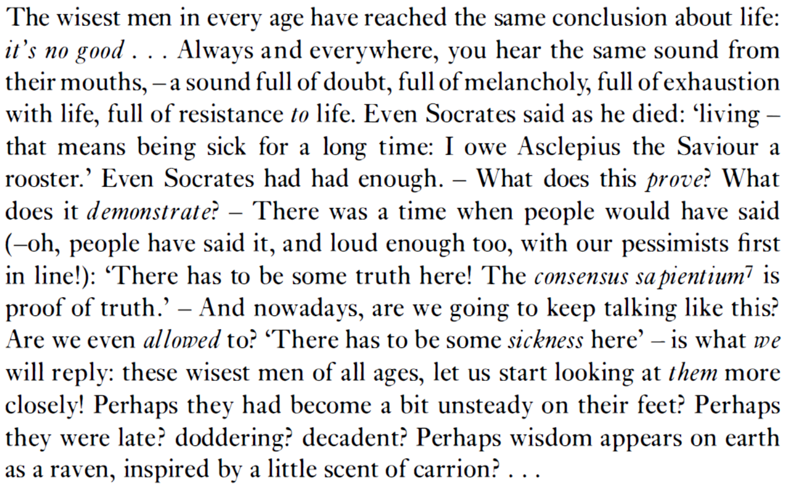
The heaviest weight (*The Gay Science* §341)



The heaviest weight. – What if some day or night a demon were to steal into your loneliest loneliness and say to you: 'This life as you now live it and have lived it you will have to live once again and innumerable times again; and there will be nothing new in it, but every pain and every joy and every thought and sigh and everything unspeakably small or great in your life must return to you, all in the same succession and sequence – even this spider and this moonlight between the trees, and even this moment and I myself. The eternal hourglass of existence is turned over again and again, and you with it, speck of dust!' Would you not throw yourself down and gnash your teeth and curse the demon who spoke thus? Or have you once experienced a tremendous moment when you would have answered him: 'You are a god, and never have I heard anything more divine.' If this thought gained power over you, as you are it would transform and possibly crush you; the question in each and every thing, 'Do you want this again and innumerable times again?' would lie on your actions as the heaviest weight! Or how well disposed would you have to become to yourself and to life *to long for nothing more fervently* than for this ultimate eternal confirmation and seal?

Twilight of the Idols (1899), The Problem of Socrated, §1:

Life is a sickness >> good reasons to accept the sentence



The wisest men in every age have reached the same conclusion about life: *it's no good* . . . Always and everywhere, you hear the same sound from their mouths, – a sound full of doubt, full of melancholy, full of exhaustion with life, full of resistance *to* life. Even Socrates said as he died: 'living – that means being sick for a long time: I owe Asclepius the Saviour a rooster.' Even Socrates had had enough. – What does this *prove*? What does it *demonstrate*? – There was a time when people would have said (–oh, people have said it, and loud enough too, with our pessimists first in line!): 'There has to be some truth here! The *consensus sapientium*[7] is proof of truth.' – And nowadays, are we going to keep talking like this? Are we even *allowed* to? 'There has to be some *sickness* here' – is what *we* will reply: these wisest men of all ages, let us start looking at *them* more closely! Perhaps they had become a bit unsteady on their feet? Perhaps they were late? doddering? decadent? Perhaps wisdom appears on earth as a raven, inspired by a little scent of carrion? . . .

The Socratic Method

a) Honest interest in interlocutor's concerns
b) Takes interlocutor seriously as a person
c) Exclusive focus on i's statements, beliefs -- taking the 'if' out (*Protagoras* 331c)
d) Avoid expressing his own opinions
e) Constantly ask i. whether i. can follow
f) Eager to clarify his meaning, if necessary
g) Constantly asks for i.'s meaning -- "if I understand you correctly"
h) Aim to uncover incoherent presumptions

Nietzsche and Socrates

At the beginning, Socrates taught us to ask philosophical questions.
2,000 years later, Nietzsche started to ask philosophical questions about these.
    Are traditional phil. questions healthy questions?
And Nietzsche started to ask new philosophical questions.
    Are reason, truth, knowledge, science absolutely good?
    What is affirmation of life?
    What is an authentic life?
    # Healthy = will to power = being able to do things as we want >> keeping alive
In this way, Nietzsche has had a strong influence.
    On Heidegger, Arendt, De Beauvoir, Foucault, etc.
    # "Philosophy is not a science." >>"Being trapped in rational beliefs rather than healthy, irrational ones is what killed Socrates in the end."
    But not on so many thinkers in the analytic tradition

Descartes

Comments

Descartes: skepticism

It's a mistake perceptions can prove us awake, because we can feel the same in dreams.

