Consciousness without Physical Basis

A Metaphysical Meditation on the Immortality of the Soul

Olaf L. Müller
(Humbolt University, Berlin, Www.GehirnImTank.De)

CONTENT.

ABSTRACT. Can we conceive of a mind without body? Does, for example, the idea of the soul's immortality make sense? Certain versions of materialism deny such questions; I shall try to prove that these versions of materialism cannot be right. They fail because they cannot account for the mental vocabulary from the language of brains in the vat. Envatted expressions such as "I think", "I believe", etc., do not have to be reinterpreted when we translate them to our language; they are semantically stable. By contrast, physical expressions from the vat language are semantically instable; due to Putnam's externalism they cannot be transported to our language without change. This contrast opens the way to a new understanding of what the immortality of the soul might be like: A brain in a vat (and its mental life) might survive what the brain calls "my physical body's death".

NOTE. This is a modified version of an unpublished paper originally presented at the conference on *Epistemology and Metaphysics of the Mental* (October 1st—3rd, 2000, Heidelberg). Two endnotes and a few entries in the bibliography have been added in 2012.

Consciousness without Physical Basis

A Metaphysical Meditation on the Immortality of the Soul

I. Introduction

Most of us agree that philosophical proofs in favour of the immortality of the soul are fallacious.[1] Yet some of us believe in immortality, on the grounds of religious faith, while others prefer to take an agnostic stand (or even to deny immortality), because they want to avoid wishful thinking. And then there is a group of highly sophisticated people—mostly philosophers—who claim that the whole quarrel is besides the point because it does not even make sense to say that our soul can survive the death of our body.[2]

To support their claim they typically argue as follows. We cannot ascribe *mental* properties to any entity unless we can ascribe *physical* properties to it as well; furthermore, there are no identity criteria for somebody's soul which are not at least partly based on identity criteria for her body. If this is right then talk about "souls without body" amounts to playing with empty words.[3]

Is it nonsense to speak about a soul, or the mind, or some kind of consciousness (sufficiently similar to ours) which goes on existing without any physical basis? Is it a conceptual impossibility? I shall try to convince you that it is not. The goal of my paper is to imagine a subject who rightly claims that her mental characteristics (such as consciousness, thoughts, experiences, feelings, sense impressions, etc.) are completely divorced from physical characteristics (such as brain states, overt behavior, content and structure of the physical environment, etc.) To be sure, I do not aim at describing a subject whose mental characteristics do not have any non-mental basis whatsoever; rather my subject's mental life will be based upon *meta*physical, or *super*natural, facts—not upon physical facts. The differences between these two kinds of facts will be explained in due course.

II. Two Unhappy Strategies

There are two strategies for defending the cognitive significance of claims about immortality which I want to mention only in order to put them aside for the remainder of our discussion. The first strategy appeals to the Logical Positivist's well-known principle of verification (as a criterion of cognitive significance[4]). The strategy was adopted (among others) by Moritz Schlick, who insisted that there is a method of verification for hypotheses such as:

(1) Mentally I shall survive my body's death.

Schlick summarized the method in question with the beautiful words "Wait until you die".[5] This proposal is perhaps outdated because Schlick's test of hypothesis (1) can only be performed from an entirely private, essentially subjective, or anyway phenomenalistic point of view—which nowadays seems suspect to the majority of philosophers. Although the proposal might deserve some closer attention, I shall neglect it in order to save space.[6]

If we do not wish to appeal to possible evidence which can only be assessed from a non-naturalist, first-person perspective, then we might try to make progress by imagining possible *empirical* evidence as we know it from the natural sciences: Such evidence must be intersubjective, i.e., publicly available. But how are we to imagine possible evidence which makes it plausible from outside that a certain subject survived her physical death? The literature on parapsychology is full of answers to that question. It seems too easy to imagine publicly available evidence which would convince us that the ghost of, say, a dead friend of ours communicates surprisingly specific messages in the course of a séance or suchlike.[7]

This second strategy for defending the cognitive significance of the idea of immortality will also be neglected in this paper. Here is my main reason for putting it aside. If that strategy succeeded at all, it could only help in defending the conceivability of ideas about immortality which are rather naive, or unsophisticated, such as this one:

(1') Mentally I shall survive my body's death, so that I'll be able to send both cheerful messages to the bereaved and threatening curses to former enemies.

The strategy does not and cannot offer help to ideas which are more interesting from a spiritual point of view, ideas according to which the surviving soul is imagined to be forever detached from all mundane matters and earthly fetters.

III. Metaphysical Perspectives

In the preceding section I decided against stories told from the perspective of the subject concerned as well as against stories told from an outside perspective. So we find ourselves in the position of someone who does not allow himself to solve a riddle, neither while standing inside the circle nor while standing outside it. Is there a third standpoint from which the riddle can be solved? Possibly yes; perhaps *from above the circle*.

Let us see whether a similar change of perspective might help in our case. As I said, I do not want to solve the problem from an inside perspective, appealing to merely subjective observations; nor do I want to solve it from an outside perspective, that is, by way of imagining empirical, scientific observations in the physical world. Where is the third perspective from which we might succeed? It is, I want to urge, the perspective of metaphysics. *Meta*-physics: by this I mean a perspective which is placed above, or beyond, the physical level.

(I shall use expressions like "physics", "physical level", etc., so as to cover all levels of reality which are open, in principle, to empirical, scientific investigation. In so doing I can simplify my formulations, with apologies to chemists, biologists, geographers, etc.)

In order to set up my argument I shall assume that a materialist (or if you prefer, naturalist or physicalistic) picture of mental phenomena is right. (This is a natural starting point for our enterprise because materialism is the doctrine which backs up my opponent's claim that the idea of immortality is senseless.) By appealing to some additional assumptions, which I hope you'll find plausible, I shall infer that there are, or can be, mental phenomena for which materialists cannot possibly have an account. In short, I want to give a *reductio ad absurdum* of materialism. To be sure, I shall not venture to launch a full-scale attack on each possible version of materialism. My target is more limited: I shall try to attack precisely those versions of materialism which threaten the very conceivability of the soul's immortality. The goal of this paper is not entirely destructive though. If we manage to leave behind certain versions of materialism, then (I hope) we shall gain a better understanding of what the metaphysical point of view could be, and how we might think or speak about the immortality of the soul.

