

# The Externalist and the Structuralist Responses To Skepticism

David Chalmers

# Overview

- In *Reason, Truth, and History*, Hilary Putnam mounts an externalist response to skepticism.
- In ‘The Matrix as Metaphysics’ (and *Constructing the World*), I mount a structuralist response to skepticism.
- Aim: compare these and argue that the structuralist response succeeds where the externalist response fails.

# Content Externalism

- Externalism = content externalism
  - mental content is not in the head
- Focus on Putnam-style externalism
  - *water* refers to H<sub>2</sub>O for Oscar
  - *water* refers to XYZ for Twin Oscar

# Global Skepticism

- Skepticism = global external-world skepticism
- For all we know, few or none of our positive beliefs about the external world are true
- The sort of skepticism suggested by evil demon, brain-in-vat, Matrix arguments

# Local Skepticism

- Global skepticism is stronger than local skepticism: for most or all  $p$ ,  $p$  is not true for all we know.
- The sort of skepticism suggested by e.g. painted mule scenarios (casting doubt on our beliefs one at a time)
- Local skepticism suffices for many skeptical purposes, but it's not the focus here.

# Externalist Replies to Skepticism

- Externalist replies to skepticism argue from content externalism to anti-skeptical conclusions.
- More specifically: from specific externalist theories of content (e.g. causal theories) to the denial of global external-world skepticism.

# Putnam's Externalist Replies

- In *Reason, Truth, and History* (chapter I), Hilary Putnam argues from causal theories of content to an anti-skeptical conclusion.
- He gives at least two arguments, which should be distinguished.

# Putnam's Two Arguments

- Argument 1: We can rule out the hypothesis that we are brains in vats (BIVs), since BIVs can't refer to BIVs with 'BIV'.
- Argument 2: Even if we are BIVs, most of our beliefs are true, since BIVs have mostly true beliefs.



# The Structuralist Reply

- In 'The Matrix as Metaphysics', I argued against global skepticism via the claim that even if we're in a Matrix, most of our beliefs are true.
- This argument is often assimilated to Putnam's externalist arguments.
- My view: it has some connections to Putnam's second argument, but it turns on different structuralist considerations.

# Plan

- Discuss Putnam's two externalist arguments, their similarities and differences and diagnose where they fail (spoiler: they require global externalism).
- Discuss the structuralist argument, its similarities and differences with Putnam's arguments, and make the case that it succeeds where these arguments fail.

# Putnam's First Argument

1. If I'm a BIV, 'I am a BIV' is not true.
2. If I'm not a BIV, 'I am a BIV' is not true.
3. So: 'I am a BIV' is not true.
4. So: I am not a BIV.

# Language and Thought

- I'll go back and forth between language and thought.
- To simplify we can suppose a mental language in which 'brain' is the term expressed by 'brain' in English.

# Argument for First Premise

1. BIVs aren't causally connected to brains
2. If BIVs aren't causally connected to brains, BIVs can't refer to brains with 'brain'.
3. If BIVs can't refer to brains with 'brain', then a BIVs 'I am a BIV' is not true.
4. So a BIV's 'I am a BIV' is not true.

# What Does 'BIV' Refer to for a BIV?

- Depending on the correct causal theory of reference, a BIV's term 'brain' either refers either to nothing or to brains-in-the image:
  - e.g. experiences, vat-chemicals, or computer processes
  - brains aren't brains-in-the-image
- Either way: a BIV's 'I am a BIV' is not true.

# A Smallish Hole

- A hole: we may be BLVs inside a simulation.
- Then our brains — both in the vat and the image — are computational, and arguably they share a natural kind *virtual brain*.
- Our term ‘brain’ will refer to virtual brains.
- Then our ‘I am a BLV’ is true (BLVs are causally connected to brains!).
- So I can’t know I’m not a BLV (in a Matrix).