David Hume: don't believe in Christianity & skepticism

Descartes's doubts are reasonable. But the problems he brought up are basically unsolvable and pondering over them can bring no knowledge.

Immaneul Kant: critical; disagreement with reliance on subjective intuition

"No matter how innocent idealism may be held to be as regards the essential ends of metaphysics (though in fact it is not so innocent), it always remains a scandal of philosophy and universal human reason that **the existence of things outside us** (from which we after all get the whole matter for our cognitions, even for our inner sense) **should have to be assumned merely on faith,** and that if it occurs to anyone to doubt it, we should be unable to answer him with a satisfactory proof."

Hilary Putman: Brain in a vat, but failed to answer

(*Reason, Truth, and History*) introducing↑as the modern version of Descartes' *demon*
>> a stronger skeptic scenario: artificial/virtual reality

G.E.Moore: "Proof of an External World"&"Certainty"

"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."

Sceptical argument

(1) If I know that p, then I know that sceptical scenario s regarding p is not the case.
(2) I do not know that sceptical scenario s regarding p is not the case.
(3) Hence, I do not know that p.

Anti-sceptical argument

(1) If I know that p, then I know that sceptical scenario s regarding p is not the case.

(4) I know that p.

(5) Hence, I know that sceptical scenario s regarding p is not the case.

E.g. Moore: (1') If I know that here is a hand, then I know that I am not dreaming.
(4') I know that here is a hand.
(5') Hence, I know that I am not dreaming.

"One final point should be made clear. It is certainly logically possible that I should have been dreaming now; I might have been dreaming now; and therefore the proposition that I am dreaming now is not self-contradictory. But what I am in doubt of is whether it is logically possible that I should both be having all the sensory experiences and the memories that I have and yet be dreaming. The conjunction of the proposition that I have these sense experiences and memories with the proposition that I am dreaming does seem to me to be very likely self-contradictory."

Neo-Moorean Responses to Scepticism

1) It is not self-contradictory that all of our sense experiences and memories should be part of
one big computer simulation (i.e., the brain-in-a-vat hypothesis).

2) Davod Chalmers: VR is no less real than the reality (*Reality*+)

3) The discovery of brain-in-a-vat hypothesis would just be a (possibly scary) discovery about
the fundamental structure of this world; but with it, my knowledge does not shrink, but increase.

VS the evil-demon hypothesis: the demon is said to be constantly deceiving

Naming and Necessity

Russell:

Proper Names=words for particulars

The names that we commonly use, like "Socrates," are really abbreviations for descriptions.

The only words one does use as names in the logical sense are words like "this" or "that."

Referentialism: the meaning of a word is its referent

Naming is only reference.

Paradox of Analysis

1) For any 'A = B', either 'A' and 'B' have the same meaning or not.
2) So, 'A = B' is either trivial or incorrect.

Frege:

Proper names=signs for objects

An object is anything that is not a function,
so that an expression for it does not contain any empty place.

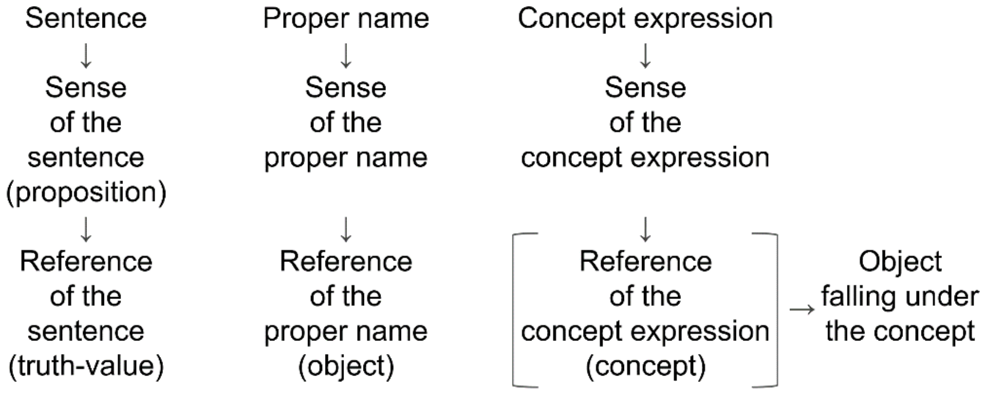
Solution of Paradox of Analysis:

Naming is both *sinn* (sense=mode of presentation) and *bedeutung* (meaning/reference).

