

## IN DEFENSE OF THE TWIN EARTH–THE STAR WARS CONTINUE\*

NENAD MIŠČEVIĆ

University of Maribor, Slovenia, Central European University, Hungary

### ABSTRACT

*The paper discusses the meta-philosophy of thought-experiments, in particular its neglected diachronic aspect, focusing on Putnam's work as the paradigm case, and on the trail(s) that developed out of the Twin Earth thought-experiment. Putnam's experiment is discussed from a perspective that combines metaphilosophy and actual history of analytic philosophy. Peter Unger has branded the whole debate around it as empty and fruitless. A meta-philosophical analysis shows him to be wrong. The experiment as originally proposed already appeals to a broad variation of examples and intuitive induction; the variation continues in other works addressing the issue, and produces interesting results. The second aspect is the search for reflective equilibrium, lasting till the present day. The internal logic of these processes is discussed, in order to show that the accusation for emptiness turns against Unger himself. In general, debates around thought-experiments, the already famous and also around new ones, make a large part of contemporary analytic philosophy. The way to understand a large part of this, and of debates surrounding it, is to link it to the internal understanding of a typical thought-experiment; stages of a particular experiment get discussed, developed and changed in the history of a particular trail produced by it. This is an important way in which a philosophical tradition is born, and we need to combine synchronic and diachronic approaches in order to understand it.*

**Keywords:** *thought-experiments, Twin Earth, meta-philosophy*

## 1. Introduction

Putnam is one of the most successful thought-experimenters in twentieth-century philosophy; he has produced two great thought experiments, Twin Earth one and the Brain-in-a-vat one, and many related smaller ones. The two great ones have marked analytic philosophy of our time, from philosophy of language through philosophy of mind to philosophy of science and general metaphysics. Here we shall look at his Twin Earth thought experiment (Twin Earth TE for short), anticipated by Putnam in his 1970 paper „Is semantics possible?“ and formulated in 1975 in „The Meaning of ‘Meaning’“, and at a recent criticism of it, due to Peter Unger (2014). We shall briefly defend Putnam from the criticism, and concentrate on Unger’s complaint about endless “spates of papers” generated by TEs in recent analytic philosophy. We shall argue that the prolonged discussions of fundamental thought experiments (TEs) in various (sub-) disciplines of philosophy define big chunks of the history of analytic philosophy (not to mention three centuries of debates over Social contract TEs and Evil Demon TE). The meta-philosophy and epistemology of TEs has to take this phenomenon into account, thus connecting history of philosophy with meta-philosophical theorizing. We shall argue for this suggestion in connection with Twin Earth TE, and the ensuing debate.

Let me just mention that recently the publication of such discussions of TE is becoming systematic. Andrew Pessin and Sanford Goldberg have collected important papers on Twin Earth TE in their *The Twin earth chronicles* from 1996 (see References). Two years ago Cambridge University Press has started a series, Classic Philosophical Arguments, pretty much concentrated on TEs, like *The Brain in a vat* (Goldberg 2016), *The Prisoner’s dilemma* (Peterson 2015) and *The Original position* (Hinton 2015). The meta-philosophy (or epistemology) of TEs has to join on the theoretical side, connecting history of philosophy with meta-philosophical theorizing.

Here is then the preview. Section II focuses on Twin Earth thought experiment, starting with a short reminder, and continuing with a brief proposal of how to distinguish stages in thought-experimenting, in particular within the Twin Earth example(s). Section III assesses Unger’s criticism (from his *Empty Ideas*, (Unger 2014)) and then turns to the defense of Twin Earth TE. Section IV continues the discussion with Unger by turning to the trails-traditions that have sprung out Twin Earth TE, and connecting the trails to various aspects of the experiment itself. Thus the short-term and long-term patterns of Putnam’s thought-experimental reasoning and its continuation are brought together. Section VI attempts to generalize the morals of the defense beyond Twin Earth cases, first mentioning Putnam’s other famous TE, the Brain-in-the-Vat, and then moving very briefly to other examples central in the history of

philosophy. It then summarizes our answer to one of the central questions of the paper: why go historical in the meta-philosophy of TEs, rather than stay with the usual style of topical discussion?

## 2. Reminder: The Twin Earth TE and its stages

Let me start with a general characterization of thought experiments. So, what is a TE? It is an investigative procedure „in the armchair“, which normally involves

0. the formulation of experimental design pointing to
  - i) the determination of the goal(s), in particular the thesis/theory to be tested, and
  - ii) the construction of a (typically) counterfactual scenario to be considered
1. the presentation of the scenario thus constructed to the experimental subject (either the author of the scenario herself, or an interlocutor),<sup>1</sup>
2. understanding done by the experimental subject
3. the (typically imaginative) contemplation of the scenario and some piece of reasoning,
4. the decision („intuition“) concerning the thesis/theory to be tested, and then the variations and generalizations from the result.

Once this result is achieved, it can be, and often is compared with results of other thought experiments in the vicinity. We shall see that the characterization captures the TE we are interested in.

