Descartes's Meditations
The Matrix
Putnam's brains in vats
Moore's direct realism

Scepticism, brains in vats, and the matrix

Christian Wüthrich

http://philosophy.ucsd.edu/faculty/wuthrich/

15 Introduction to Philosophy: Theory of Knowledge
Spring 2010

Descartes's Meditations | First Meditation
The Matrix | Second Meditation
Putnam's brains in vats | Third through Sixth Meditation
Moore's direct realism | Conclusion and appraisal

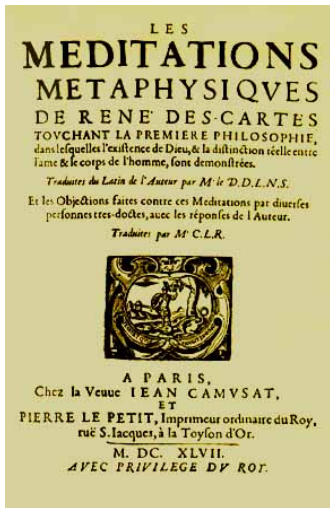# René Descartes (1596-1650)
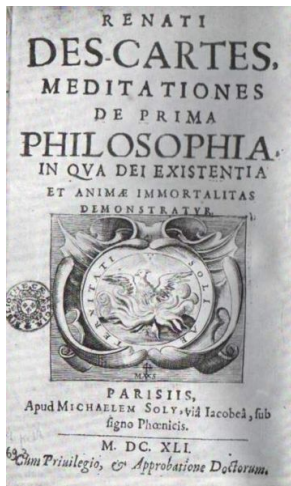
Recommended: `Philosophy Bites Interview with Anthony Grayling on Descartes' Cogito`

Meditationes de prima philosophia, *in qua Dei existentia et animae immortalitas demonstratur* (1641)

(*Meditations on First Philosophy, in which are demonstrated the existence of God and the distinction between the human soul and the body*)

Descartes's Meditations	First Meditation
The Matrix	Second Meditation
Putnam's brains in vats	Third through Sixth Meditation
Moore's direct realism	Conclusion and appraisal

Les Méditations Metaphysiques, *de René Descartes touchant la première philosophie, dans lesquelles l'existence de Dieu, et la distinction réelle entre l'âme et le corps de l'homme, sont demonstrées* (1647)

Descartes's Meditations | First Meditation
The Matrix | Second Meditation
Putnam's brains in vats | Third through Sixth Meditation
Moore's direct realism | Conclusion and appraisal

Preliminary remarks

- *Meditations* often considered starting point of modern Western philosophy
- *Meditations* tie together metaphysics, epistemology, philosophy of religion, philosophy of mind
- understanding of universe as governed by small number of fundamental abstract (often mathematical) principles
- metaphysics developed in *Meditations* constructed to offer foundation for new physics of pre-Newtonian C17
- *Meditations* reminiscent of Ignatius of Loyola's *Spiritual Exercises*
- like latter, *Meditations* consist of three stages:
  1. purgation (skeptical doubt)
  2. illumination (proof of the existence of the self, of God)
  3. union (connecting illumation to material world)

Descartes's Meditations
The Matrix
Putnam's brains in vats
Moore's direct realism

First Meditation
Second Meditation
Third through Sixth Meditation
Conclusion and appraisal

First Meditation

"What can be called into doubt"

- resolution of meditator: sweep away all he thinks he knows and start from scratch by building knowledge on certain foundation

- rather than doubt each of his beliefs individually, he doubts foundations and basic principles upon which these beliefs are founded

- most of his beliefs have been acquired either from or through the senses

- foundations of senses may be doubted in several ways: insanity, dreaming, deceitful demon

- first, meditator establishes his sanity (otherwise we might doubt rationality of his arguments)

Descartes's Meditations
The Matrix
Putnam's brains in vats
Moore's direct realism

First Meditation
Second Meditation
Third through Sixth Meditation
Conclusion and appraisal

- dreams: although we can doubt existence of particulars and "composite things", it seems as if certain simple and universal things (redness, doghood, betweenness) are real

- reasonable conclusion: physics, medicine etc relying on particulars in their study of composite objects are doubtful; while mathematics dealing only with general and simple (!) things generates more certain knowledge ("painter's analogy")

- even existence of simple things can be doubted: malicious demon might employ all its energies to deceive meditator

⇒ all knowledge, even of universals, could be mere delusions devised by demon to ensnare meditator's judgment

Descartes's Meditations
The Matrix
Putnam's brains in vats
Moore's direct realism

First Meditation
Second Meditation
Third through Sixth Meditation
Conclusion and appraisal

The argument structure

1. If I am dreaming/deceived, then my beliefs will in general not be true.

2. I do not know whether I am dreaming/deceived.

3. Therefore, I do not know whether my beliefs are true or not.

- Dream argument: undermines reliability of senses if read as suggesting universal possibility of dreaming, and only *of senses* (questions Aristotelian empiricism)

- Evil Demon argument: suggests *all we know* may be false, *and* we cannot trust senses

Descartes's Meditations
The Matrix
Putnam's brains in vats
Moore's direct realism

First Meditation
Second Meditation
Third through Sixth Meditation
Conclusion and appraisal

Second Meditation

"The nature of the human mind, and how it is better known than the body"

Summary of first part (cogito and sum res cogitans):

- Archimedes: can move entire Earth if only given one immovable point

- similarly: Descartes hopes to achieve great things if he could only have one certain vantage point

- Even though he concluded that he may have no senses and no body, does this mean that he doesn't exist?