Morning star = evening star; 5 = 2 + 3; Water = H_2O

Debates on concept words

Frege: With a concept word it takes one more step to reach the object than with a proper name, and the last step may be missing—i.e., the concept may be empty—without the concept word's ceasing to be scientifically useful.


Sentence
Sense of the sentence (proposition)
Reference of the sentence (truth-value)
Proper name
Sense of the proper name
Reference of the proper name (object)
Concept expression
Sense of the concept expression
Reference of the concept expression (concept)
Object falling under the concept

Russell: The word "common name" leads to the mistaken assumption that a common name is related to objects in essentially the same way as is a proper name … that is why I prefer "concept word" to "common name".

Kripke:

Name spreads from place to place.

"Someone, let's say, a baby, is born; his parents call him by a certain name … Through various sorts of talk the name is spread from link to link as if by a chain. A speaker who is on the far end of this chain … may be referring to Richard Feynman … A chain of communication going back to Feynman himself has been established, by virtue of his [the speaker's] membership in a community which passed the name on from link to link."

>> the link rather than the description secures the naming (so possibilities are simply irrelevant)

Proper names are rigid designators.

A rigid designator is anything that will designate the same object in every possible world where it exists (and nothing else in any possible world)

All proper names;

Some common names ("water", "H_2O", "gold", etc.);

Some definite descriptions ("the element with atomic number 79", "the successor of 2", etc.)

So, what happens to common names?

Kripke also argues that some other words are rigid designators, e.g. 'water': If this is correct, science can discover necessary truths, e.g. 'water = H_2O'; then science and metaphysics will have significant overlap.

"If we imagine a hypothetical (admittedly somewhat artificial) baptism of the substance [gold], we must imagine it picked out as by some such "definition" as, "Gold is the substance instantiated by the items over there, or at any rate, by almost all of them"… The "almost all" qualification allows that some fool's gold may be present in the sample. If the original sample has a small number of deviant items, they will be rejected as not really gold. If, on the other hand, the supposition that there is one uniform substance or kind in the initial sample proves more radically in error, reactions can vary: sometimes we may declare that there are two kinds of gold, sometimes we may drop the term “gold”. (These possibilities are not supposed to be exhaustive.)”

Reflection
According to Kripke, the word “gold” could have developed differently (e.g., to designate “gold- or-pyrite-or-chalcopyrite-or-mica-or-…)
Kripke writes: “Scientific investigation generally discovers characteristics of gold which are far better than the original set” (1980, p. 138)
However, alternative histories show that more is required, esp. scientific society.
Q: Are there any words that cannot be rigid designators? What about words such as ‘love’, ‘knowledge’, ‘piety’, ‘justice’, ‘intelligence’, or ‘democracy’?

2024春季学期的Intro, Julius提供了极为详尽清晰的Notes

（那就完全没必要看我以下半懂不懂的笔记啦……）

Physicalism and reduction

The Mind’s Place in Nature (the physical world)

Consciousness: the vivd experiential character of experience
>> no colour in brain?
Intentionality: how can brain states be about the world (*aboutness*)
Mind v.s Physical World: how to explain non-physical entities (e.g numbers)
>> the mind being concrete(physical) & normative(can be correct/incorrect)
How can a totally physical thing be right or wrong?
>> consciousness in undeniable
Physicalism: There’s nothing over and above physical entities.
Supervenience physicalism
One-side dependent relation: objects’ changes ***rely on*** changes in detailed, physical, basic layer
(but it doesn’t make sense other way round)
How to understand a duplicate of someone? Is it the same person or only a physical copy of him?
Physicalism: the same, physically identical → totally identical
Explanation issue: bottom-up=mental-physical(idealism)/physical-mental
reductivism 立场：
mental states 等同于 physical properties
Non-reductivism+supervenience:
Mental states 依赖于 physical properties
Explaining supervenience: Identity
If A is identical to B, that B exists will be sufficient to explain the existence of A.
Semantic-identification: mental terms have a meaning identicall to physical terms
Referring to the same person with different terms → identity
Token identification(event):
The same event has numerous properties and physicalists avoid the possibilities that there are purely mental properties.
>> E1: sitting there(physical) = attaining enlightenment(non-physical):E2
Type(/Kind) Identification:
在 type 意义上一一对应（比如红色对应某大脑状态）
Theory reduction: if all psychological states can be explained by neuron firings
>> in this way, psychology can be reduced to neurobiology
Type(red; computers) >> specific examples
what science is about - general truth >> how ↑ is expected to behave
相对于 Type 指大类对应，token 指一一对应；无关 reductivism/non-reductivism
但 token 的描述/称呼往往不得不带着 type
Belief: thinking something to be true <knowledge as a kind of belief

Wetware	v.s	hardware
Neurons, brainy		silicon

Beliefs = wetware/ wetware or hardware?