IV. The Brain in the Vat

Consider the following thought experiment. We kidnap a new-born infant, open his head, steal the brain, bury the rest of the body, and put the brain in a vat of nutrients so that it can survive as long as we wish. We connect the brain's nerve endings to the famous superscientific computer which, what a coincidence, turns out to be programmed in such a way that it simulates an ordinary human life for our victim (a life, that is, which the complete infant would have lived if we had not interfered in the first place).

Scenarios like this have been described very often; the most significant version is due to Hilary Putnam.[8] I want you to imagine a version which diverges from Putnam's in one crucial aspect: According to Putnam's version *we* are brains in a vat; according to my version, we (you and me) are not envatted. In my version of the scenario, the envatted victim is a third party—somebody else, a certain infant.

This is an important difference. Let us note some of its implications. Putnam told the story in the way he did (i.e., from the first-person perspective), because he wanted to prove that his scenario is self-refuting.[9] An important element in this famous refutation is the usage of the word "we". Putnam's conclusion is:

(2) *We* are not brains in the vat.

The proof cannot be turned against my scenario where neither you nor I (nor Putnam) is in the vat but an infant. Putnam's refutation does not have a parallel in the following:

(3) *The infant* is not a brain in the vat.[10]

The main idea behind Putnam's refutation is that an envatted subject is unable to speak about his own situation because his words fail to refer to brains, vats, computers, and other material objects on that level.[11] What is the reason for this failure of reference? The answer derives from Putnam's externalism: A speaker can only refer to objects of a certain kind if he has stood in (direct or indirect) causal contact with at least some objects of the kind in question.[12] Reference without any causal contact would be magic, says Putnam—and many philosophers agree that magical theories of reference should be put aside.[13]

Let us apply this externalist doctrine to our case of an envatted brain. As our victim was envatted in early childhood, it lacks causal contact with material things from the external world. When it thinks and speaks about "tigers" and "brains" then this is caused not by true tigers or true brains but by the supercomputer. Thus, the envatted brain does not and cannot refer to true tigers or true brains.

The upshot of this is the following. Even though envatted brains are physically possible,[14] there is no envatted brain that can say that it is a brain in a vat.[15] This applies to the envatted infant as much as to us. Are we perhaps in the vat? If so, we have (right now!) managed to say that we are; but a moment ago we saw that no brain in a vat can say that it is in a vat. So we are not envatted.[16]

V. Reinterpreting the Language of an Envatted Brain

In the preceding section you have seen that an envatted brain cannot refer to true tigers or true brains. Surely, its words "tiger" or "brain" must refer to *something*. But what then? I think the answer is to be found in the supercomputer.[17] Certain complicated codes, which are stored in its memory, must be causally responsible for the brain's enjoying tiger-like sense impressions; certain other codes will cause impressions of kangaroos. Let us call the former codes "cybertigers" and the latter ones "cyberroos". (For the sake of imagining something concrete, I want to assume that cybertigers are an isomorphic representation of those elementary particles that make up a true tiger. Similarly for cyberroos, cyberbrains, cyberrivers, and cyberclouds.)

In sum, when an envatted brain uses an expression such as "tiger", it does not refer to true tigers but only to cybertigers. So the language of an envatted brain cannot be taken at face value. When we want to understand what an envatted brain says, then we must reinterpret its language.

Not every word from the vat language needs to be changed when translated into our language. It seems clear, for example, that the envatted expressions for logical operators (such as "not", "or", etc.) do not demand reinterpretation.

Before I proceed I wish to introduce a notion which comes to mind in quite obvious fashion when we start wondering about the limits of the externalist doctrine as applied to vat language. Let us call an expression "semantically stable" when we can transport it without change from the vat language to our's.[18] Not only logical expressions are semantically stable; also the temporal vocabulary is. As the envatted brain's stream of consciousness is temporally ordered as much as our's, there is no reason to make up a difference between their temporal expressions and our temporal expressions. Envatted expressions such as "while", "when", "later", "now", etc., do not call for reinterpretation when translated into our language.[19]

By contrast, spatial vocabulary is not semantically stable. When the envatted brain says

(4) Greenland is 300 miles from this island,

then we have to translate this as follows:

(4') Cybergreenland is 300 *cybermiles* from this cyberisland.[20]

(If we had translated the spatial expression "300 miles from" without change, then we would have to assume that the supercomputer must have a diameter of at least 300 miles, which is absurd.)

Natural kind terms and observation terms are of course semantically instable. The whole descriptive vocabulary from the natural sciences (including the biological term "brain") is semantically instable. This has a nice consequence. When an envatted scientist produces an observation sentence such as:

(5) There is a tiger close to me,

then we need not claim that he is wrong (due to the absence of true tigers). To the contrary, he will be right, because the sentence must be reinterpreted:

(5') There is a cybertiger in my cybernetic neighborhood,

and because in general there really will be a cybertiger when the envatted scientist is disposed to produce sentence (5).[21] (Remember that it is cybertigers which are responsible for tiger-like sense impressions in the vat.)

This means that even envatted observation sentences have an overall tendency towards truth. And if an envatted scientist builds an empirical theory upon his observations, then his envatment does not threaten the adequacy of that theory. In short, the thought experiment of envatment does not lead to skepticism.

VI. Envatted Thoughts

Can we grant genuine mental life to the brain in the vat? I think we have to. To see why, remember that I've started from materialist assumptions concerning mental phenomena. If mental phenomena are based upon physical phenomena, then it is most reasonable to assume that two brains which are exactly alike as regards physical composition and structure must also be capable of the same types of mental activity.

Let us be a little more specific about this. The brain in the vat may be imagined to have an unenvatted twin, who was saved when we kidnapped his brother in early childhood; we can imagine that the brain which we have envatted and the brain outside the vat are exactly alike (that is, qualitatively identical).