- Yet in order to have the doubts he raised in the first meditation, he must exist, i.e. there must be an "I" that exists if that "I" doubts, can be deceived, etc.

Descartes's Meditations	First Meditation
The Matrix	Second Meditation
Putnam's brains in vats	Third through Sixth Meditation
Moore's direct realism	Conclusion and appraisal

- Famous cogito "argument": "So after considering everything very thoroughly, I must finally conclude that this proposition, *I am, I exist*, is necessarily true whenever it is put forward by me or conceived in my mind."

- But what kind of thing is this *I*?

- all its potential attributes are doubtful except one: that it thinks

- meditator concludes that he is only a thing that thinks

- mind is res cogitans (as opposed to bodies which are defined as *extended things*, i.e. res extensa)

- **Careful:** what really is established is that there are thoughts now, not that there is a thinking thing, let alone a "self" or "I"

Descartes's Meditations | First Meditation
The Matrix | **Second Meditation**
Putnam's brains in vats | Third through Sixth Meditation
Moore's direct realism | Conclusion and appraisal

Analysis of first part

- Argument derives its name from Latin version in the *Discourse on Method*: "cogito ergo sum" ("I think, therefore I am")

- plays role of Archimedean certain vantage point which overcomes radical skepticism of First Meditation

- turning point in Western philosophy: mind is no longer something that helps us grasp world, but instead becomes something inside which we are epistemically locked.

- Important: distinction between *I think, therefore I am* of the *Discourse* and the *I am, I exist* of the *Meditations* where "therefore" is absent;

Descartes's Meditations	First Meditation
The Matrix	Second Meditation
Putnam's brains in vats	Third through Sixth Meditation
Moore's direct realism	Conclusion and appraisal

$\Rightarrow$ we shouldn't read *cogito* as a three-step argument of the following form:

> **1** Whatever thinks exists.
> **2** I think.
> $\therefore$ I exist.

- Descartes explicitly denies this reading elsewhere; why?

- Two reasons: (i) first premise not immune from systematic doubt, and (ii) cogito shouldn't be read as reasoned inference at a point when even reasoned inference might be doubted.

- How then can the meditator know the *cogito*?

- intuition or performative utterance rather than inference, i.e. as something that is known "at once" or is confirmed by its being uttered

Descartes's Meditations	First Meditation
The Matrix	Second Meditation
Putnam's brains in vats	Third through Sixth Meditation
Moore's direct realism	Conclusion and appraisal

- *sum res cogitans* argument (mind is thinking thing):
- epistemological reading: I only *know* that I am a thing that thinks
- metaphysical reading: I *am* only a thing that thinks
- "thing": likely used to mean "substance", i.e. fundamental and indivisable elements of existence
- "thinking"?

Descartes's Meditations
The Matrix
Putnam's brains in vats
Moore's direct realism

First Meditation
Second Meditation
Third through Sixth Meditation
Conclusion and appraisal

Summary of second part (wax argument)

- mind also desires, wills, imagines, senses etc. because although he may be deceived (and the sensory perceptions thus not be veridical), but these are still activities of the mind

- If meditator is *res cogitans*, why does he have distinctive perception of his *res extensa*, i.e. his body, yet such nebulous idea of this "I" that thinks?

- wax argument: how do we know a piece of wax?

- by sensory perception: color, smell, taste, shape, temperature, hardness...

- but if melted, wax will change all its sensible qualities *although it's still the same piece of wax*

- ⇒ knowledge of this sameness cannot come through senses

Descartes's Meditations   First Meditation
The Matrix   Second Meditation
Putnam's brains in vats   Third through Sixth Meditation
Moore's direct realism   Conclusion and appraisal

- can only know of wax that it is extended, flexible, and changeable–and this we can only know through intellect

- mental perception can be confused and obscure (when led by senses and imagination) or clear and distinct (when careful mental scrutiny is applied)

⇒ mind as knower superior to body; also, mind knows itself far better than anything else

- in conclusion, meditator can know at least that he exists, that he is a thinking thing, that his mind is better knower than his body, and that all clear and distinct perceptions are mental, rather than sensory

Descartes's Meditations
The Matrix
Putnam's brains in vats
Moore's direct realism

First Meditation
Second Meditation
Third through Sixth Meditation
Conclusion and appraisal

Analysis of wax argument

- ninth para of Second Meditation: birth of modern mind

- Cartesian mind sharply distinguished from "world," and philosophy has been concerned how to connect the two ever since

- I have visual perceptions of things in world, but visual sensations are in my mind, and objects I perceive in world, so how can we connect the two?

- wax argument: senses only perceive disorganized influx of information, only intellect can help us organize and understand it

- must be read against backdrop of then predominant Aristotelian epistemology that all knowledge comes from the senses

Descartes's Meditations
The Matrix
Putnam's brains in vats
Moore's direct realism

First Meditation
Second Meditation
Third through Sixth Meditation
Conclusion and appraisal

Third through Sixth Meditation

3. two arguments for existence of God

   - "design" argument: "just as the objective intricacy of idea of very perfect machine must have a cause in scientific knowledge of engineer, so must idea of God in us have God himself as cause."