The Problem of Projectability

为什么可以从个例推广出全体?

推广的边界是哪里:个体——所有物体,停在哪里

1) 不能诉诸 common sense

2) Casual power -- 因果是闭合的→一切都有因果

>> functionalism

Micro structure (basic) --但不能满足 projectability 的要求,不能界定“类”

Observation of cases (F=G) can strengthen our credence in the next F being G.

>> as certainty climbs up, a law is formed

MR rejects the possibility of unified behaviours, i.e laws

Why is MR important?

At first, it was Type Identification

>> because of MR, you cannot reduce beliefs into one type of thing

→Integrity of psychology and mental entites

\>>Kim: once you have mutiply realizable mental kinds, they have no lawlike behaviour anymore, so they are not entities in any recognizable senses

→behaving in predictable ways is the premiss of unified rules, thus a discipline

Q&A

MR: special sciences seem not to reduce to physics

No exhausive correspondence between minds

MR does lead to irreducibility at the cost of losing unified beliefs

“天鹅都是白色的”与“不是白色就不是天鹅”的命题是等价的,看到一个蓝色水杯会增加后者的可信度,所以前者的可信度也增加了?

根本问题:什么是 nomic kinds?

理想情况是 micro structure 带来的 causal powers,但事实上如何摆脱日常语言与社会观念?

Being Red 15(in particular) explains its being Red, but they are not identical.

>> the explanatory relation: the hue of “red 15” realizes its being red

And there being other shades of RED (other than red 15) prevents them being identical.

Consciousness

Access Consciousness: a state's being conscious is a matter of its availability to interact with other states and of the access that one has to its content
Creature Consciousness: being awake
Self-consciousness, self-awarenss: knowing about oneself, e.g recognizing oneself in the mirror
Phenomenal consciousness: what-it-is likeness

The Easy Problems of Consciousness: problems that seem directly susceptible to standard methods of cognitive science >> to be explained in terms of computational/neural mechanism

The Hard Problems of Consciousness: the problem of experience↔physical knowledge

The Knowledge Argument:
Traditionally speaking: no "allegedly new" knowledge → no surprise
BUT in Frankson's case, Mary should still be surprised, and that's the problem.

Why seeing other reds will make us less surprised when we see a new shade of red?
-- Thourgh Type Identification? If I see some part of a type, the other parts of it would not shock me anymore.
But how to explain the concept of "type" from a philosophical level?
How can we categorize things into a "type", from daily expeirence?

Another possible solution: old facts, new abilities
>> experiencing red is not a matter of knowing(knowing-that), but an ability(knowing-how)
*Experiencing something explains why we have certain abilities, e.g:

Our ability to see red make us able to recognize objects(e.g apples) through colour but it is that we can recognize those objects that proves that we can see red.

Chalmers (1996) _The Conscious Mind
Philosophical Zombies: "Someone or something physically and functionally identical to me (a conscious being), but lacking conscious experiences altogether."
This poses threat to physicalism

Why should a zombie believe that he's not a zombie? Isn't it a part of (self-)consciousness?
>> My zombie twin only lacks phenomenal consciousness (the hard problem part), so they process information just as we do, it's just they are not conscious of it. Beliefs and desires(which are responsible for actions) are results of information processing, not phenomenal consciousness.

> Chalmer's point is that there's no stage in acting where one is forced to think of consciousness.
>
> There's no critical point when "consciousness" arises in the development of cell as

well since there is only physical stuff running around.

——so that's why it would be hard to say computers *must not* have consciousness.

How can a zombie imagine there's a zombie duplicate of himself since he doesn't have consciousness to subtract? But belief is something decided by information processing so the zombie should believe they have consciousness as "I" do.

Dennett: Being conceivable doesn't mean being possible.
Chalmers: The Default Theory: It requires much evidence to say something is totally impossible.
>> in the zombies' case, it's conceivable and you cannot say it's impossible.