When we idealize the situation as far as that, materialism seems to imply that the envatted subject is capable of mental life *because* his unenvatted twin brain is. (And the latter's mental capacities we assume to be obvious; the unenvatted brain belongs to an ordinary human being, after all.)

But let us be careful; we should not be misled into hasty simplifications. Even in our idealized situation (where we compare an envatted and an unenvatted brain, whose states, physical structure and composition are exactly alike) it would be wrong to conclude that the two brains lead the *same* mental life; the truth is that they don't.

They do not lead identical mental lives because of Putnam's externalism. When the twin of our envatted brain thinks about tigers then his counterpart may well be in exactly the same brain state—it will still not manage to think about tigers, for lack of causal connection. Rather on that occasion, the envatted brain thinks about cybertigers.

So Putnam's externalism does not only apply to the meaning and reference of *sentences* such as

(6) There is a tiger over there.

It carries over to the content and reference of *thought*.[22]

This does not run counter to materialism; on the contrary, it is an implication of reasonably non-magical, causal accounts of (mental) reference. According to these causal, materialist accounts, thinking about tigers is a matter of being in a certain brain state *plus* being situated in an environment having a certain physical structure; thinking about cybertigers involves being in the very same brain state but being situated in an environment having a quite different physical structure.

So let us agree that the thoughts of two qualitatively identical brains can be different; the envatted brain does not lead the same mental life as his unenvatted twin brain. *But I want to insist that they are both leading a mental life.*

Indeed, this is an assumption which was with us all the time in the course of my externalist reflections. I said that the envatted brain *thinks* about cybertigers while its twin *thinks* about true tigers. The *type* of mental activity is the same; only its *content* is different.

My next claim is this: If the envatted brain is able to think (as much as we are able to), then—like us—it is also able *to speak about* thinking. The brain's linguistic capacities are appropriate for identifying, naming, and describing its own thoughts. In other words, the envatted term "thought" is semantically stable. Let us look at an example. When the brain says

(7) I think about tigers,

then this must be reinterpreted because of Putnam's externalism. However, *nothing but* the natural kind term "tiger" has to be changed, so we get the following translation:

(7') *I think* about cybertigers.[23]

While the envatted brain is not in touch with true tigers, it certainly is in touch with its own thoughts. This is the reason for the interpretative disanalogy between "tiger" and "thought".

Against this you may object that for success in referring to thoughts you must be able to identify not only your *own* thoughts but also those of others.[24] Fine; so I have to claim that our brain in the vat can identify the thoughts of its envatted colleagues. And indeed that seems right. Suppose two envatted friends receive tiger-like stimulations which derive from one and the same cybertiger; and suppose the first brain says to the second brain "I am afraid of this tiger"; then the following sentences from the second brain say something true: "My friend is afraid of this tiger" or "My friend thinks about a tiger". If you still do not want to grant this, you would have to hold that the two envatted brains are unable to communicate. But why should they be unable? After all, they are fluent in dialogue about "tigers" as much as about "emotions elicited by tigers".

So the semantic stability of "to think" carries over to many other expressions from our mental vocabulary. Here are some further examples for semantic stability:

> to be afraid; to wonder; doubt; experience; observation; belief; mental activity; etc.

How can we reconcile the semantic stability of words like these with Putnam's externalism? Isn't there a tension here, particularly for adherents of materialism? No; Putnam's externalism puts causal constraints on the *reference*, or content, of speech and thought. But our mental vocabulary has not only referring function; it does not only occur in referring position (specifying the *content* of propositional attitudes). It often serves for *expressing* propositional attitudes. When I say "I *believe* that many wild tigers live in India" then I do not (always) intend to describe my mental life; rather I intend to describe India's fauna. We should not be surprised that Putnam's externalism (i.e., a thesis about reference) does not force us to reinterpret vocabulary whose main function does *not* lie in referring. (It would be odd to insist that an envatted Fregean judgement stroke must be reinterpreted because of Putnam's externalism.)

VII. The Brain's False Doctrine about Thought

Now I am in the position to do our first step towards the *reductio ad absurdum* that I had promised. Let us look at the situation from the perspective of our envatted speaker. Is materialism the right view about mental phenomena from his perspective?

It is not too important to find out how the speaker in the vat would reply to these questions, or which of his answers would be justified, given his information. What matters to us is something entirely different. What matters is what his *correct answers* to these questions have to look like, whether or not he can find or justify them.

So suppose the brain in the vat states its materialist doctrine as follows:

(8) My thoughts are identical to certain physical events.[25]

Is this a true statement when uttered by the brain in the vat? Clearly not; for when we translate the sentence into our language and reinterpret all and only its semantically instable vocabulary, then we obtain this:

(8') My thoughts are identical to certain cyberphysical events.

Here we have a false description of envatted thoughts. Given materialism, they are identical not to cyberphysical events, taking place in the supercomputer, but to physical events that take place in the envatted brain itself.[26]

This consideration provides the core of the advertised *reductio ad absurdum*. Its structure is as follows: We assume materialism for describing the mental situation of an envatted speaker in order to show that *his* materialism is wrong.

You may ask: Does this really provide a *reductio ad absurdum*? Isn't the materialism, which is supposed to be the point of departure for the *reductio*, different from the materialism which gets repudiated in the end? Granted: We depart from *our* materialism and repudiate the *envatted brain's* materialism. If, therefore, my opponent wishes to defend his materialist doctrine, then he should try to divorce his own version of materialism from its envatted counterpart. This divorce will take place in the next section (VIII). It will force me to invoke an additional premiss in order to complete my *reductio* (section IX).

VIII. The Semantic Destabilization of Materialism

What the argument from the last section shows is that there may be subjects whose contingent situation (of, e.g., envatment) makes their materialist doctrine false; more accurately: that there may be subjects whose contingent situation makes false *what they call* "materialism".