4. perfect God wouldn't deceive me, so He doesn't deceive me about what I have a clear and distinct perception

5. I can define material objects via mathematics, and insofar this is possible I can have certain knowledge of them; however, this certainty doesn't make them exist, which only God can (and in fact does).

6. mind-body dualism, "but such that they still partake in unity"

Descartes's Meditations
The Matrix
Putnam's brains in vats
Moore's direct realism

First Meditation
Second Meditation
Third through Sixth Meditation
Conclusion and appraisal

Conclusion and appraisal

Descartes concludes:

> *"The great benefit of these arguments is not, in my view, that they prove what they establish—namely that there really is a world, and that human beings have bodies and so on—since no sane person has ever seriously doubted these things. The point is that in considering these arguments we come to realize that they are not as solid or as transparent as the arguments which led us to knowledge of our own minds and of God, so that the latter are the most certain and evident of all possible objects of knowledge for the human intellect."*

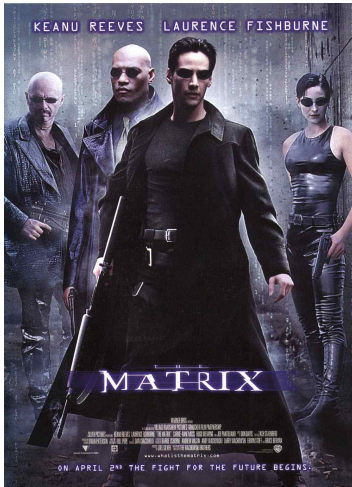| Descartes's Meditations | First Meditation |
The Matrix	Second Meditation
Putnam's brains in vats	Third through Sixth Meditation
Moore's direct realism	**Conclusion and appraisal**

Some early objections

Descartes circulated manuscript of *Meditations* prior to publication among philosophers and theologians; their objections and his replies were included in publication

Here are some of the most substantial objections included:

- On what grounds can we be certain that what we perceive as clear and distinct is indeed clear and distinct?

- Circularity: if we have no basis on which we can be certain that clear and distinct ideas are true before we prove God's existence, then

  1. we can't be certain that God exists, since we use clear and distinct ideas to prove God's existence; and
  2. we can't be certain that we are a thinking thing.

- We are certain that bodies exist or that perception faithfully represents reality.

Descartes's Meditations
The Matrix
Putnam's brains in vats
Moore's direct realism

The movie
Chalmers's anti-scepticism

The Matrix: The movie trilogy

- Wikipedia entry for *The Matrix* (1999)
- Wikipedia entry for *The Matrix Reloaded* (2003)
- Wikipedia entry for *The Matrix Revolutions* (2003)

Descartes's Meditations
The Matrix
Putnam's brains in vats
Moore's direct realism

The movie
Chalmers's anti-scepticism

David Chalmers (*1966)

- Professor of Philosophy at Australian National University in Canberra
- PhD 1993 in Philosophy and Cognitive Science at Indiana University (Bloomington) under Douglas Hofstadter
- works primarily in philosophy of mind
- *The Conscious Mind: In Search of a Fundamental Theory* (1996): non-reductive science of consciousness, defends version of property dualism
- Wikipedia entry

Descartes's Meditations
The Matrix
Putnam's brains in vats
Moore's direct realism

The movie
Chalmers's anti-scepticism

Chalmers: "The Matrix as Metaphysics"

in Christopher Grau (ed.), *Philosophers Explore the Matrix*, Oxford University Press (2005), pp. 132-176.

First, a bit of terminology:
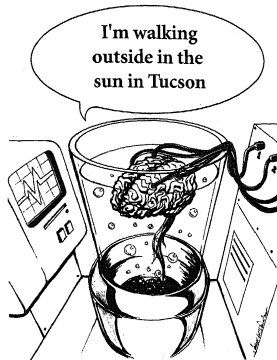
Definition (matrix (lower case))

A matrix *(lower case) is "an artificially designed computer simulation of a world." (133)*

Thus, the Matrix is a particular version of a matrix.

Definition (envatment)

Someone is envatted *(or* in a matrix*) "if they have a cognitive system which receives its inputs from and sends its outputs to a matrix." (ibid.)*

The question that now arises is, of course, how do I know whether or not I am in a matrix...

Descartes's Meditations
The Matrix
Putnam's brains in vats
Moore's direct realism

The movie
Chalmers's anti-scepticism

I'm walking outside in the sun in Tucson

Hypothesis (Matrix)

I am in a matrix; or, equivalently, I am envatted and have always been envatted.

- Matrix Hypothesis seems to be a skeptical hypothesis in that (if true) it seems to render almost all of my beliefs concerning the world false

- Reasoning: "I don't know that I'm not in a matrix. If I'm in a matrix, I'm probably not in Tucson. The same goes for almost everything else I think I know about the external world." (135)

Descartes's Meditations
The Matrix
Putnam's brains in vats
Moore's direct realism

The movie
Chalmers's anti-scepticism

Chalmers:

- I cannot rule out the Matrix Hypothesis.

- But: it is not a skeptical hypothesis because even if it's true, most of my beliefs will still be true!