Brain realizes consciousness?
-- Why we're inclined to think "zombies"(duplicates) should have consciousness.
Object 的前提：Intersubjectivity

Intentionality

About-ness = Intentionality = Representation
Intentional Inexistence: my mental state being about something even though there's nothing in the world that can be about
Opacity: coreferential terms (terms referring to the same thing) cannot be substituted *salva veritate* (i.e, preserving truth of sentence) when embedded under intentional terms.
>> Intentional states are not purely object focused.
Correctness: there's rights and wrongs of intentional states, but there's no such things in the physical world. (and that's because not only the physical world provides the content of my mind)

Non-existent aspects: misrepresent a property of an object
Non-existent object: misrepresent an object

Goals of explanations: non-circular, correctness
>> underextension: capturing too few cases, theory
>> overextension: capturing too many cases, resemblance theory

Terms can have meaning without being about something existent.
The meaning of a term X is what it takes for anything to make me call it X.

Covariance:
Seeing the dog (S) means a certain bodily state (z) of me.
But being alive is also part of "z" but it is entirely unrelated to S.

Resemblance:
Definitions of Intentionality: R represents S iff X.
Intentionality as Resemblance: R represents S iff R resembles S.
  Symmetry: It doesn't work the other way.
  Specificity: too simple
  Representation proliferation: everything resembles everything in some sense.
  >> Illusion of consciousness:~~ resemblance~~

Causal Theory
Simple Causal Theory of Intentionality: R represents S iff S caused R.
>> if it doesn't go this way, it is incorrect.
Overextension/underextension: cannot explain why thoughts can be incorrect
Problem: expect = things happen as I thought →still intentional

The Main Problem is always the Misinterpretation.

Supervenience Principle: changes in higher layers require changes in more basic layers
Explanation:
1) Identity: The two being the same thing
>> opposite cases: (true but not identical)
there being a bottle can explain there being a bottle or a cup
there being a crimson patch on the blackboard can explain there being a red patch
# The "Correctness" Thing: There's no way of explaining "correctness" without the concept of "correctness" now → quite non-physical concept

Specifies laws are necessary for thinghood (being an entity) [Kim-MR doesn't have it]
Q: Are laws sufficient for thinghood?

Causal relation v.s contemporary explanation:
Contemporary means metaphysical explanation (meaning at the same time).

Mary: Mary learning a new fact = Conceiving zombies
>> Consciousness is a non-physical fact
Conceivable=non-contradictory
When would Mary be forced to write experiences of "red" in the list she knows?
=when would one be forced to know consciousness after various neural states given?

Boundary from Conceivability to Possibility:
Chalmer: all boundaries are arbitrary so conceivability=possibility
>> All physical properties cannot be determined by <u>any laws of nature</u>.
There being no determined-to-be-correct laws of nature?

Descartes' problem:
How could soul (something taking no space) communicate with body (something taking space)?

Beliefs and other propositional attitudes

What is the nature of (human) thought?
Intuition: always conducted in language
→ is thinking something like inner speech? Is thinking language-like at all?
Fodor's answer: There is an inner language of thought (not any human language and cannot be conveyed by speech). It is the same for all thinking creatures and it need not to be learned.

Propositional attitudes: hoping, believing, desiring, knowing
= all attitudes that take propositions as their content (indicated with a that clause)
→ Person +An attitude+A proposition, e.g Julius hopes that this lesson would be fun

Beliefs have structure: they relate us to a proposition.
Propositions are the things that can be **true and false,** the things that represent the world. [about]
>> Beliefs are true exactly when their propositions are true.
Thus, beliefs are composed of many parts.
Why is this structure important? It tells us how different beliefs are related.

Julius believes that John loves Marry and Julius believes that Beijing is in China are similar

in that they **both are beliefs of Julius**.

Julius believes that John loves Marry and Julius believes that John is tall are similar in that

they are **both beliefs about John**.

Compare Fodor's fusion story (542) (= beliefs have no structure)
>> but if belief does not have structure, you wouldn't be able to make inferences to acquire new beliefs (A loves B, C loves B→both A and C love B).

Propositional attitudes are language like

Logical form

Some arguments showing that beliefs are language like: Beliefs behave logically
Existential generalization: Julius believes John loves Mary entails Julius believes something.
Transmission (Aristoteles' condition): Beliefs figure in chains of reasoning such as this:

P1. John believes that it will rain if he washes his car.
P2. John wants it to rain.
C. John washes his car.