In his “The Meaning of ‘Meaning’” Putnam famously introduced his Twin Earth TE as a „science-fiction example” (!):

For the purpose of the following science-fiction examples, we shall suppose that somewhere in the galaxy there is a planet we shall call ‘Twin Earth’. Twin Earth is much like Earth; in fact, people on Twin Earth even speak English. In fact, apart from the differences we shall specify in our science-fiction examples, the reader may suppose that Twin Earth is exactly like 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 in our science-fiction examples, the reader may suppose that Twin Earth is exactly like Earth.

---

<sup>1</sup> In a philosophy seminar the design of the scenario is usually given through a text formulated in a paper. This is what is colloquially then called „TE“.

He may even suppose that he has a *Doppelgänger* - an identical copy – on Twin Earth, if he wishes, although my stories will not depend on this. (Putnam 1975, 223)

The next move can be described as separating chemical composition of the liquid known as “water” from its stereotype.

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. (Ibid.)

Let me note that I have asked Putnam in conversation whether we can use Kant’s idea of “separating” in “experiments with our mind” of elements that normally go together to characterize what he is doing in Twin Earth scenario, and he very emphatically agreed.

Next comes the question for the reader: when a Twin Earthling uses the word “water” does he refer to water? So, after the stage of presentation and question raising, we presumably have reader’s understanding. The reader is expected to imagine the situation, to build a model, so to speak. The next comes the reader’s intuitive answer. The expected answer is negative: the Twin Earthling refers to what is in fact XYZ, and it is not water!

Let me mention some obvious problems with the performance of the TE at these early stages. First, consider the construction of a (typically) counterfactual scenario to be considered, and its presentation at the stage one. The proponent, say Putnam, is testing people’s intuitions, and hopefully, the non-specialists will be included, since he wants ordinary intuitions from non-experts (as well as from his colleagues, the experts). So, the story has to be *relatively short and not too technical*; the best would be if it were not technical at all. This is the first sore point: any expert will be able to come with some counterexample to the short story. Note the parallel with “real” experiments: typically a single experiment concerns just one variable, and taken in isolation proves very little. What is being done is rich controlled variation; real laboratory is in this respect very similar to the laboratory of the mind.

One feature of experiments closely connected to learning is their variability: scientists perform them over and over again with modifications which may be systematic and intentional or intuitive and exploratory. Variation is one of the key factors in the success of experiment. Frequency of variation in Faraday’s experiments is apparent from his laboratory notebooks: these records suggest a lot of unrecorded and unpremeditated variation. (James 1989, 67)

The construction and the presentation culminate with a question. We might assume that the main question in the original version of the TE concerns reference: does the word “water” in the Twin Earth language refer to water or not?

Next, the understanding. It is happening within the conversation, or at least the reading. Strong pragmatic considerations might intervene. First, the motivational ones. The recipient might wonder why she is being asked a particular question, what are the expectations, and so on. Next comes the understanding of the content received; no non-technical, ordinary piece of discourse is free of potential ambiguity, and we know these days that ambiguities often do show their teeth.

Unfortunately, philosophers sometimes suggest more than is needed for mere understanding. Putnam does it. And here is the simple illustration, in the passage we just quoted. “I shall suppose that the oceans and lakes and seas of Twin Earth contain XYZ and not water...”, he writes at the very beginning of his presentation of Twin Earth. But does this not assume that XYZ is not water? Does this not dictate the intuition that the reader is suppose to contribute himself? What if the reader has the opposite feeling: if it looks like water, tastes like water, and so on, than it is water; I don’t care for chemistry! Too bad, the TE was supposed to test intuitions, not to suggest them!

I shall assume that the recipient next builds her model of the situation; let it be the stage Three of the initial TE. This stage is not controllable by the proponent any more, and the model-building depends a lot on specific skills, memories and interest of the recipient. But Putnam does suggest the answer, although he should not be doing this. Let me assume that the recipient goes through some unconscious reasoning; we shall not need it here. So much about the problems.

The stage Four is then the rise of intuition concerning the thesis/theory to be tested.: “It seems to me, very obviously, that the Twin Earthling does not refer to water”.

Normally, the recipient is expected to generalize: is what is valid for water also valid for gold, silver, and so on? The process has been called ‘intuitive induction’ by Roderick Chisholm (1977, *Theory of Knowledge*, ch. 4, section ‘Intuitive induction’), who borrowed the term from W.D. Ross. Anyway, the intuitive induction ends with a general judgment; in this case “what is valid for water should be valid for gold, silver, and so on.” I would add some obvious and immediate consequences, for instance: it is the deep structure of the matter that ultimately determines reference for typical mass terms.

Again, this stage, our number Five, is rife with problems. Assume that the intuition is valid for gold; what about “wine” or “brandy”? Is it really the chemical composition that counts, rather than drinkability, taste and the effects on the consumer?

What normally follows is theorizing. The new generalization (and its immediate consequences) should confront the rest of our theoretical commitments, in this case ranging from the ones in chemistry to the ones in semantics. And it should explain the intuitions discovered in the course

of TEs. Let me borrow the famous term from Rawls, “reflective equilibrium” and call this, sixth stage, the search for reflective equilibrium. Putnam did find one, and a quite radical one, for that matter, namely a restructuring of the whole of semantics. This is the final stage, number seven, the reflective equilibrium itself. We shall return to it later.