- For instance: even if I'm envatted, I am still walking outside in the sun in Tucson

- Instead, it's a metaphysical hypothesis, i.e. it concerns the nature of the most fundamental level of reality.

- Chalmers: in fact, the Matrix Hypothesis is equivalent to the conjunction of three hypotheses:
  1. Computational Hypothesis
  2. Creation Hypothesis
  3. Mind-Body Hypothesis

Descartes's Meditations
The Matrix
Putnam's brains in vats
Moore's direct realism

The movie
Chalmers's anti-scepticism

The Computational Hypothesis

Atoms

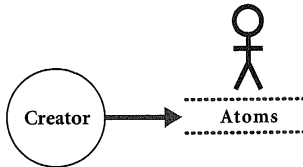**110101101011**
011010000111

The Computational Hypothesis

Hypothesis (Computational)

"Microphysical processes throughout space-time are constituted by underlying computational processes." (137)

- don't know that it is true, don't know that it is false, but it's coherent

- not skeptical: elementary particles etc are just more like tables and chairs, and so fundamental reality is different from what we thought, but it still exists and most of our ordinary beliefs are not affected by its truth or falsity

Descartes's Meditations
The Matrix
Putnam's brains in vats
Moore's direct realism

The movie
Chalmers's anti-scepticism
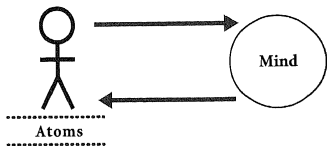
The Creation Hypothesis

The Creation Hypothesis

Hypothesis (Creation)

"Physical space-time and its contents were created by beings outside physical space-time." (139)

- don't know that it is true, don't know that it is false, but it's coherent

- not skeptical: even if true, most of my ordinary beliefs remain valid

Descartes's Meditations
The Matrix
Putnam's brains in vats
Moore's direct realism

The movie
Chalmers's anti-scepticism

The Mind-Body Hypothesis

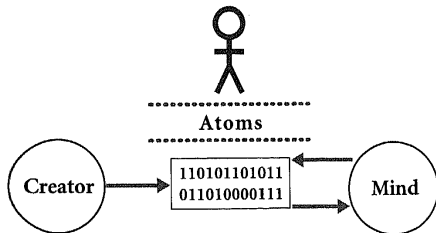*The Mind-Body Hypothesis*

Hypothesis (Mind-Body)

"My mind is (and has always been) constituted by processes outside physical space-time and receives its perceptual inputs from and sends its outputs to processes in physical space-time." (140)

- don't know that it is true, don't know that it is false, but it's coherent
- not skeptical: even if true, most of my ordinary beliefs remain valid

Descartes's Meditations
The Matrix
Putnam's brains in vats
Moore's direct realism

The movie
Chalmers's anti-scepticism

The Metaphysical Hypothesis

The Metaphysical Hypothesis

Hypothesis (Metaphysical)

Physical space-time and its contents were created by beings outside physical space-time. Both physical space-time and the microphysical processes it contains are constituted by computational processes that were designed as a computer simulation of the world. Also, our minds are outside physical space-time but interact with it. (Cf. 141)

Descartes's Meditations
The Matrix
Putnam's brains in vats
Moore's direct realism

The movie
Chalmers's anti-scepticism

The Matrix Hypothesis as a Metaphysical Hypothesis

- Again, Metaphysical Hypothesis is coherent and not skeptical.

- Chalmers: Matrix Hypothesis is equivalent to Metaphysical Hypothesis, i.e. they imply one another

- Metaphysical → Matrix: from Mind-Body, Computational and Creation Hypotheses, it follows that "I have (and always had) a cognitive system that receives its input from and sends its output to an artifically designed computer simulation of the world" (142), but that's just the Matrix Hypothesis

- Matrix → Metaphysical: accepting Matrix means to accept that whatever underlies apparent reality is really just as Metaphysical Hyp claims, viz. that there is a domain containing my mind, which causally interacts with an artificially created computer simulation of physical space-time and its contents (cf. 143)

Descartes's Meditations
The Matrix
Putnam's brains in vats
Moore's direct realism

The movie
Chalmers's anti-scepticism

Conclusion: anti-skepticism

- If this is right, then the Matrix Hypothesis is not a skeptical hypothesis as the resulting picture is one of a "full-blooded external world" (144), even though it entails that fundamental reality is quite a bit different from what our currently best scientific theories tell us.

- (Study the qualifications on pages 145 and 146)

- Let's consider a few objections.

  1. Envatted brain may think it is in Tucson when in fact it is in Sydney. Response: the envatted brain's concept of "Tucson" does not refer to Tucson, but to something else entirely (call it "Tucson*")

  2. But what sort of thing does the envatted being refer to? Response: entities constituted by computational processes

Descartes's Meditations
The Matrix
Putnam's brains in vats
Moore's direct realism

The movie
Chalmers's anti-scepticism

A truly skeptical hypothesis

Hypothesis (Chaos)

"I do not receive inputs from anywhere in the world. Instead, I have random, uncaused experience. Through a huge coincidence, they are exactly the sort of regular, structured experiences with which I am familiar." (158)

- coherent, but has minuscule probability
- truly skeptical: if true, almost all of our beliefs would be true, and almost none of our concepts could refer (to physical objects, or patterns of bits in computational processes)

    "[I]f we are granted the assumption that there is some explanation for the regularities in our experience, then it is safe to say that some of our beliefs about the external world are correct. This is not much, but it is something.
    In conclusion: It's not so bad to be a brain in a vat." (159)

Descartes's Meditations   **Brains in vats: The skeptical argument reconsidered**
The Matrix   Semantic externalism and twin earth
Putnam's brains in vats   Putnam's anti-skeptical stance
Moore's direct realism   Upshot

The skeptical argument reconsidered

- Suppose you inhabit a world consisting just of your mind and of a malicious demon.