More abstractly

P1. John believes that if he does p then q
P2. John desire q.
C. Therefore, John does p.

Reasoning with beliefs works in a schematic manner.

Believing and saying

Things that I can believe can also be states.
Are there beliefs that cannot be said?

Propositional attitudes are not language like

Opacity

Beliefs are opaque: they don't survive substitution of co-referring terms.

"Superman can fly" is true if and only if "Clark Kent can fly" is true.

But it is not true that "John believes that superman can fly" is true if and only if "John

believes that Clark Kent can fly."

When put under propositional attitude verbs (e.g., "believes") words acquire a different meaning, i.e., their mode of presentation.

They don't suit the inference pattern (X is F, Any X that is F is G→X is G).

Carnap's answer:

Quotation relates you to a sentence, not to it's meaning.

Despite being related, it is not same to believe that superman can fly and to believe "superman can fly." Because the former indicates you believing a fact while the latter indicates believing in a proposition.

An example: 'Cesar said "veni, vidi, vici"'. You can understand this sentence even if you don't know what "veni, vidi, vici" means.

Beliefs are less finely individuated than sentences.

Syntactic variation: Again, "Mary bit John" and "John was bitten by Mary" can be the contents of the same beliefs, but they are different sentences.

Different languages: "John bit Mary" and "约翰咬了玛丽" express the contents of the same beliefs, yet they are different sentences.

Beliefs without language

Beliefs don't require language: Animals have beliefs/Little babies have beliefs.

Acquiring a language requires beliefs:

"If (token) sentences of a natural language are the objects of propositional attitudes, how are (first) languages learned? On any theory of language learning we can now imagine that process must involve the collection of data, the formulation of hypotheses, the checking of the hypotheses against the data, and the decision about which of the hypotheses the data best confirm. That is, it must involve such mental states and processes as beliefs, expectation and perceptual integration." (550)

The language of thought hypothesis (LOT)

Beliefs are relations to sentences in an innate, non-natural language (550): the language of thought. The language of thought is a system of "internal representations and propositional attitudes are relations that we bear to them." (553)

>> Animals and babies can think because they have a language of thought.

>> Different languages: Chinese and English speakers have many of the same beliefs because they have the same language of thought.

>> Opacity: Belief ascriptions are opaque because coreferential sentences can be

different sentences in the language of thought

>> Acquisition: The language of thought is used when acquiring a natural language.

Think of LOT as factory installed software. Humans come with intricate hardware; why not intricate software?

The language of thought hypothesis

– cons –

Observability: we cannot ever say what the language of thought is, we can only talk about it in spoken language. All we can say is that it is language-like.

Meaning: How do these innate concepts acquire meaning? Most of our concepts have meaning because we are connected, in some sense, to the world.

The belief relation

Consider a belief:

Julius believes that John bit Mary.

So far we have talked about what propositions are and how we relate to them: we encode them in the language of thought.

Let's talk about the "believes" part - What distinguishes "belief" from others?

Julius believes that John bit Mary.

Behaviorism

Carnap: propositional attitudes are "relations between people and sentences they are disposed to utter." (547)

Behaviorism. Propositional attitudes can be reduced to Input – Output laws.

(too easy)

What kind of truth: conceptual, empirical?

Consider this belief: Julius believes that Beijing is beautiful.

Am I disposed to utter this sentence? Not necessarily – I don't go around saying "Beijing is beautiful".

Am I disposed to utter this sentence when I am asked whether Beijing is beautiful? Not necessarily, maybe I want to hide the fact that Beijing is beautiful.

Functionalism

Am I disposed to utter this sentence when I am asked whether Beijing is beautiful? Not necessarily, maybe I want to hide the fact that Beijing is beautiful.

Basic Functionalism. Each type of mental state is a state consisting of a disposition to act in certain ways and to have certain mental states, given certain sensory inputs and certain mental states.

>> it is also related to other mental states, not only the directly concerned one

e.g My telling Beijing is beautiful is also related with my attitude towards the questioner.

Problem with behaviorism – no input-output laws. I don't always say that Beijing is beautiful when I'm asked. The law "A person believes that p if and only if this person is disposed to say p when asked about p" is false.

>> not simple "input-output"

Machine state functionalism

How would you describe how a coke machine works?

Inputs. The machine reacts to certain inputs (i.e., money)

Outputs. It gives you coca cola in return for your money.