So much about the first (version of) Twin Earth TE. We noted the existence of challenging points at each stage. If you are pessimistic, you can call them “sensitive points” of a TE, if you are optimistic, you can describe them as “productive points”, since they obviously invite questions and discussion.

Why “productive”? Consider how the experimenting proceeds further, in Putnam’s text. After coming to the conclusion, one might ask a slightly different question, introducing history. Perhaps our intuitions concerning water are really dictated by Lavoisier’s discovery of its chemical composition. We need to introduce a more dramatic variation.

So, we arrive at Version Two: this time by introducing history of chemistry. Stages One and Two are re-done: Imagine Oscar<sub>1</sub> and Oscar<sub>2</sub> living in 1750. Did each understand the term 'water' differently from the other? Again, the intuition seems to suggest that they did: the first referred to what is in fact H<sub>2</sub>O, the second to what is in fact XYZ. So, Oscar<sub>1</sub> and Oscar<sub>2</sub> understood the term 'water' differently in 1750 although they were in the same psychological state, and although, given the state of science at the time, it would have taken their scientific communities about fifty years to discover that they understood the term 'water' differently. We perform some intuitive induction and end up with the general judgment, valid also for gold, and the like.

But what about the kinds of stuff that are both present around, perceptually indistinguishable, and strongly associated with some familiar product; or at least one of them is. This suggests further variation.

Version Three: “we will now suppose that molybdenum is as common on Twin Earth as aluminum is on Earth, and that aluminum is as rare on Twin Earth as molybdenum is on Earth. In particular, we shall assume that 'aluminum' pots and pans are made of molybdenum on Twin Earth. Finally, we shall assume that the words 'aluminum' and 'molybdenum' are switched on Twin Earth: 'aluminum' is the name of molybdenum and 'molybdenum' is the name of aluminum.”. (Putnam 1975, 225)

Version Four: the final question concerns non-exotic, not rare specimens, and moves from science fiction to ordinary examples and from mass terms to kind-terms in botanics and the like; this is the elm-beech example (Putnam 1975, 226).

We arrive now to the general conclusion: meanings can vary with psychological states remaining constant explanation: meaning depends on

external causal connections.

We claim that it is possible for two speakers to be in exactly the same psychological state (in the narrow sense), even though the extension of the term A in the idiolect of the one is different from the extension of the term A in the idiolect of the other. (Putnam 1975, 222)

How is this possible? We need explanation, and it will appeal to causal theory and division of linguistic labor. Further steps lead to the final reflective equilibrium, strongly revisionary of traditional Fregean semantics (for a historical overview see Floyd 2005, in References). Here are then the stages.

STAGES	EVENTS
Stage 0 - design	
Stage 1 - the question	The scenario presented and the question: To what does the Twin Earth speaker refer with her word “water”?
Stage 2 - understanding	The reader understand the scenario and the question.
Stage 3 - tentative conscious production: tries to imagine the arrangement and does it to one’s satisfaction.	Imagining the other planet.
Stage 3a - possible non-conscious elaboration	Possible problems: what is my body made of there? Is such a scenario possible at all?
Stage 4 - Intuition: immediate spontaneous answer	<i>The Twin Earth speaker refers to XYZ, not to water.</i>
Stage 5 - Intuitive induction: varying and generalizing	Variations: First, other materials. Next: History of chemistry, Aluminum-molybdenum, Elm-beech
Stage 6 - general belief	Extensions can differ, although the speakers are in the same psychological state.
Stage 6a - explanation	“Meanings are not in the head”, causal theory, division of linguistic labor
Stage 7 - search for reflective equilibrium	Discussion of description theory in semantics, criticism of it
Stage 7a - reflective equilibrium achieved	Keep the general belief and the explanation, and restructure semantics.

This concludes the short reminder of the original TE(s). Very soon after the original presentation, several philosophers, and Tyler Burge in particular in 1979 paper “Individualism and the mental”, famously noted that the story about linguistic meaning can be extended to mental content as well. Burge developed it into an externalist theory of mental content. Again, we have a variation, but rather dramatic one, primarily at the level of immediate conclusions from Twin Earth and similar TEs, and perhaps even deeper, at the level of initial question. The reader is presented with variations Oscar<sub>1</sub> and Oscar<sub>2</sub> scenario, and now, in the ‘laboratory of her mind she has to test a bit different assumption: is the content of Oscar<sub>1</sub>’s mental state when thinking of the relevant metal the same as Oscar<sub>2</sub>’s or not? Burge famously answered in the negative. Here is Putnam’s comment on Burge from 1996: “...I have come to believe that he is right” and “...I agree with his paper” (Goldberg 1996, xxi).