- Nothing physical lives in this world, all experiences are directly caused by demon.

- Materialists/Monists: mind is complex physical system, i.e. mind couldn't possibly exist in matterless world.

- Can we come up with a skeptical argument which is consistent with materialism?

Descartes's Meditations
The Matrix
Putnam's brains in vats
Moore's direct realism

Brains in vats: The skeptical argument reconsidered
Semantic externalism and twin earth
Putnam's anti-skeptical stance
Upshot

Brains in vats... (BIVs)

Descartes's Meditations
The Matrix
Putnam's brains in vats
Moore's direct realism

Brains in vats: The skeptical argument reconsidered
Semantic externalism and twin earth
Putnam's anti-skeptical stance
Upshot

The skeptical argument

- Brain floating in nutrient fluids is disembodied. ("The Matrix" depicts embodied brains).

- Important: the brain has experiences which are qualitatively indistinguishable from those of normal perceiver.

- Skeptical challenge reformulated: on what grounds can you rule out this possibility? Skeptic argues:

  1. If you know that *p*, then you know that you are not a brain in a vat.
  2. You don't know that you are not a brain in a vat.
  ∴ You don't know that *p*.

Descartes's Meditations
The Matrix
Putnam's brains in vats
Moore's direct realism

Brains in vats: The skeptical argument reconsidered
Semantic externalism and twin earth
Putnam's anti-skeptical stance
Upshot

Epistemic closure under entailment

Principle (Epistemic closure under entailment)

For all epistemic agents S and for all propositions p, q, if S knows that p, and S knows that p entails q, then S knows that q.

- Premise (1) in foregoing argument supported by closure principle: you know that *p* and you know that *p* entails that you are not a brain in a vat (BIV), therefore you know that you are not a BIV.

- Premise (2) is supported by claim that experiences in both cases are qualitatively indistinguishable *ex hypothesi*.

Descartes's Meditations
The Matrix
Putnam's brains in vats
Moore's direct realism

Brains in vats: The skeptical argument reconsidered
Semantic externalism and twin earth
Putnam's anti-skeptical stance
Upshot

Hilary Putnam (*1926)
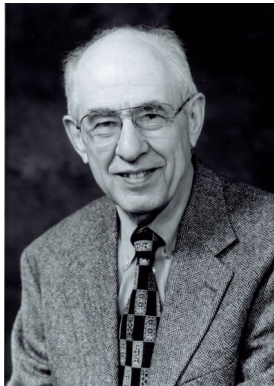
- phil of mind: hypothesis of multiple realizability, concept of functionalism
- phil of language: causal thy of reference, semantic externalism
- phil of mathematics: Quine-Putnam indispensability thesis
- phil of physics: disproof of openness of future based on special relativity
- contributions to metaphysics, mathematics, computer science
- epistemology: brain-in-vat argument against skeptic ("Brains in a vat", 1981)

Descartes's Meditations
The Matrix
Putnam's brains in vats
Moore's direct realism

Brains in vats: The skeptical argument reconsidered
Semantic externalism and twin earth
Putnam's anti-skeptical stance
Upshot

Crucial assumption: semantic externalism (SE)

Characterization (Semantic externalism)

Semantic externalism *is the view that the meaning of a term, word, symbol etc is at least partially determined by factors external to the speaker, in particular insofar as the term's reference is concerned.*

- For a semantic externalist, meaning isn't just in words; rather, the speaker must stand in the appropriate causal relationship to referent.

- ⇒ There could be two speakers in exactly the same state, uttering the exact same combination of words, yet meaning rather different things by that utterance.

- Generally considered a necessary consequence of a causal thy of reference, i.e. account assuming initial act of fixing reference ("naming") and subsequent usage causally linked to this original act. (Ex. elm vs. birch trees and division of linguistic labour)

Descartes's Meditations
The Matrix
Putnam's brains in vats
Moore's direct realism

Brains in vats: The skeptical argument reconsidered
Semantic externalism and twin earth
Putnam's anti-skeptical stance
Upshot

Master argument for SE: Twin Earth thought experiment

Hilary Putnam, "Meaning and reference", *Journal of Philosophy* **70** (1973): 699-711.