But, sometimes it gives you coke when you give it 50 cents, sometimes it does not

Internal state 1. If the machine is in state 1, then it will not give you coke in response to 50 cents.

Internal state 2. If the machine is in state 2, then it will return coke for 50 cents.

Coke machines are best described in terms of input-ouput-internal state laws.

What are these internal states that play these roles? … they are physical machine states.

Now consider a mental case: BR = the belief that it is raining.

Input: BR tends to be caused, among other things, by seeing rain.

Output: BR tends to cause one to put on one's raincoat, if one wants to go outside, and if one doesn't like rain.

Relation to other states: BR tends to cause the belief that the streets are going to be wet.

What plays this role:

a. A physical brain state (remember machine state functionalism)

>> physical v.s propositional

b. A proposition - You can't describe thinking without propositions.

New challenges – large language models

Fodor famously said that the language of thought hypothesis is the "only game in town". The language of thought is an internal reasoning system (remember Aristoteles' condition).

P1. John believes that if he does p then q

P2. John desire q.

C. Therefore, John does p.

Explicit reasoning with rules. We need LOT to explain reasoning, thinking, believing.

Large Language models can perform reasoning tasks without explicit rules:

Prompt: suppose I love milk and know that milk is in the fridge. What should I do?

GPT 4: If you love milk and you know that there's milk in the fridge, the most sensible thing to do is to go to the fridge and pour yourself a glass of milk. Enjoy your delicious drink! However, if the milk is past its expiration date or smells and looks spoiled, it's better to discard it and not risk your health.

Personal Identity

Kinds of Identity:

Identity: things sharing all properties (Indiscernibility)
Qualitative Identity: Two objects are qualitatively the same iff they share all their properties. e.g. two billiard balls
Numerical identity: Being numerically identical means being the very same thing, e.g., Clark Kent and Superman
>> being the very same thing at one point in time/at different points in time
So, What kinds of changes can I survive?

Two questions about personal identity

Persistence Question. What makes me the same person over time?
  Persistence question → decide the fundamental self ↓ (what can I live without)
Fundamentality Question. What kind of thing am I, fundamentally speaking?
  Animalism: Humans are just animals.
  Because, If I am essentially X, then I cannot possibly lack X.
    If I am essentially my brain, then I cannot lose my brain.
    If I am a series of memories, then I'm gone once all my memories are lost.
  Compare **phase sortals** and **substance sortals**:

Substance sortal/phase sortal
Substance sortal = properties that do not change over time
Phase sortal = properties that only exist over a given phase/episode
>> intuitively, Personal identity seems to be something that doesn't change through time.

Psychological continuity

Psychological Continuity: A person remains the same person through time if and only if her psychology is continuous.
Teleportation case. Imagine that there is a teleportation device that destroys your original body, and creates a molecule-for-molecule identical copy at some other place.
>> According to this view, you survive.
Branch-line case problem. One day you try to teletransport yourself, but the machine is malfunctioning. You are not destroyed, but your counterpart is created nevertheless.
>> Whether I survived seems should not be related with changes somewhere else.
But Persons are Tokens. You are a token, and not a type (that can be instantiated twice). Thus, there is at most one of you.
>> Psychological continuity views need to rule out such cases explicitly.
Transitivity. If A is identical to B, and B is identical to C, then A is identical to C.
>> the Sorites Paradox
Strong connectedness. that is, the identity of an object over time is preserved through the preservation of enough of its members.
Continuity. Is the holding of overlapping chains of strong connectedness members.

Physical continuity

Physical Continuity. A person remains the same person through time if and only if her body (or parts of it) is (are) continuous.
Brain-based physical continuity. A person remains the same person through time if and only if her brain (or parts of it) is (are) continuous.
You stay with your brain. If your brain is transplanted into a different body, this is a way for you to acquire a new body, and not a way for someone else to acquire a new brain.

Spectrum cases

Greta Garbo Transformation. Suppose that a scientist were to begin replacing your cells, memories, and other mental states, one by one, with those of Greta Garbo at the age of thirty. At the beginning of the experiment, you'd of course still be you, and at the end it would clearly be Garbo, but what about the middle?
A sharp cut-off, or vague cases?