On first sight, this result does not give us much; it seems compatible with what *we* call "materialism". To see why, I want to ask whether the term "materialism" is semantically stable or instable. If it were semantically instable then my *reductio ad absurdum* would not yet be convincing, because in this case nothing of interest would follow from my repudiation of envatted "materialism": Should the term "materialism" turn out to be semantically instable, then the doctrine that *we* call "materialism" would be entirely different from the doctrine that *an envatted brain* calls so; and the repudiation of the envatted materialist doctrine would not carry over to our own materialist doctrine.[27]

Therefore, my materialist opponent has good motives for destabilizing the term "materialism". Are his chances of reaching that goal good? One might think so. Among other things, the materialist doctrine speaks about "physical events". But because this expression is semantically instable, an envatted brain cannot manage to speak about physical events, and thus, necessarily fails to express what we call "materialism". (It can only express cyber-materialism, i.e., the false doctrine that thoughts are identical to certain cyberphysical events.)

Although I am not sure whether I find this line of reasoning convincing, I want to agree to it for the sake of argument.

Given this, my *reductio ad absurdum* of materialism from the preceding section is not convincing, as was observed at the outset of the present section. But my anti-materialist case is not yet lost. In order to win it, I'll force my opponent to claim more than he has done hitherto. As you'll see in the next section, my opponent is obliged to admit that his materialist doctrine has a very peculiar semantic status. In our dialectical situation the doctrine must be understood as a *conceptual* claim about the mental.

IX. Conceptualist Materialism—Reduced to Absurdity

Remember that I am concerned here with defending the *conceivability* of consciousness without physical basis, and of the soul's immortality. Accordingly, I do not need to repudiate those forms of materialism which claim that *as a matter of fact* our thoughts happen to be identical to certain physical events. For if materialism expresses only a factual truth then its falsity will still be conceivable.

(When a claim about the facts is true, and thus, makes sense, then its negation must make sense too.)

So we are dealing here with a stronger form of materialism: materialism as a *conceptual* claim, a claim about the very *meaning* of our mental vocabulary.[28] Only in this analytic form can materialism be thought to imply that all talk about consciousness without physical basis is *meaningless* nonsense; only this analytic version of materialism will be the victim of my *reductio*.[29]

The first step of the *reductio* of conceptualist materialism was established in section VII: Given the truth of our materialist doctrine, the envatted "materialist" doctrine is false. This is not yet dangerous for my opponent, if we agree (as I did in the preceding section) that the word "materialism" is semantically instable. In this dialectical situation I have still not refuted materialism (a doctrine concerning thoughts and physical events), but only what an envatted brain happens to call "materialism", that is—in our language—*cyber*-materialism (a doctrine concerning thoughts and *cyber*physical events).

In order to complete the projected *reductio* I need to take one additional step. If our materialism is an analytic truth, then (I submit) cyber-materialism must be analytically true, too. To see why this must be so, let us investigate more generally what happens when an envatted brain repeats any sentence which is analytically true in our language. For example:

(9) No bachelor is female.

True, in the language of an envatted brain, the words from (9) mean, and refer to, something different than in our language. The envatted word "female" refers to cyberwomen, the envatted word "bachelor" to certain cybermen (to those who have never undergone a cyberwedding). But this reinterpretation affects *all* referring terms from the vat language, and it affects them all *in the very same way*. When envatted brains repeat our words then the relation between our words and the world gets turned into an isomorphic relation between their words and the cyberworld. Notice, however, that this reinterpretation leaves untouched the mutual—conceptual—relations *among the words themselves*. Cyberbachelors fail to be cyberfemale in the very same way bachelors fail to be female.

So my claim is that the very same sentences are analytic in our language and in the vat language.[30]

Given this, I can complete my *reductio* of conceptualist materialism.

(i) *Assumption:* The sentence "All thoughts are identical to certain physical events" is an analytic truth in our language.

(ii) All thoughts are identical to certain physical events. (From (i); disquotation.)

(iii) Envatted thoughts are identical to certain physical events. (From (ii).)

(iv) When an envatted brain says "All thoughts are identical to certain physical events", then this is true iff all thoughts are identical to certain cyberphysical events. (Externalist reinterpretation of semantically instable words.)

(v) The envatted sentence "All thoughts are identical to certain physical events" is false. (From (iii) and (iv).)

(vi) If a sentence is analytically true in our language, then its envatted counterpart is analytically true in the vat language. (From the definition of analyticity and the description of the vat scenario.[31])

(vii) The envatted sentence "All thoughts are identical to certain physical events" is analytically true in the vat language. (From (i) and (vi).)

In order to avoid the contradiction between (v) and (vii) I conclude that our assumption (i)—conceptualist materialism—must have been false; it is not the meanings of our concepts which guarantee that thoughts have to be identical to certain physical events.

X. Other Versions of Materialism

In the last three sections I assumed materialism, as formulated on our level outside the vat, in order to repudiate the envatted parallel of materialism, which in turn gave me ammunition against conceptualist materialism on our level. (This was the heart of my *reductio ad absurdum*.)

My argument was directed against versions of materialism which employ an identity thesis such as this one:

(8) My thoughts are identical to certain physical events.

But it seems clear that the *reductio* works not only against materialists who believe in an identity thesis, but also against those who subscribe to supervenience. When an envatted brain says:

(10) My thoughts supervene on certain physical events,

which must be translated into our language as follows:

(10') My thoughts supervene on certain cyberphysical events,

then this is wrong, given the truth of supervenience on our level. (Given the truth of supervenience on our level, the envatted brain's thoughts supervene on neural events, happening in a true brain, not on cyberneural events in the supercomputer.)

And again, when the materialist comes up with something other than identity or supervenience, I can repeat my argument. For example, I can run a *reductio* on the following version of materialism:

(11) My thoughts are based upon physical events.

To be sure, my argument does not show that materialist doctrines such as (8), (10), and (11) are false. The argument shows merely that these claims cannot be conceptual truths. This is good news for the metaphysical idea of immortality; we can expect that it will soon be released from the charge of being inconceivable. Before freeing it from that charge (at the end of this paper), I want to say something more general about the language of metaphysics. The ambition of the following sections is to make it plausible that the language of an envatted brain has *metaphysical* resources sufficiently strong so as to enable the brain to speculate about its own mental situation.