Once the new question was around, a new round of reactions followed, represented by ten famous paper reproduced in (Goldberg 1996) mostly from the eighties, when the action was at its peak. Searle (1983) insisted that the difference between contents is internal, due to the nature of intentionality, and reflexive functioning of the indexical(s) allegedly involved (“this metal”). Fodor (1991) proposed two kinds of content, and argued for the primacy of the internal, “narrow“, one. Dennett came up with the idea of a “notional content” (see his contribution in Goldberg 1996) and Jackson (in Goldberg 1996) proposed a stable, compromise solution, that later developed into “two-dimensionalism”. The difference in the question asked, or at least the difference in the weight assigned to the mental in contrast to the linguistic, resulted in differences in intuitions, then in generalizations arrived at, and then projected deep into the area of explanation and reflective equilibrium (see Goldberg 1996) for a nice collection of historically crucial papers). So, hopefully, the attention to stages might help us understand the further history of the Twin Earth TE.

But why care about such matters? Why go historical in the meta-philosophy of TEs, rather than stay with the usual style of topical discussion? We shall say more about the process after we take a look at the recent criticism of Putnam and of methodology of analytic philosophy in general.



### 3. Empty thoughts: Peter Unger's criticism of Putnam and analytic methodology

Let me start by quoting Peter Unger's interview in which he nicely summarizes his view of contemporary philosophy:

Philosophers easily get the idea that somehow or other, just by considering things about the world that they already know, they can write up deep stories which are true, or pretty nearly true, about how it is with the world. By that I especially mean the world of things that includes themselves, and everything that's spatio-temporally related to them, or anything that has a causal effect on anything else, and so on. They think they can tell a deep story about how it is that all of this stuff really hangs together, that's much deeper, more enlightening and more comprehensive than anything that any scientist can do.

And so philosophers proceed to write up these stories, and they're under the impression that they're saying something new and interesting about how it is about the world, when in fact this is all an illusion. To say new and interesting things about the world — and that's very hard, things of any generality I mean, or even anything interesting — you really have to engage with a lot of science. And very few philosophers do any of that, at least in any relevant way. (Boey and Unger 2014)

All this, Unger thinks, applies to Putnam, to whom several sections, and the whole chapter three of his book *Empty Ideas: A Critique of Analytic Philosophy*, titled Earth, Twin Earth and History, are dedicated.

His first criticism is simple: there is too much material, just spates of articles. He said he entered “twin earth”, coming up with 1,941 items, comprising 1,571 articles and 261 reviews, and the remainder in other categories. Is it good or bad? Depends on the quality of the articles. And here comes his second criticism: “...a fair lot of much of this remarkably empty literature” (Ibid.)

And Unger comes, surprisingly, with an alternative proposal, a new Twin Earth thought experiment. Here is just the barest sketch. He invites us to imagine someone like Lavoisier, „on the verge of discovering that earthly water is composed of hydrogen and oxygen.” At the same time, the twin Lavoisier is on the verge of discovering the chemical composition of twin water. „In all their real mental powers, our two Joseph Antoinies will be precisely similar, however we sensibly construe “mental powers”. not only as regards what behavior each is apt to produce (in any encountered environment) but also as concerns what experiencing each is apt to enjoy, each of our two chemists is, at our start, precisely similar to the other.” (2014: 64). “.../N/ow, suppose each ... of them „ to switch places with the other, almost instantaneously. with this switch, twinchemist will

be here on earth, in 1750". Of course, each proceeds with his discovery, only that it is now the twinchemist who discovers the composition of water. So, what is the big deal, Unger asks. "The capacity to „think of water" seems not to be needed for the job." Having beliefs about water appears irrelevant to being able to discover water's composition. Our belief-states have full general propensities, they react in the same way to water and twin water, Unger claims. Therefore, Putnam's line is irrelevant!!! Unger than offers a very complicated and sophisticated TE with elementary particles, to introduce what he sees as the relevant alternative scenario. We shall not reproduce it here.

Let us now pass to a brief defense of Twin Earth. First, Unger's own methodology suggests the indispensability of TEs: even if you see analytic philosophy as a bunch of empty ideas, you will have to propose TEs of your own. This is a deep inconsistency in what Unger is doing. Alternatively, if he had looked at alternatives outside of analytic philosophy and had turned to recent continental authors, he would, with his criteria of clarity, probably see it as passing from contradiction to contradiction; not a very attractive option.<sup>2</sup>

Unger sees the biggest part of analytic philosophy as being empty of concrete content, and proposes an alternative relying on extremely sophisticated scientific knowledge as the background for analytic philosopher. Emptiness is for him tied to the endless production of tons of empty material. So, let me briefly address his objection, and then concentrate upon the first one.

---

<sup>2</sup> Think of famous continental thinkers, writing in the second half of 20th century. Heidegger suggests that philosophy should become poetic. Here is a typical quote on human condition:

But what is it that touches us directly out of the widest orbit? What is it that remains blocked off, withdrawn from us by ourselves in our ordinary willing to objectify the world? It is the other draft: Death. Death is what touches mortals in their nature, and so sets them on their way to the other side of life, and so into the whole of the pure draft. Death thus gathers into the whole of what is already posited, into the posited of the whole draft. As this gathering of positing, death is the laying-down, the Law, just as the mountain chain is the gathering of the mountains into the whole of its cabin. (Heidegger 1971, 123).