# Weak Conclusion

- Bigger problem: the conclusion is very weak. It merely rules out one global skeptical hypothesis.
- Leaves plenty of others open:
  - I am being globally deceived by an evil demon
  - I am in an artificially designed computer simulation
  - I am in a situation structurally analogous to that of a brain in a vat.



# Generalization?

- Putnam could say: the same goes here!
- Just as BLVs can't refer to brains with 'brain'
  - Those deceived by evil geniuses can't refer to deception with 'deceive'
  - Those in a computer simulation can't refer to computers with 'computer'.
- Externalism holds for 'evil', 'computer', etc.

# Global Externalism

- This reply requires *global externalism*: externalism (and causal theory of reference) for every term/concept in our language.
- Problem: global externalism is implausible, or at least isn't remotely established by the standard arguments for externalism.

# Twin-Earth Experiments

- One can construct Putnam-style Twin Earth experiments for 'water' and 'brain', but it's much harder for many other terms: e.g. 'and', 'zero', 'deceive', 'computer'.
- Arguably: any twin of mine refers to zero with 'zero', to deception with 'deceive' and to computers with 'computer'.
- Same goes for twins in skeptical scenarios!

# Twin-Earthability

- I.e. ‘water’ and ‘brain’ are *Twin-Earthable*, but ‘deceive’ and ‘computer’ are not.
- A term is Twin-Earthable when it can be used nondeferentially by twins with different referents.
- A term as used by an individual is Twin-Earthable when it can be used by a twin with a different referent.

# Social Externalism

- Arguably any term (including ‘deceive’ and ‘computer’) can be subject to Burge-style twin cases involving differences in linguistic community
- But these all involve deferential uses.
- Plausibly there are also nondeferential uses of these terms that aren’t subject to such cases.

# Dialectical Situation

- Putnam's Twin Earth arguments don't establish global externalism: prima facie they don't work for 'plus', 'deceive', etc.
- Burge's arguments don't either: prima facie they don't work for nondeferential uses.
- So the widely accepted arguments for externalism don't establish global externalism.

# Other Arguments?

- Maybe there are other arguments for global externalism?
- Davidson? Putnam against 'magical theories of reference'?
- These arguments are at best obscure and not widely accepted.

# Internalism

- Furthermore, it's fairly plausible that some terms (e.g. 'and', 'deceive', 'computer') aren't not Twin-Earthable.
- We can use these terms (nondeferentially) to formulate global skeptical hypotheses:  
e.g. *I'm in a computer simulation.*
- Absent global externalism, Putnam's argument doesn't refute these hypotheses.



# Recent Envatment?

- N.B. these are (prima facie) global skeptical hypotheses, unlike 'I am a recently envatted brain' (which affects only some beliefs, and so makes a case only for local skepticism).
- The recent envatment hypothesis doesn't undercut the argument against global skepticism, whereas the non-externalist hypotheses seem to.

# Putnam's Second Argument

"By what was just said, when the brain in a vat (in the world where every sentient being is and always was a brain in a vat) thinks 'There is a tree in front of me', his thought does not refer to actual trees. On some theories that we shall discuss it might refer to trees in the image, or to the electronic impulses that cause tree experiences, or to the features of the program that are responsible for those electronic impulses. These theories are not ruled out by what was just said, for there is a close causal connection between the use of the word 'tree' in vat-English and the presence of trees in the image, the presence of electronic impulses of a certain kind, and the presence of certain features in the machine's program. *On these theories the brain is right, not wrong in thinking 'There is a tree in front of me.'* Given what 'tree' refers to in vat-English and what 'in front of' refers to, assuming one of these theories is correct, then the truth conditions for 'There is a tree in front of me' when it occurs in vat-English are simply that a tree in the image be 'in front of' the 'me' in question in the image or, perhaps, that the kind of electronic impulse that normally produces this experience be coming from the automatic machinery, or, perhaps, that the feature of the machinery that is supposed to produce the 'tree in front of one' experience be operating. *And these truth conditions are certainly fulfilled.*" (Reason, Truth, and History, p. 14)

# Argument

If certain causal theories of reference are true:

1. BIVs 'tree' refers to trees-in-the-image.
2. If BIVs 'tree' refers to trees-in-the-image, BIV 'There are trees in front of me' is true.
3. So if I'm a BIV, there are trees in front of me.
4. So even if I can't know I'm not a BIV, this doesn't lead to global skepticism.