> *"...[W]e shall suppose that somewhere there is a planet we shall call Twin Earth. Twin Earth is very much like Earth: in fact, people on Twin Earth even speak English. In fact, apart from the differences we shall specify..., the reader may suppose that Twin Earth is exactly like Earth... One of the peculiarities of Twin Earth is that the liquid called 'water' is not H_2O but a different liquid whose chemical formula is very long and complicated. I shall abbreviate this chemical formula simply as XYZ. I shall suppose that XYZ is indistinguishable from water at normal temperatures and pressures. Also, I shall suppose that the oceans and lakes and seas of Twin Earth contain XYZ and not water, that it rains XYZ on Twin Earth and not water, etc." (700f)*

Descartes's Meditations | Brains in vats: The skeptical argument reconsidered
The Matrix | Semantic externalism and twin earth
Putnam's brains in vats | Putnam's anti-skeptical stance
Moore's direct realism | Upshot

"...Now let us roll the time back to about 1750. The typical Earthian speaker of English did not know that water consisted of hydrogen and oxygen, and the typical Twin-Earthian speaker of English did not know that 'water' consisted of XYZ." (701f)

- In fact, the experiences that earthians have of H_2O and those that twin-earthians have of XYZ are qualitatively indistinguishable.

- Question: when an earthling (say, Oscar) and his counterpart on Twin Earth utter "water" do they mean the same thing?

Descartes's Meditations
The Matrix
Putnam's brains in vats
Moore's direct realism

Brains in vats: The skeptical argument reconsidered
Semantic externalism and twin earth
Putnam's anti-skeptical stance
Upshot

- Twin is also called "Oscar" on his planet; in fact, planet is also called "Earth"

- Naming convention: planet is referred to as "Twin Earth," Oscar's twin as "Twin-Oscar" or "Toscar," twin-water as "twater"

- *ex hypothesi*: Oscar and Toscar molecule-by-molecule identical

- yet: when Oscar utters "water," he refers to H_2O, whereas when Toscar utters "water," he refers to XYZ

⇒ contents of a person's brain insufficient to determine reference of terms used uniquely

- causal history of how this individual has acquired usage of term must be taken into account

- Putnam: " 'meanings' just ain't in the head"

  ("The meaning of 'meaning' ", in *Mind, Language and Reality. Philosophical Papers, vol. 2.* Cambridge: Cambridge University Press, (1975), p. 227.)

Descartes's Meditations
The Matrix
Putnam's brains in vats
Moore's direct realism

Brains in vats: The skeptical argument reconsidered
Semantic externalism and twin earth
Putnam's anti-skeptical stance
Upshot

Apply SE to BIVs...

- Consider Mars (instead of Twin Earth), a planet entirely without trees populated by inhabitants who lack any causal connection to trees.

- So when a Martian utters "tree," it doesn't refer to anything.

- If the Martian has mental image of tree, this image doesn't represent a tree, etc...

- (Similarly, ant's path doesn't represent Churchill in any way).

- Similarly, the mental image or utterance of a BIV who lacks a causal connection to tree doesn't represent a tree.

- So what does the BIV's mental image refer to? Putnam sees three possibilities...

Descartes's Meditations | Brains in vats: The skeptical argument reconsidered
The Matrix | Semantic externalism and twin earth
Putnam's brains in vats | **Putnam's anti-skeptical stance**
Moore's direct realism | Upshot

1. to "trees-in-the-image" (image $\approx$ succession of experiences had by BIV)

2. to electric impulses that stimulate brain s.t. it has experience of tree

3. to computer program features which are causally responsible for stimuli described in (2)

$\Rightarrow$ E.g. on account (1), a BIV's utterance "Here is a tree" is true iff the BIV is having experiences as of being near a tree.

$\Rightarrow$ SE denies crucial Cartesian assumption about relation between mind and world, viz. that BIV's utterances express systematically mistaken beliefs about external world.

- Generally: BIV's utterances differ in reference and truth conditions from standard utterances.

Descartes's Meditations | Brains in vats: The skeptical argument reconsidered
The Matrix | Semantic externalism and twin earth
Putnam's brains in vats | **Putnam's anti-skeptical stance**
Moore's direct realism | Upshot

Sketch of Putnam's argument

- causal thy of reference ⇒ if there were a BIV, it couldn't refer to, and thus think about, brains or vats

- Its use of "brain" and "vat" couldn't possibly refer to brains and vats bc it has neither seen a brain or a vat, nor communicated with someone who has, nor could it in some other way be causally linked to original naming act.

⇒ A BIV cannot entertain a brain-in-a-vat scenario.

⇒ If we are entertaining a brain-in-a-vat scenario, then we are not BIVs.

Descartes's Meditations   Brains in vats: The skeptical argument reconsidered
The Matrix   Semantic externalism and twin earth
Putnam's brains in vats   Putnam's anti-skeptical stance
Moore's direct realism   Upshot

The point of Putnam's argument

In an interview with *The Harvard Review of Philosophy* **2** (Spring 1992): p. 22.

> *My discussion of the 'brain in a vat' model of Cartesian*
> *scepticism is too long to summarize here, but I can say*
> *what my purpose was: my purpose was to argue that*
> *concepts and world involve each other, that the concepts*
> *you have depend on the world you inhabit and how you are*
> *related to it. The idea that we first have concepts in some*
> *purely 'private' medium and we must then proceed to see if*
> *anything corresponds to them has had a powerful grip on*
> *our thinking ever since Descartes, but it is at bottom*
> *completely incoherent.*

Descartes's Meditations | Brains in vats: The skeptical argument reconsidered
The Matrix | Semantic externalism and twin earth
Putnam's brains in vats | **Putnam's anti-skeptical stance**
Moore's direct realism | Upshot

Simple Reconstruction of Putnam's argument

Brueckner Tony, "Brains in a vat", *Stanford Encyclopedia of Philosophy* (2004).