Somebody like Unger, who finds Putnam not rigorous enough, would be probably shocked by the very idea that this claim of Heidegger is a piece of philosophy. Similarly with Derrida. Here is how Derrida expresses his view that a non-figurative treatment of metaphor is impossible:

I am obliged to speak of [metaphor] *more metaphorico*, to it in its own manner. I cannot treat it (entraîner) without dealing with it (sans traîner avec elle) ... I do not succeed in producing a treatise (une traite) on metaphor which is not treated with (traite avec) metaphor which suddenly appears intractable (intraitable). (Derrida 1998, 102–3).

For more examples and a systematic discussion see my [https://www.academia.edu/1828410/PHILOSOPHIZING\\_WITHOUT\\_ARGUMENT](https://www.academia.edu/1828410/PHILOSOPHIZING_WITHOUT_ARGUMENT)

The sources of potential philosophical knowledge are not mysterious as Unger sees them. Remember: “Philosophers easily get the idea that somehow or other, just by considering things about the world that they already know, they can write up deep stories which are true, or pretty nearly true, about how it is with the world.” This simply doesn’t fit the Twin Earth TE. The semantic part is about us. We might be unaware of the structure of our semantic habits – the reflection on cases can make us aware of it. As far as the non-human world is concerned, the info in the theory of Twin Earth TE comes from science, chemistry etc. and is then brought together with our understanding of meanings.

This is, it seems to me, quite a general pattern with famous TEs in descriptive-explanatory theoretical philosophy. For instance, Jackson’s Black and White Mary makes sense only on the background of quite developed neurological picture of human mind, and questions one central ambition of it. The same with Putnam’s Brain-in-a-vat TE. One element that made it interesting was his replacing of the Cartesian thinker, ego, or whatever, with brain, and his connecting this to the up-to-date debates concerning Turing machines and Skolem-Lowenheim theorem (in chapter one of *Reason, Truth and History*; see the paper by Danilo Šuster in this volume). I think Tim Williamson is right about the abuse of the epithet “empty” by Unger:

Unger’s use of the term “empty” is just an advertising trick. It’s like a competitor who defines “empty” as “containing nothing but brand X fruit juice” and then puts up posters warning that cartons of brand X fruit juice are empty. To read Empty Ideas, one must get through the equivalent of numerous elaborate descriptions of cartons of brand X fruit juice of various types, each concluding that the carton was empty, and for contrast some elaborate descriptions of cartons of brand Y fruit juice of various other types, each concluding that the carton was full. The reader’s task is made no easier by Unger’s loquacious, attention-seeking prose. (Williamson 2015, 22-23).

We should then turn to meta-philosophy and history of philosophy. What about Unger’s first criticism? Does the spate of articles make any sense? Obviously it does for Unger, for he tries to add to it another version of Twin Earth TE. At the same time, it goes on his nerves. Here is a therapy: let us try to understand where the spates are coming from.

#### 4. Stages and traditions - from the synchronic to the diachronic

Unger is obviously right about TEs generating spates of articles, full of comments, criticisms and alternative proposals. Where do these spates come from? How does a spate begin? In order to answer the questions we have to dig deeper into the matter, and ultimately, we need some meta-philosophy explaining the recent history of analytic philosophy. First, we can note that spates come in trails. Unger himself notes the following:

„Just a few years later and largely thanks to their contemplating Putnam’s Twin Earth scenarios, several other philosophers did endorse this further idea, concerning who thinks what, the two most timely being, perhaps, Tyler Burge and Colin McGinn” (Unger 2014 ,77- 60).

But the spate of articles that come out of Putnam’s and Burge’s efforts constitutes a trail, the externalist one. We can even call it a tradition. There is the contrary trail-tradition, represented by more internalistically oriented authors, from Searle to Jackson.

The first thing to note is that the trails start as reactions to particular features of the original Twin Earth TE, and the second, and a surprising one, that reactions target particular stages of the TE, particular sore-fruitful spots which one can use as starting points from one’s own proposals. Some authors, perhaps most famously Burge, widen the range of examples (from water to arthritis, and from true beliefs to systematically false ones), thus targeting the initial design of the TE and changing a bit the question asked, others stay with original example, but, at the stage of intuition-generation suggest a different intuition: some content of Twin Earthling’s thought is identical to the content of Earthling’s thought. The new intuition demands new explanation, and commands a different process of intuitive induction.

In other words, *there is a connection between the synchronic structure of stages and the diachronic process of generation of trails, even traditions.* Let us take a closer look at the externalist and internalist trails.