# Similarities and Differences

- Like argument 1, this takes an externalist premise, though a much stronger one.
- The upshot is completely different.
  - Argument 1: we know we aren't BIVs.
  - Argument 2: even if we can't know this, this doesn't lead to skepticism.

# Textual Observations

- Putnam doesn't commit to the theories of reference on which the argument depends.
- He doesn't argue for premise 2.
- He doesn't take steps 3 and 4 (he doesn't turn this into an anti-skeptical argument).
- This is the only discussion of the argument in *RTH*.
  - [Question: Does he discuss it elsewhere?]

# Problem I: Crude Causal Theory

- First problem: the case for premise I requires an especially crude causal theory.
- Standard causal theories won't suffice. They say causal connections are necessary for reference, not sufficient!
- Prima facie, 'tree' for a BIV may be like 'phlogiston': it has causes, but not the right sort for it to refer.

# Problem 2: Truth and Charity

- Problem 2: It's not clear how a causal theory of reference will ground premise 2 about truth.
- At least this may need some sort of charity-based principle as well.

# Problem 3: Global Externalism

- The argument seems to require global externalism.
- Suppose 'square', 'object', 'front', 'talk', 'philosopher' are not Twin-Earthable.
- Then the argument won't apply to 'There is a square object in front of me' or 'I am talking with three philosophers'.



# Non-Twin-Earthable Beliefs

- If 'square' is non-Twin-Earthable, BLVs refer to squares (etc) with 'square' (etc).
- So: BLV's 'There is a square in front of me' is true iff there is a square in front of it.
- Prima facie: there is no square object in front of it. If so, the BLV's belief is false.
- If beliefs like this are false, other beliefs like 'there is a tree in front of me' are also false.

# Upshot

- So (assuming global externalism is false): Putnam's hypothesis does nothing to save the truth of our non-Twin-Earthable beliefs under the BIV hypothesis.
- And if these are false, then plausibly most of our Twin-Earthable beliefs are false too.

# Summary So Far

- So: both of Putnam's arguments require an implausible global externalism (and the second requires an implausibly crude causal theory of reference).
- Neither of them succeed.

# The Matrix as Metaphysics

- In ‘The Matrix as Metaphysics’, I argue for a similar conclusion as Putnam’s argument 2 (via an entirely different argument):
- People in a Matrix (and BLVs) have mostly true beliefs; so the Matrix hypothesis doesn’t lead to skepticism.

# Matrix Hypothesis

- Matrix Hypothesis: I'm in a matrix.
- I.e. my cognitive system receives its inputs from and sends its outputs to an artificially designed computer simulation of a world.
- Unlike the BLV hypothesis, the relevant terms are plausibly non-Twin-Earthable.
- Neo refers to matrices with 'matrix'.

# The Argument

- Argument: The Matrix Hypothesis is equivalent to a certain Metaphysical Hypothesis.
- The Metaphysical Hypothesis is not a skeptical hypothesis.
- So the Matrix Hypothesis is not a skeptical hypothesis.

# Metaphysical Hypothesis

- Computational hypothesis: Physical processes are constituted by computational processes
- Plus creation hypothesis (physical processes were created by a designer) and a hypothesis about interaction between physical and cognitive processes.

# Equivalence

- The Metaphysical Hypothesis epistemically entails the Matrix Hypothesis (easy).
- The Matrix Hypothesis epistemically entails the Metaphysical Hypothesis (trickier).
- The latter requires structuralism about computational physics: if physics is computational, the computation can be realized in arbitrary ways.



# Matrix Realism

- If we're in a Matrix, there are still tables and trees: they're just made of bits (constituted by computations).
- In effect, if we're in a Matrix, tables are virtual tables (computational tables).
- The Matrix hypothesis isn't eliminativist about tables, any more than computational physics or quantum mechanics is.