(1) If I am a BIV, then my word "tree" does not refer to trees.

(2) My word "tree" refers to trees.

∴ I am not a BIV.

- Account (3) [= trees are computer program features] arguably offers most plausible externalist assigment of reference, so use it to construct the following argument.

- Premise (1) directly from SE, but (2) seems question-begging ⇒ modify simple reconstruction...

Descartes's Meditations | Brains in vats: The skeptical argument reconsidered
The Matrix | Semantic externalism and twin earth
Putnam's brains in vats | **Putnam's anti-skeptical stance**
Moore's direct realism | Upshot

Modified Simple Reconstruction

- (1)* If I am a BIV, then it is not the case that if my word "tree" refers, then it refers to trees.

- (2)* If my word "tree" refers, then it refers to trees.

  ∴ I am not a BIV.

- (1)*: from SE

- (2)*: knowledge that there are trees in my world not required in order to be justified

  - But: still need to presuppose *a priori* rules about disquotation, i.e. need to know which language I am operating in (English or vat-English).

  - Important constraint on anti-skeptical arguments: premises must be known a priori (not through senses).

Descartes's Meditations   Brains in vats: The skeptical argument reconsidered
The Matrix   Semantic externalism and twin earth
Putnam's brains in vats   Putnam's anti-skeptical stance
Moore's direct realism   Upshot

Let's turn the table on the skeptic...

- Given SE, skeptical hypotheses such as "I am a BIV" are not compatible with external-world propositions such as "I see a hand".

- I know *a priori* that either

  (A) Trees are features of computer programs; or
  (B) Trees are not features of computer programs.

- If (A), then skeptical position is incoherent.

- If (B), then skeptical challenge is withdrawn.

Descartes's Meditations
The Matrix
Putnam's brains in vats
Moore's direct realism

Brains in vats: The skeptical argument reconsidered
Semantic externalism and twin earth
Putnam's anti-skeptical stance
Upshot

What's the upshot?

- It seems impossible to positively establish that we are not BIVs.

- But it seems equally impossible to coherently establish that we are, or even could be, BIVs.

    "...the BIV hypothesis may well be refutable, given semantic externalism and given the assumption that one has a priori *knowledge of some key semantic properties of one's language [...] Even if Putnamian arguments fail to rule out all versions of the brain-in-a-vat hypotheses, their success against the radical BIV hypothesis would be significant."*

Brueckner (2004)

Descartes's Meditations
The Matrix
Putnam's brains in vats
Moore's direct realism

Direct realism
Moore and the external world
Moore and Hume's principles

The skeptic once again

1. If you know that *p*, then you know that you are not a brain in a vat.

2. You don't know that you are not a brain in a vat.

∴ You don't know that *p*.

- Huemer: if sound, only disproves *indirect realism* about the external world, i.e. inferential knowledge about external world.

Characterization (Inferential knowledge)

A subject S knows a proposition p inferentially *iff S's knowledge of p "is based on (or is constituted by) the fact that one can legitimately infer [p] from some other different proposition [q]"* which is known.

Richard Fumerton, "Knowledge by Acquaintance vs. Description", *Stanford Encyclopedia of Philosophy*, 2004.

Christian Wüthrich   Topic 3

Descartes's Meditations
The Matrix
Putnam's brains in vats
Moore's direct realism

Direct realism
Moore and the external world
Moore and Hume's principles

Direct realism

- Huemer: skeptic hasn't ruled out *direct realism* about the external world, i.e. the possibility of non-inferential knowledge or of non-inferentially justified beliefs about the external world.

- Non-inferential knowledge is directly "given," i.e. we are directly acquainted with the truthmakers of a non-inferentially known proposition, usually by means of perception (but sometimes *a priori*).

- Examples:
  - non-inferential: "7 equals 7"
  - inferential: "7 is the cube-root of 343"

Descartes's Meditations
The Matrix
Putnam's brains in vats
Moore's direct realism

Direct realism
Moore and the external world
Moore and Hume's principles

G[eorge] E[dward] Moore (1873-1958)

- studied and taught at Cambridge
- one of founders of analytic tradition
- ethical non-naturalism (*Principia Ethica*, 1903)
- common sense philosophy

Descartes's Meditations
The Matrix
Putnam's brains in vats
Moore's direct realism

Direct realism
Moore and the external world
Moore and Hume's principles

Kant's lament

> *"It still remains a scandal to philosophy... that the existence of things outside of us... must be accepted merely on* faith*, and that, if anyone thinks good to doubt their existence, we are unable to counter his doubts by any satisfactory proofs."*

(*Critique of Pure Reason*, B xxxix)

Descartes's Meditations
The Matrix
Putnam's brains in vats
Moore's direct realism

Direct realism
Moore and the external world
Moore and Hume's principles

Moore, "Proof of an external world"

Perfectly rigorous proof of objects of external world:

1. "[S]omething which I expressed by showing you my hands, making certain gestures, and saying the words 'Here is one hand, and here is another.' " (603)

∴ "Two human hands exist at this moment." (602)

Proof is rigorous:

- Premise and conclusion inequivalent: ✓
- Premise is known with certainty: ✓
- Certain that conclusion follows from premise: ✓

Descartes's Meditations
The Matrix
Putnam's brains in vats
Moore's direct realism

Direct realism
Moore and the external world
Moore and Hume's principles

- If someone doubts the existence of ant eaters or space aliens, it would surely be sufficient evidence for their existence if we were able to present an exemplar to the person who doubted their existence.