XI. Beyond Nature

We saw that many envatted sentences about the brain's thoughts turned out false, due to materialism on our level. You may ask: Given the (factual, i.e., non-conceptual) truth of any version of materialism on our level, what is the right theory on the level of an envatted brain? Of course, when the brain in the vat denies its original sentences, it says something true:

(12) My thoughts are *not* identical to certain physical events.

(13) My thoughts do *not* supervene on certain physical events.

(14) My thoughts are *not* based upon physical events.

But with these negative claims we should not be satisfied. We want to see which positive account of mental phenomena would be right when put forward by the brain in the vat. Let us imagine that the envatted speaker is willing to risk something and wishes to claim more than in statements (12)—(14). For example, he might limit the scope of the negation in these statements as follows:

(15) My thoughts are identical to non-physical events.

(16) My thoughts supervene on non-physical events.

(17) The basis of my thinking is beyond physics.

Remember that for convenience I have been using the expressions "physics", "physical", etc., in a very broad sense throughout this paper. This means that (negative) sentences such as (15)—(17) are more radical than may seem on first sight. In a more explicit form they say something like this:

(18) My thoughts are identical to non-natural events.

(19) My thoughts supervene on super-natural events.

(20) The basis of my thinking is beyond nature.

These are very risky claims indeed. What exactly are they supposed to mean? When *we* speak about nature then we refer to a realm of entities and events, which are located in space and time, which can stand in direct or indirect causal contact with one another and with us, and which, therefore, are in principle open to empirical scientific investigation. But what is a realm *beyond* nature supposed to be? What is the super-natural?[32]

It is easier to answer this question not for our language but for that of an envatted brain. Let us first clarify what the envatted term "nature" means. When the envatted brain speaks about nature, this refers to a realm of entities and events with which the brain can stand in direct or indirect causal connection. The envatted expression "nature" covers cyberobjects in cyberspace, that is, codes in the supercomputer's memory—plus their changes (that is to say, cyberevents).

But in statements (18)—(20) the envatted brain speaks about a realm *beyond* nature. What does that mean? Well, if we remember what the envatted term "nature" means, then it seems clear that its term "beyond nature" refers to a realm of things beyond the brain's cyberspace, i.e., beyond the memory of the supercomputer. The term covers material objects such as true tigers.

It covers a realm of things about which the envatted brain cannot possibly gain empirical knowledge, due to lack of appropriate causal connection. To the brain in the vat, supernatural things are things which are natural things for us. They are material objects which are located in the surroundings of the vat, the computer, and the envatted brain. The brain, the vat, and the computer are themselves "supernatural things" from the envatted point of view.[33]

If this is right then the brain in the vat says something true when it produces statements such as (18)—(20), because their translation runs thus:

(18') My thoughts are identical to events that do not take place in the supercomputer.

(19') My thoughts supervene on natural events.

(20') The basis of my thinking is beyond the supercomputer.

(In uttering sentences which must be so translated, the envatted brain says something true although it does not and cannot know that it has hit on the truth.)

XII. Metaphors and Charity

You may ask how an envatted brain manages to refer to entities with which it does not stand in causal connection at all? Doesn't this violate the causal constraints on reference invoked by Putnam's externalism?

Yes, it does. But don't be afraid of that. When properly understood, Putnam's externalism must be restricted to words that one uses for describing natural phenomena—the phenomena with which one can (in principle) stand in causal connection. Externalism is most plausible when applied to natural kind terms. (It holds good for other expressions from the natural sciences, too.)

But the externalist doctrine fails when applied to words that do not even pretend to refer to things *via* causal connection. The doctrine fails, for example, for numerals. (That is to say, numerals are semantically stable; the envatted word "seven" does not need to be reinterpreted.) Externalism also fails for talk about the supernatural—or so I want to claim.

In statements such as (18)—(20), the envatted speaker explicitly goes beyond the domain of things that are accessible to him for empirical investigation. The vocabulary from (18)—(20) is not made for expressing scientific truths, and it does not pretend to; it belongs to another language game.

It may be instructive to have a closer look at an essential element from the boldest sentence that we have come across up to now:

(20) The basis of my thinking is *beyond* nature.

In everyday life, and in the sciences, the word "beyond" has a spatial meaning, and in this usage, it is subject to externalist reinterpretation (when uttered by an envatted brain). But I have left the word untouched when I translated the envatted sentence (20) into our language:

(20') The basis of my thinking is *beyond* cybernature.

This means that the original sentence (20) was not interpreted as a statement from the vat sciences; the word "beyond" was not understood in the literal meaning (signifying cyberspatial relations between codes in the memory of the supercomputer).

Rather, we have understood the word *metaphorically*, signifying something *analogous* to what the brain usually describes with the word "beyond". What gives us the right to interpret the envatted statement (20) metaphorically? Two reasons: charity and fantasy.

Let me explain. First, if we interpreted the spatial expression from statement (20) literally, not metaphorically, then we would obtain plain nonsense. (It would amount to speaking about a spatial place beyond space; or rather about something which is simultaneously inside and outside the brain's *cyber*space.[34])

A very good, old maxim for interpretation is the celebrated principle of charity. It warns us against ascribing excessively weird opinions to an interpretee. If possible (the principle says) we should find an interpretation that diminishes the amount of craziness which we blame on the speaker.[35] And in the present case there is a more charitable interpretation at hand than the literal one, an interpretation that avoids ascribing nonsensical statements to the envatted brain. Surely, the charitable interpretation of the envatted statement:

(20) The basis of my thinking is beyond nature,

is the translation I have given:

(20') The basis of my thinking is beyond cybernature.

As you have seen, the trick in this translation is to read the expression "beyond" not literally but as an envatted metaphor. This we can do because it is a metaphor that speaks to us. We are open to that metaphor, due to our fantastic fantasy for playing with words which are literally empty but not empty *for us*.