Start with Tyler Burge. Here is his question, stage 1: Does the ignorant person who believes that he has arthritis in his thighs refer to arthritis? Yes, the intuition suggests (at stages 4-5). The same for sofa and contract. We can see his initial proposal as a variation of the crucial example, replacing water with items like sofa, arthritis or contract. This changes the TE already in its initial stages. Now, intuitive induction, our stage 6: the contents of the thought of ignorant persons all depend on the public meaning of the terms involved. The very content of thought is determined externally; a strong externalism about content, not just about linguistic meaning, follows. Of course, a lot of work is to be done at the next stages to relate the view to other accepted ideas in philosophy of mind. Burge and his followers will be adamant about the externalist intuition, its

externalist explanation, and the general externalist stance; all else can be changed, but not these results. Some kind of reflective equilibrium follows (of course, it can be challenged by the internalist, who might point out that on this account we don't know contents of our thoughts, as Boghossian famously did-see References).

Incredibly enough, the trail generated by the externalist proposal will be crossing another trail, the Demon TE that comes from a Cartesian tradition, and the fans of the trail will try to subvert the results from the Cartesian one. Consider a brain-in-a-vat: does it have false thoughts about the external world, a Cartesian thinker will ask. Well, does it have thoughts about the external world at all? The externalist line suggests a negative answer. But, if it has no thoughts about the external world, it does not have any false thoughts about the external world; skepticism is inconsistent. No wonder, there is a recent book bringing together the papers from the crossing, Sanford C. Goldberg's 2016 collection *The Brain in a Vat*. A discussion of this crossing of the two traditions would demand a separate paper. However, mere pointing out the phenomenon of crossing is all we can do here.

To conclude, the history of semantic externalism can be reconstructed to a large extent as history of the Twin Earth TE(s) debate. Attention to the internal structure of the TE(s), especially to the sore-fruitful spots at each stage of the TE that prompt questioning and the debate make see the discussion and the 'spate of articles' produced as rational response to the crucial issue involved. The elementary reconstruction of the TE(s) in terms of stages, and of problems connected with the performance of the experiment combine with diachronic rational reconstruction of the history of externalism as a whole and make us understand the important trail of contemporary analytic semantics and theory of mental content in rational and philosophically relevant terms.

Let me now pass to the opposite, internalist trail, extremely popular, with a lot of followers. First, a moderate proposal. Searle insisted that the difference between contents is internal, due to the nature of intentionality, and reflexive functioning of the indexical(s) allegedly involved ("this metal") (Searle 1983, 206). Next, consider a more radical proposal, involving a deeper analysis of earthling vs. Twin earthling thoughts. There is a kind of mental content, narrow content, that is the same across Earth-Twin Earth contrast. Just consider causal powers of the thoughts: „being a water thinker is the same causal power as being a twater thinker, only instantiated in a person with a different causal history.” (Fodor 1991, 5-26, reprinted in Pessin and Goldberg (eds.) 1996, 275 ff). His further conclusion is that the narrow content is the right content for psychology (and psychological generalizations), the view known as methodological solipsism. The final reflective equilibrium reached is completely opposite from Burge's and later Putnam's, and it has had a lot of followers.

So, we may look at the longer term history of different stages within the history of a TE or its variants. It starts with initial experiment, including variation, general belief-conclusion, explanation, and early reflective equilibrium, and can for example proceed to a new question. Is the linguistic meaning determined by external factors is supplemented with a more dramatic question: isn't the mental state itself then determined by external factors? There will be a new example (arthritis in the thigh), or a new reading of the old one (or one can accept two kinds of content, like Fodor, and Jackson). The new process of 'intuitive induction' will lead to a new conclusion, say that one extended to mental content that now varies with the surrounding. Then we need a new explanation: causal theory is widened to the mental content, and the division of linguistic labor has internal psychological consequences. This is a new reflective equilibrium, different from the earlier one(s). With some luck a new trail is created, which can become a real new tradition. And this can be repeated many times, passing, for instance to analysis of ethical discourse, and proposing Moral Twin Earth TE (see T. Horgan and M. Timmons 1992). But, there is space for other kinds of variation, for instance, including information about scientific treatment of kinds (chemistry, biology, etc.), what Unger would describe as Substantial Scientiphicalism. Or, one can proceed to psychological questioning as has been done in experimental philosophy (see Haukioja 2015 for some discussion of Putnam).

What is crucial of us is the link between short-term and long-term pattern of (thought-) experimental reasoning. This has not been noticed in the literature. Let me summarize what has been said in this section in the following table:

<b>STAGES</b>	<b>VARIATION AND NEW TRAILS</b>	<b>EXAMPLES</b>
Stage 0 - design	Accommodating data from stages 6, 7, 8	Redesigning the TE Tyler Burge: externalist proposal Jackson: varying which world is actual Fodor: methodological solipsism
Stage 1 - the question	New question: reformulation  or  complete transformation	Widening: From language to the mind  or  From mind to ethics: moral Twin Earth