# Nonsemantic Argument

- Unlike Putnam's argument, this argument does not take any general theories of meaning or content as a premise.
- It proceeds from consideration of cases and from a structuralist premise about physics: roughly, what matters for the truth of certain physical theories is causal structure.

# The Role of Externalism

- Externalism comes in later to rebut an objection: e.g. there are no tables in the Matrix computer, so Neo's claim 'That is a table' is false.
- Reply: Neo refers to virtual tables with 'table', so his claim is true. Its content differs from ours (externalism), so the reasoning above doesn't go through.

# Non-Twin-Earthable Beliefs

- What about Non-Twin-Earthable claims, like ‘I am talking with three philosophers’ or ‘I am using a computer’?
- Structuralist line: These have the same content as used in the Matrix, and these contents are true in the Matrix.
- There really are philosophers and computers in the Matrix.

# Internalism

- N.B. the argument is even compatible with a global internalism on which no term is Twin-Earthable.
- We'd just need to say that virtual tables are tables (cf. twin water is water), so that there are tables in the Matrix.
  - Maybe implausible, but that's just to say that externalism is implausible in these cases.)
- So the argument does not rest on externalism.

# Objections from Space

- Biggest worry: Matrix dwellers are getting space wrong (there's a spatial constraint on the correctness of computational physics).
- Argument:
  - (1) 'Cube' is non-Twin-Earthable.
  - (2) There are no cubes in the Matrix.
  - So (3) Neo's belief 'That's a cube' is false.

# Reply

- Reply: spatial concepts are Twin-Earthable.
  - Doubled Earth: twin case for ‘one meter’
  - El Greco World and Lorentz Earth: twin cases for ‘cube’.
- Spatial functionalism: space is what plays the space role.
- Also suggested by practice in physics (e.g. quantum mechanics and string theory).

# Structuralism

- If there is a semantic view associated with the argument, it's some sort of *structuralism*
- Roughly, the truth of our claims turns on the causal structure of the external world (or: causal, mental, mathematical, logical structure).
- This structure is present in the Matrix



# Structuralist Response to Skepticism

1. Our beliefs have structural content.
2. The structure present in nonskeptical scenarios is also present in corresponding Cartesian scenarios.
3. If our beliefs are true in nonskeptical scenarios, they're true in corresponding Cartesian scenarios.

[See *Constructing the World*, excursus 15.]

# Internalism and Externalism

- This structuralism is consistent with externalism, but also with internalism.
- Also, the original argument doesn't assume structuralism, but it emerges as a sort of consequence along the way.

# Relation to Putnam

- Unlike both Putnam's arguments, this argument doesn't require global externalism.
- Unlike argument 1, it works just as well for 'I am in a computer simulation'.
- Unlike argument 2, it doesn't require a crude causal theory.

# Combining with Putnam

- A fan of Putnam's argument 2 could combine it with some of this to rebut the objection from non-Twin-Earthable beliefs
- use spatial functionalism/externalism to show that spatial beliefs aren't a problem
- use structuralism to deal with the rest
- But I think structuralism is doing the work.

# Residual Skeptical Hypotheses

- ‘I am recently in a matrix’: locally skeptical but not globally skeptical.
- ‘I am stimulated by evil genius’: like matrix.
- ‘A deceiver is globally deceiving me’: incoherent.
- ‘There is no external world’: globally skeptical, requires e.g. abduction to rule out

# Local and Global Skepticism

- So: the structuralist reply doesn't refute local skepticism (which may be all the skeptic ever wanted).
- With some help from abduction, it might refute Cartesian arguments for global skepticism.
- Combining with abduction or other replies may also help with local skepticism.

# Conclusion

- Putnam's externalist replies to global skepticism fail, largely because they require an implausible global externalism.
- Structuralist replies to global skepticism do not require global externalism (or any antecedent theory of content) and they may even succeed.