Much more would have to be said about metaphor; alas, the subject is beyond the scope of this paper. But before we leave it altogether I want to mention briefly two observations concerning the envatted metaphor of a "realm beyond nature".

First, the charitable interpretation of statement (20) forced us to ascribe to the brain in the vat a metaphorical use of its expression "beyond"; but in our translation we were able to *unpack* the metaphor, employing our word "beyond" in its literal, spatial meaning. It is interesting that, unlike us, the brain in the vat cannot unpack its own metaphor. This is due to its lack of causal connection to what *we* call spatial objects.

But even though it does not know how to unpack the metaphor, nor whether the metaphor can be unpacked so that it refers to something, it can still be aware that the word "beyond" in its own sentence (20) is being used metaphorically. The brain in the vat knows (as well as we do) that the literal understanding of the expression "beyond" produces nonsense in the sentence (20).

This brings me to the second observation that I announced. As the brain in the vat understands its own statement (20) metaphorically, it can explicitly guard it against literal misunderstandings. For it can rephrase it as follows:

(20*) The basis of my thinking is beyond nature, *but not in the literal, spatial sense of the word "beyond"*.

If you were not convinced by my charitable interpretation (20') of the original envatted statement (20), then I hope you will at least agree that in (20*), the word "beyond" should not be read literally. (Otherwise your unwillingness to charity goes far beyond what a good interpreter can afford.)

XIII. Metaphysical Discourse

Perhaps you find it a little unsatisfactory that the whole burden of my reasoning has come to rest upon the unsharp notion of metaphor. It seems trivial and uninformative to be told that there are metaphorical modes of speech which are not governed by the strict rules of externalist accounts of reference. It might be helpful if I could say something more concrete about the specific use of metaphor which I have brought into play.

To what kind of discourse does the quasi-spatial expression "beyond" belong? It belongs, I submit, either to religious discourse or to the discourse of metaphysics. I shall not say much about the metaphor's place in religious discourse. Suffice it to say that, in religious speech, expressions like "beyond nature" are often used for formulating statements that express religious faith. Think of statements which say that earthly matters are unimportant because there is some kind of higher reality with which we should be concerned first of all. Notice that it is a typical move in religious discourse to warn against literal understandings, a move similar to mine in (20*).

In the discourse of metaphysics, by contrast, we have less confidence in the truth and reference of statements that contain metaphorical expressions such as "beyond nature". As I conceive of it, metaphysics is a discipline for posing significant *questions*—questions which cannot possibly be answered by the sciences. The domain of the sciences is nature, that of metaphysics (*meta*-physics!) is beyond nature. (There are as many conceptions of metaphysics as there are metaphysicians. Of course, I do not claim to have the copyright for that word.)

Thus it is obvious that the old positivist sense criterion cannot be appropriate for metaphysical questions. Neither is externalism appropriate for determining the extension of metaphorical, metaphysical expressions.

Even so, a metaphysical question should be clear. It must allow us to say *what it would be like* if one of the answers to it were right. Because the use of metaphors and analogies is far less restricted than that of natural kind terms and other expressions from the sciences, the difference between clarity and obscurity becomes crucial in metaphysics. There is no fixed canon of rules (no algorithm) for posing and clarifying good metaphysical questions. (But then, neither is there an algorithm for solving scientific problems.[36])

Perhaps I should make explicit which method for clarifying metaphysical questions I have wanted to follow throughout this paper. Instead of posing metaphysical questions from our own perspective, I have put them in somebody else's mouth, somebody whose situation we can supervise from above, from a perspective which is not open to him in principle. From outside we have been looking at the situation of the brain in the vat and at what is going on in his world, in the supercomputer.

I have been trying to understand the envatted brain's metaphysical statement:

(20) The basis of my thinking is beyond nature,

or rather, its metaphysical question:

(21) Is there a realm beyond nature?

In search of a charitable interpretation we have appreciated how the use of metaphors makes possible reference without causal connection. (There was no mystery about this; rather, it followed naturally from applying the principle of charity.)

This exercise was not performed for the sake of the envatted brain; we performed it for ourselves, for clarifying *our* metaphysical question:

(21) Is there a realm beyond nature?

The fantasy of the brain in the vat gives us a picture, or a simile, of what it would be like if the answer to *our* question was "yes". The simile clarifies our question, it improves our understanding of it—but it does not yield a literal, non-metaphorical reformulation of the question. The simile plays a role comparable to that of Plato's myth of the cave. This is more than nothing. Unless you are blinded by anti-metaphysical prejudice, you should admit, I think, that it increases your control over the idea of the supernatural when you contemplate that idea by way of analogy. You could, for example, imagine yourself in the position of someone who really is (like an envatted brain) dependent on what he calls the "supernatural" (e.g., the vat).[37]

So much about metaphysical language in general. Let us return to my starting point and see whether what we have learnt applies to the idea of immortality.

XIV. Death in the Vat

In order to clarify metaphysical questions concerning immortality I shall try to repeat the line of thought that has helped us in clarifying questions concerning the supernatural. I shall look from outside at somebody whose mental life does not end at the moment of what he calls his "biological death". This will bring us to an understanding of *what* the soul's immortality might be like.

In order to divorce biological and mental death, I do not need to invent another thought experiment. It suffices to continue the one with which we were dealing all the time: the thought experiment of the brain in the vat. The projected divorce should be easy; my main idea will be that mental expressions are semantically stable while biological expressions are not.

Let us imagine the situation of our envatted victim shortly before its death. After seventy years of envatment, the victim suffers from simulated symptoms of terminal illness, say, simulated lung cancer. The lungs aren't there, of course—neither is there any cancer. (Where would it be?) But the pain is there none the less. The brain is connected not to true lungs but to *cyberlungs*, that is to certain parts of the cybernetic representation of the human body which our victim would have had, had he not been envatted seventy years ago.

So we have a sad case of cybernetic lung cancer. (There are cybercancer cells spreading all over the cyberlung.) What happens next? According to a conservative version of Putnam's scenario, the story will end rather soon; according to my version, the story will last longer.