Stage 2 - understanding	(adding explanatory material)	
Stage 3 - tentative conscious production: tries to imagine the arrangement and does it to one's satisfaction.  Stage 4 - possible non-conscious elaboration  Stage 5 - Intuition: immediate spontaneous answer	Questioning intuition, suggesting a different judgment	Jackson: a very different intuition XYZ is water in one sense, is not water in another  Fodor: We and Twin Earthlings share one content (narrow), but not the other (wide)
Stage 6 - Intuitive induction: varying and generalizing	Conservative:  Widening the range  Revolutionary:  Problematizing the original examples  ????	Jackson and Fodor both widen the range enormously, so as to cover all Putnam's cases (and perhaps more)
Stage 7 - general belief    7a explanation	A new general belief G: Conservative extension Revision ???  A new explanation	A new general belief G*: Jackson: Two meanings Fodor: 2 kinds of content methodological solipsism
Stage 8 - search for reflective equilibrium	Traditionalist: revise G and keep the inherited beliefs  revolutionary: keep G and change the inherited beliefs	
Stage 8a - reflective equilibrium achieved	The new theory: conservative revolutionary	

But why do philosophers go through all these efforts. One answer is that TEs are indispensable. Philosophy does not use laboratory to test its theories; the only experiments available here are those in thought. TEs play in philosophy the crucial role that laboratory experiments play in science. Philosophers are vitally interested in *connections between our spontaneous understanding of important items, like meaning and content of our thoughts, and the results of science*. In order to answer the question about the relation between, say, cognitive science-cum-neurology and our feeling of having contentful thought, we need to confront the two, and we cannot do it within science alone. We need the bridge, and a TE is a perfect candidate. TEs are our laboratory, and philosophers return to their experiments, as scientists do to theirs.

We now understand why one should go historical in the meta-philosophy of TEs, rather than stay exclusively with the usual style of topical discussion. The spates of articles, lamented by Unger, are trails, promising to become traditions, and mark the present-day analytic philosophy, as their ancestors marked the early twentieth century efforts.

## **5. Trails-traditions: towards a meta-philosophy of analytic efforts**

It is time to generalize, and draw the morals from our discussion of Putnam's TEs. We obviously need a historically informed meta-philosophy of analytic efforts, in this case ones connected to thought-experimenting, and combining the awareness of the inner structure of a TE with insights into historical, diachronic process of generation of trails leading from various components of a TE to further and further developments. Some famous examples, much older than the analytic tradition, like Plato's TE of ideal state, and Descartes's Evil Demon TE present the same trail-forming tendency.

Here are the elements we were able to identify. First, the indispensability of thought-experimenting. There is often no other way to address central philosophical questions. So, there is no wonder that challenges TEs are revisited time and again

Second, performing a TE involves asking a question and then letting the subject decide; like in laboratory TEs, the designer of the experiment is not the absolute master. There are performance challenges, issuing in hot spots: the reader responds differently from the expectations of the designer. So, the designer has to go back to the drawing board, and vary the initial proposed scenario. Or, a colleague with different background opinions might find the intended answer incredible. Then she has to propose an alternative scenario that will vindicate her intuitions. This produces endless variation in scenarios and questions and production of new intuitions and immediate generalizations.



Next, the history sets in, with long term development of these generalizations, attempts to explain them, and new candidates for reflective equilibrium, from very conservative to very radical ones. Trails are being born, that, if there is luck enough, turn into solid tradition.

Note that laboratory experiments are discussed and varied in the same way as TEs. Take Edison's discovery that direct current can do marvelous things. But then comes Tesla: "Yes, boss, but alternating current can do many more, and much more marvelous ones." What followed was "War of Currents" in the 1880s over whose electrical system would power the world — Tesla's alternating-current (AC) system or Edison's rival direct-current (DC) electric power.

So with laboratory experiments we have the same pattern: first, the initial experiment, followed by small, routine variations, then, in case of really important ones, a spectacular variation. Scientists discuss the prospects of both, and propose further variations, if needed. Remember the quote from F. James who stresses the variability of experiments, and compare his diagnosis to what happened with Twin Earth TE.

Back to the diachronic developments in a long-term life of a TE. Let me make a first step towards generalizing the morals of it and offer a new proposal for the understanding of the history of debates around important TEs that make up a significant part of analytic philosophy. Let me again mention two to three famous TEs: Evil demon-Brain in a vat TE and the Social contract. These have engendered long term traditions, each lasting around three hundred years, if we count modern philosophy only, and set aside the Greeks and Saint Augustine. If we could bring together short-term and long-term stages, this would make the historical process more intelligible. For instance, take the relatively recent variations of the two mentioned thought-experimental oldies. The Evil Demon has been innovatively transformed into (or replaced with) the Brain-in-a vat (see Goldberg,2016), and Social contract has generated the Original position proposal (see Hinton, 2015) and its relatives, due to Scanlon and Habermas.

The patterns seems recognizable. The tradition starts with an initial TE, and proceeds through three possible kinds of reactions. First, a new variant of the TE or a new, but related TE is proposed. In the Social Contract tradition, one can change the characterization of the parties involved (males only, females as well, what about mentally challenged persons, and so on). The dramatic example of the latter is Brain-in-a-vat: replace the thinking person from the original Cartesian scenario with a brain wired to a control panel.

Or, there can be a non-thought experimental reaction to the initial TE: think of the numerous objections to Descartes First Meditation, collected in *Meditations with Objections and replies*, or of the "negative program" within experimental philosophy, or, concerning Original position and its

consequences, the criticisms of B. Barry (1973).