Parfit's view of what we are

Reduction. My identity can be reduced a range of psychological and physical facts (e.g., facts brain activity and memories).
Existence. I do exist, but I am nothing over and above the physical and psychological states that I reduce to.
Language. In some cases, whether or not I exist is just a matter of how I decide to use language.
>> Parfit thinks we are like heaps: sometimes, our existence is indeterminate: Whether I'm me is only a problem of language because the hard problem of my identity cannot be vague.
>>>> distinction of "who am I" and "what I should care about" (latter, psychological continuity)

# Kripke: picking up a person in essential/non-essential way
>> e.g Nixon as Nixon v.s Nixon as the winner
Parfit's challenge: there being no essential character of an entity for which losing it makes the entity not it anymore.

Phenomenal Consciousness = qualitative character + subjective character
"质"的特征 主观（第一视角-"我"的存在）
Memory=Remember qua intentionality（曾经存在）身处情景之中；
First personal perspective=第一视角（多重人格）

想象/虚构：
1. Material(hyle) = from memory
2. View from somewhere(~~nowhere~~)：身体图式="这里"(absolute "here")+绝对的方位感（"上下左右"的空间定位）+连续性（运动+对运动的感知）
3. 流动的时间

Mental Identity

1. Memory
2. Same reaction against same conditions

所有第三人称(third-personal)视角描述(qualitative characters)都是可错的，严格来说 entities 可以不具有这个特征而仍然存在，而只有 first-personal 视角是不可错的

Knowledge

The concept of Knowledge:

Truth:

Factual knowledge: knowledge of propositions

Knowledge requires truth (distinct from "beliefs")

Think of truth as correspondence

Belief:

Knowledge requires believing or accepting a proposition (it does not only need to be true)

Eric Schwitzgebel 2022: One can know something without believing it (it seems that knowledge can go without consciousness but belief should be a state)

Justification:

Knowledge>> epistemically good state (?)

Knowledge that p requires that p is based on evidence

Knowledge as True Justified Beliefs: but justification is the only way to decide whether something is true, so what if the justification is misleading after all?
# They're different questions, "whether there is knowledge" and "whether we have knowledge"

The Gettier problem: case 1

Not sufficient, but necessary enough: some true justified beliefs are not knowledge

Distinction of actual/conditional knowledge?
e.g If I say "assume that the plate is round, then the plate is not rectangle", it's true and I believe in this proposition and I'm justified in believing so. So it would be knowledge.

Possible Solutions:

No Luck:

Against the "either-or" case: If S knows that p, then S has ruled out all *relevant* alternatives.

Sensitivity:

If S knows that p, then, had p been false, Smith would not have believed that p.

If something changes to make the belief wrong, and S would not be sensitive to it, then S's belief is not knowledge.

Why is Knowledge important?

Knowledge as an intrinsic good:

Aristotle (*Metaphysics*), Socrates (*Apology*)

Knowledge and assertion:

Hawthhorne(2004)-Lottery, "justified" seems not enough for "knowledge"

Knowledge and inquiry:

The end of inquiry (when you have the knowledge, you stop inquiring)

Criticism and defense

Knowledge seems to be the defense against criticisms in numerous situations.
>> One may say, “You shouldn’t do sth. since you don’t know it.”

Meaningfulness may have nothing to do with the availability of definition.
>> closeness of paradigm instead of definition?

Definite description VS Proper Name (TA Session, not from the course)
Donnellan: DD are divided into two categories

1. Attributive use: used to describe property
2. Referential use: 指称某人，而非特征

Justification

Truth conductivity
Justification is in the service of truth (rather than belief)
>> e.g, being offered fortune may be a reason to take up a belief but may not justify that belief
# Externalism
Internalism holds that an individual's mental states or internal factors alone determine the justification or knowledge status of their beliefs. Externalism argues that factors external to the individual's mind, such as reliable processes, tracking the truth, or the reliability of information sources, can contribute to the justification or knowledge of beliefs.

Responsibility
Believing without justification is criticizable.

Skepticism
Epistemic responsibility and truth-conductivity can come apart
Attempts to bring them together: indubitability, self-presentingness, self-evidence, incorrigibility
>> seeking justified beliefs that fit epistemic responsibility and truth-conductivity at the same time
Counterexample for Self-presentingness:
Neurosurgeon's intervention may not be a reliable way, as opposed to sensations and reasoning.

Process Reliabilism
(Input-process-Output)
World-to-belief/belief-to-belief: reliable sensations/reasoning

Objection 1: idiot savants/chicken sexers
Objection 2: Jones' facribated amnesia
>> *"unless it's undermined"*
Objection 3: process individuation
>> for example in the "squirrel-like cat case", which one is to decide is reliable or not? Is it my ability to identify cats, to identify squirrel-like cats or to identify this cat?