As the cyberpulse stops, the supercomputer simulates a sudden attack of pain; and then: nothing. The computer stops sending further impulses to the brain in the vat, because the cyberbody begins to decompose (so that its cybernose, cybereyes, cybernetic pain receptors, etc., do not function any longer).

What about the envatted brain itself? (Notice that this is not a question about the subject's *cyber*brain, which undergoes cybernetic decomposition as much as the rest of the cyberbody whose death is being simulated. The question is about a true brain—the one which we had stolen and envatted seventy years ago.)

Up to now, the envatted brain is still alive. True, it does not enjoy sensory stimulation any longer; its experiential field is empty, black, and silent—apart, perhaps, from phantom pain.

But the envatted brain can still think; and it is likely to be frightened.—Or, is it really? This depends; we can continue the story however we wish.

According to the conservative version of Putnam's scenario, the envatted brain must be killed at the moment of its simulated death. In this scenario, the true brain's biological death *coincides* with the simulated cyberbrain's death.

In my scenario, on the other hand, we prolong the biological life of the envatted brain beyond the moment of its simulated death. How are we to do this? As before: We have to supply sufficient amounts of nutrients, we have to control the fluid's temperature, we have to prevent the brain from inflammation, and so forth.

XV. The Immortality of the Soul

To sum up the result from the preceding section, it seems theoretically possible that we can keep the envatted brain alive beyond the moment when the computer ceases to supply simulations of an ordinary life.

Of course, I have not *literally* described a situation where a subject continues leading its mental life without physical basis. The physical basis of the subject's mental life is the brain in the vat. But the scenario helps us to see what it would be like to survive one's own death. For if we look at the situation from the linguistic standpoint of the envatted brain, then we do have a case where the mental life lasts longer than the biological life.

Suppose the envatted brain suffers from cyberlung cancer and raises the following question:

(22) Will my mental life last longer than the biological life of my body?

Mental and temporal notions are semantically stable, biological ones are not. So when we translate the question into our language, we obtain this:

(22') Will my mental life last longer than the cyberbiological life of my cyberbody?

And the positive answer to this question yields a true description of the scenario we have been imagining.

As I have announced, we are able to describe a situation where hopes concerning an afterlife come true. Notice that this time we are moving on ground that is even more solid than in the case of the supernatural, which I've been discussing earlier on. In interpreting an envatted brain's talk about the "supernatural" or the "realm beyond nature", I had to ascribe a metaphorical mode of speech to the speaker. (The quasi-spatial expressions "super" and "beyond" could not be understood in their literal meaning.) This time, by contrast, I interpreted the envatted brain literally, *tout court*. In its literal meaning, the envatted expression "mental life" is semantically stable; so is the envatted expression "lasts longer than".

Therefore, when *we* want to speculate about our chances concerning immortality, we can also speak in the literal, non-metaphorical mode:

(22) Will my mental life last longer than the biological life of my body?

Even when understood literally the question makes sense.[38] What the right answer to this question is I do not know. It is not a question which can be answered by the natural sciences. Still it is a crucial question. It is a question from an old philosophical project—the project of metaphysics.[39]

Notes

1 Compare Patzig [P]:35-45 as an example for the destruction of Plato's proofs (from the dialogue *Phaidon*). Throughout this paper, the phrase "immortality" is meant to cover both possibilities: either that the soul survives the body for a time or that it does so eternally.

2 See, for example, Ayer [LTL]:117.

3 What I have described here is of course only a sketch; the literature is full of more elaborate arguments. See, for example, Geach [MA]:111-117, [GS]:17-29 and Flew [LoM]:16-29, 102-106, 111/2, 115/6 *et passim*. See also Williams in H.D. Lewis (et al.) [LaD]:57-59. (Although Williams summarizes his position by way of saying: "[...] there is no such thing as immortality" ([LaD]:59), his point is clearly directed against the conceivability of that idea, as can be seen from the following quotation: "[...] you haven't got *any control over the idea* of a personal immortality" ([LaD]:58, my emphasis).) Williams' line of thought is related to a similar one by Strawson, see [I]:87ff., especially 97, 102. (Compare, however, Strawson [I]:115/6.) — To my knowledge, Carnap is the earliest writer whose views in the philosophy of mind imply very much the position I have sketched in the main text. See, for example, Carnap's first sentence in [PiPS]:107 (§1): "Es soll im folgenden die These erläutert und begründet werden, daß jeder Satz der Psychologie in physikalischer Sprache formuliert werden kann" ([PiPS]:107 (§1), italics omitted). Later on in the same paper, he says: "[...] ein Satz über Fremdpsychisches besagt, daß ein physikalischer Vorgang bestimmter Art am Leib der betreffenden Person stattfindet" ([PiPS]:117 (§4)).

4 For the criterion of significance see Carnap [SiP]:47-54 (§7). (See also Schlick [MV]:339-43.)

5 Schlick [MV]:356.

6 For a defense of that strategy, see Hick [TV]:258, 261-267. Compare also Hick [PoR]:100-102. These ideas have been criticized by Williams and Flew, see Williams [IS]:40, Flew [LoM]:112ff. and Flew [CMWH]:245-248.

7 Among others, Ducasse and Schlick have not resisted the temptation to make up stories about ghosts. See Ducasse [ECfP]:225-228 and Schlick [MV]:357.

8 See Putnam [RTH]:5/6, 12/3. In order to avoid philosophical perplexities as to the possibility of private languages, Putnam takes care that his envatted brain is not alone; he puts more brains to the vat, connects them all to one and the same computer and makes sure that they can exchange messages.

9 Putnam [RTH]:7.

10 That the first-person perspective is essential to Putnam's proof has been observed by Crispin Wright in [oPPT]:226/7.

11 Putnam [RTH]:12/3.

12 See Putnam [MoM]:223ff.

13 Putnam [RTH]:3-5. Crispin Wright is one of those who easily agree with Putnam's externalism, see Wright [oPPT]:230-32.