Finally, we can have both, criticisms plus new TE, the way we just saw in Unger's reaction to Putnam. Once we have the big picture, we understand the strange reaction from Unger: on the one hand criticizing the spate of papers discussing Twin Earth TE as empty, meaningless, and over-complicated, and on the other, adding to the spate, by proposing further, related, but much more complicated TEs. On the one hand, a philosopher is tempted to a negative reaction when one notices the sheer length of a famous TE trail. On the other, the only legitimate way to criticize is to add to the trail!

The history is in some respect analogous to short term reflection and discussion of such a TE, say in a seminar or a conference. Stages of understanding and debating normally structure the short term life of the TE. Longer lasting attempts (with variations, often fundamental) are part of the further, historical story. So, we can integrate the view of history of (analytic) philosophy with meta-philosophy of TEs, and then, we shall find Unger-like disappointments and criticisms less persuasive, and less worrisome. Twin water has become part of the philosophical ecology, and has a right to stay with us. It is part of our life as philosophers. In the case of science, understanding the history of science, we need the understanding of the longer history, of the laboratory experimental tradition tied to some important initial experiment. The same holds for TEs in philosophy and science: we need the understanding of the longer history of each given TE in the history of philosophy. We want to know what is the relation between such historical changes, and the usual micro-variations of a given TE. Once we have worked out the answer we shall be able to integrate the meta-philosophy of TEs with information from history of philosophy. Here, we have the first steps. And Putnam's work remains as the paradigmatic example of depth and fruitfulness of thought-experimenting in philosophy. The Star Wars continue.

## REFERENCES

- Barry B. 1973. *The Liberal Theory of Justice: A Critical Examination of the Principal Doctrines in A Theory of Justice by John Rawls*. London: Oxford University Press.
- Boey, G and Unger, P. 2014. *Philosophy is a bunch of empty ideas: interview with Peter Unger*, Monday, June 16, available at (<http://www.3quarksdaily.com/3quarksdaily/2014/06/philosophy-is-a-bunch-of-empty-ideas-interview-with-peter-unger.html>)

- Boghossian, P. 1998. What the Externalist Can Know A Priori. In Wright, Smith, and Macdonald (eds.), 271–284.
- Burge, T. 1996. Other bodies. Repr. in: Pessin and Goldberg, 142-160.
- Derrida, J. 1998. The Retrait of Metaphor. In Wolfreys (ed.).
- Descartes, R. 2006. *Meditations, Objections, and Replies*, edited and translated by Ariew, R. and Cress, D., Indianapolis: Hackett Publishing Company, Inc.
- Floyd, J. 2005. Putnam’s “The Meaning of ‘Meaning’: Externalism in Historical Context” in *Hilary Putnam*, ed. Y. Ben-Menahem, 17-52. Cambridge: Cambridge University Press.
- Fodor, J. A. 1991. A modal argument for narrow content, *The Journal of Philosophy*, Vol. 88, No. 1 (Jan.), 5-26.
- Goldberg, S. C., ed. 2016. *The Brain in a Vat*. Cambridge: Cambridge University Press.
- Haukioja, J., ed. 2015. *Advances in Experimental Philosophy of Language*. London: Bloomsbury Publishing.
- Heidegger, M. 1971. *Poetry, language, thought*. Harper & Row.
- Hinton, T., ed. 2015. *The Original Position*. Cambridge: Cambridge University Press.
- Jackson, F. and Pettit, Ph. 1996. Functionalism and Broad Content. Repr. Pessin and Goldberg, 219-237.
- Miščević, N. *Philosophizing Without Argument*. Available at: [https://www.academia.edu/1828410/PHILOSOPHIZING\\_WITHOUT\\_ARGUMENT](https://www.academia.edu/1828410/PHILOSOPHIZING_WITHOUT_ARGUMENT)
- Pessin, A. and S. Goldberg, eds. 1996. *The Twin earth chronicles: twenty years of reflection on Hilary Putnam's "The meaning of meaning"*. New York: Routledge and M. E. Sharpe.
- Peterson, M., ed. 2015. *The Prisoner's Dilemma*. Cambridge: Cambridge University Press.
- Putnam, H. 1975. The Meaning of ‘Meaning’. First published in K. Gunderson (ed.), *Language, Mind and Knowledge*, Minnesota, Studies in the Philosophy of Science, University of Minnesota Press. Repr. in *Mind, Language and Reality, Philosophical Papers*, Volume 2, Cambridge: Cambridge University Press.
- \_\_\_\_\_. 1981. *Reason, truth and history*. Cambridge: Cambridge University Press.

- Unger, P. 2014. *Empty Ideas: A Critique of Analytic Philosophy*. Oxford: Oxford University Press.
- Searle, J. 1983. *Intentionality*. Cambridge: Cambridge University Press.
- Williamson, T. 2015. Interview in *The Times Literary Supplement*, January 16, 22-23.
- Wolfreys, J., ed. 1998. *The Derrida Reader: Writing Performances*. Edinburgh University Press.
- Wright C, B. Smith and M. Macdonald, eds. 1998. *Knowing Our Own Minds*, Oxford: Clarendon Press.