- Moore's proof of the existence of physical objects (and thus of the external world) follows the same pattern.

- Skeptical responses anticipated by Moore:

    1. Skeptic wants proof of what is asserted when a hand is held up while uttering "Here's a hand." Moore: such a proof cannot be given; I have only given conclusive reasons for what is asserted.
    2. Skeptic may think that if such proofs of the premise cannot be given, then the proofs given by Moore are not conclusive at all. Moore: this is mistake bc I can know things, which I cannot prove (and premise is exactly such a case).

Descartes's Meditations
The Matrix
Putnam's brains in vats
Moore's direct realism

Direct realism
Moore and the external world
Moore and Hume's principles

Moore, "Hume's theory examined"

Quoting from Alice Ambrose, *Journal of Philosophy* **51** (1954): 328-331.

Principle (Hume's First Principle H_1)

"[N]o one can know of the existence of anything not directly apprehended unless he knows that something he has directly apprehended is a sign of its existence." (329)

Principle (Hume's Second Principle H_2)

"[N]o one can know [that] the existence of A is a sign of the existence of B unless he (or someone else) has experienced a general conjunction of things like A and things like B." (329)

Descartes's Meditations
The Matrix
Putnam's brains in vats
Moore's direct realism

Direct realism
Moore and the external world
Moore and Hume's principles

- existence of material objects vs. directly apprehended sense-data

- H_1: we can only know existence of material objects if we have directly apprehended something we know is a sign of their existence.

- H_2: since, in turn, we can only know this if we have experienced a general conjunction of sense-data and material objects, which isn't possible, it follows that we cannot know the existence of material objects.

- Contrast "*If H_1 and H_2 are true*, then we cannot know the existence of material objects" with "*Since H_1 and H_2 are true*, we cannot know the existence of material objects."

- This raises the question of whether H_1 and H_2 are true...

Descartes's Meditations
The Matrix
Putnam's brains in vats
Moore's direct realism

Direct realism
Moore and the external world
Moore and Hume's principles

*"If Hume's principles are true, then, I have admitted, I do
not know now that this pencil—the material object—exists.
If, therefore, I am to prove that I do know that this pencil
exists, I must prove, somehow, that Hume's principles, one
or both of them, are not true." (Moore, 606)*

Descartes's Meditations
The Matrix
Putnam's brains in vats
Moore's direct realism

Direct realism
Moore and the external world
Moore and Hume's principles

Skeptical argument structure:

1. H_1

2. H_2

∴ S (e.g. "I do not know whether this pencil exists")

Moore: if this argument is deductively valid (i.e. H_1, H_2, and $\neg S$ are mutually incompatible), then the following two arguments are equally valid:

1. H_1	1. H_2
2. $\neg S$	2. $\neg S$
∴ $\neg H_2$	∴ $\neg H_1$

In essence: from $\neg S$, we can validly infer $\neg H_1 \vee \neg H_2$.

Descartes's Meditations
The Matrix
Putnam's brains in vats
Moore's direct realism

Direct realism
Moore and the external world
Moore and Hume's principles

Moore:

1. I know that this pencil exists.

∴ At least one of Hume's principles is false.

Skeptic:

1. Hume's principles are true.

∴ I don't know that this pencil exists.

Descartes's Meditations
The Matrix
Putnam's brains in vats
Moore's direct realism

Direct realism
Moore and the external world
Moore and Hume's principles

Who has more plausible premises?

- Moore: premises of any skeptical argument less plausible than denial of its conclusion (such as "I know that there is a pencil here").

⇒ We can use particular knowledge about existence of a particular material body to discredit basic principles of Humean skeptic.

Descartes's Meditations
The Matrix
Putnam's brains in vats
Moore's direct realism

Direct realism
Moore and the external world
Moore and Hume's principles

How can we know a premise?

1 Inferential knowledge
- Any proposition is known to be true if we have a conclusive argument in its favour.
- Moore: threat of infinite regress
- ⇒ Knowledge cannot be purely inferential, i.e. ∃ non-inferential basis/foundation ("immediate knowledge")
- Challenge: how do you defend this?

2 Non-inferential knowledge ("immediate knowledge")
- A proposition may be known to be true immediately.

Moore: premise such as "I know that this pencil exists" is known immediately (unlike Hume's principles).

Descartes's Meditations
The Matrix
Putnam's brains in vats
Moore's direct realism

Direct realism
Moore and the external world
Moore and Hume's principles

Wrapping up

(Audi, "Introduction", in Huemer (ed.), p. 23)

- Distinction: rebutting skeptical case (= showing it's unsound) vs. refuting it (= showing its conclusion to be false)

- refutation suffices but is not necessary for rebuttal (e.g. show skeptical arg to be unsound without showing its conclusion to be false)

- Example: all animals can fly, birds are animals, therefore, all birds can fly.

- prospects for rebuttal better than for refutation