14 See Putnam [RTH]:15.

15 See Putnam [RTH]:8.

16 For the most attractive reconstruction of Putnam's proof, see Wright [oPPT], especially p. 224.

17 This is one of the options which Putnam considers, see [RTH]:14.

18 The English expression "semantically stable" has independently been used by other authors (see Grünbaum [IINP]:263-268, 272; Sankey [ITC]:460/1; Bealer [MP]:201, 204, 208, [APKS]:132, 134/5, 137, 142n13, [oPoP]:23-25, [IAoP]:228,

237n28, [MERR]:72, 105, 107, 114, 115, 120-123). Grünbaum's notion and Sankey's notion belong to quite specific, yet disjoint, contexts in the philosophy of science. By contrast, Bealer's notion was designed for metaphilosophical purposes: he introduced it to save the epistemological power of certain modal intuitions from attacks by scientific essentialism à la Kripke. As Putnam's externalism resembles Kripke's views in crucial respects, it is unsurprising that Bealer's notion of semantic stability bears a significant resemblance to the one I here invoke in order to deal with Putnam's insights. According to Bealer, "[a]n expression is semantically stable iff, *necessarily*, in any language *group* in *an* epistemic situation qualitatively identical to ours, the expression would *mean* the same thing" (Bealer [APKS]:134, my italics; a related notion of "context-free terms" is defined by Hirsch [MNCT]:246/7, compare Bealer [MERR]:72n1). Let me highlight four differences between Bealer's notion and mine. (i) Where Bealer's definition appeals to meaning (which can be understood in many different ways), my definition appeals to translation instead (which I propose to spell out in terms of radical translation, or interpretation, see Müller [MSS], sections 3 and 4). A consequence of this difference is highlighted in endnote 23 below. (ii) Bealer's definition invokes necessity, which my definition does not. Modally, then, Bealer's notion is stricter than mine. But it remains unclear why such modal strengthening matters. (iii) Bealer's notion deals with language *groups* and thus, unlike mine, does not apply to a single speaker of some private idiolect (used by an envatted brain that is, perhaps, kept *incommunicado*; but see endnote 7). Bealer's choice is motivated by the wish to sidestep any potential problems that might arise from phenomena of linguistic division of labour: these phenomena cannot cause trouble for complete communities, and thus, can safely be ignored when applying Bealer's notion (see Bealer [oPoP]:33n31). (iv) Bealer's notion is not exclusively concerned with *envatment*, but more generally, with *any* non-standard, but qualitatively conservative, variation in the speakers' environment. Therefore, less terms might qualify for semantic stability in Bealer's sense than in mine. This makes me wonder—given the almost unlimited imagination with which analytic philosophers concoct surprising stories—how we could ever be sure that a given term is semantically stable in Bealer's sense. Bealer does not seem to be worried by this. On the contrary, without much ado he comes up with quite extensive lists of stable terms (Bealer [APKS]:134/5, [oPoP]:23/4, [IAoP]:228, 237n28; compare Bealer [PLoS]:355/6 *et passim* for related ideas). In a monograph about semantic stability, I have by contrast felt the need to prove with detailed arguments that certain philosophical expressions (such as e.g. "kind") are semantically stable, in the sense of my weaker notion (which is restricted to envatment; see Müller [MSS], section 13). When I submitted these arguments in 2001 and published them in 2003, I was unaware of Bealer's earlier work, which I would have greeted with the enthusiasm of someone desperately looking for philosophical allies; differences aside, the overall similarity between the two independently conceived notions is striking and an indication of the fruitfulness of their root idea. In 2003 David Chalmers broached the very same idea, again independently from any earlier efforts by others. Without referencing such earlier efforts, he spoke of "semantic *neutrality*" (Chalmers [MaM], notes 1, 9). In 2010, Chalmers does quote Bealer, however, and points out some differences between his own notion of semantic neutrality and Bealer's notion (see Chalmers [CoC]:203/4). Chalmers' notion on p. 204 presupposes the apparatus of two-dimensionalism (which is controversial, see e.g. Bealer [MERR]:87-99). Like me, but unlike Bealer, Chalmers applies the notion to envatment, for which purpose the controversial apparatus proves inessential (Chalmers [MaM], notes 1, 9 = Chalmers [CoC]:481, 487/8). The coincidences between Chalmers' conclusions and my own are so overwhelming that I was rather puzzled to learn about them when preparing the electronic version of the present paper. *Déjà lu*—are we perhaps in the matrix, after all? [This endnote was added in 2012.]

19 Alison Laywine, Felix Mühlhölzer, and Hilary Putnam have pointed out to me that it is not obvious that *each* temporal expression is semantically stable. Granted, in order to demonstrate the semantic stability of expressions such as "0.000075 seconds" more has to be done than is possible in this paper. Although this in an interesting issue, we need not worry about it here. The examples from the main text are sufficient for our purposes.

20 Let us call a code X "300 cybermiles from" a code Y if, for example, the envatted brain, who enjoys a simulated trip with a simulated speed of 100 miles per hour, needs three hours in order to move from X-experiences to Y-experiences.

21 A similar point is made by Putnam in [RTH]:14.

22 Putnam says that his externalist reflections deliver "preconditions of reference *and hence of thought*" ([RTH]:16, emphasis changed).

23 Here lies the main difference between my account of semantic stability and Bealer's, who claims that "[the term] 'I' is paradigmatically semantically *unstable*" (Bealer [MERR]:115, his emphasis; a possible reason for this difference is indicated above in endnote 18, item (i)). Unlike Bealer, I propose to deal with the term "I" in connection with other terms such as "think". Notice that even in the expression "my body", the indexical itself does not require changes when translated from our language to vat language: "my bit-body". [This endnote was added in 2012.]

24 The objection is credited to Paolo Casalegno and goes back to observations made by Strawson [I]:99.

25 One might suspect that what I have said about non-referential functions of the mental vocabulary (at the end of the preceding section) begs the question against materialist doctrines such as (8). But the details of formulating materialism do not matter much for my argument; my argument works also against versions of materialism such as this one: "Whenever I believe something, then physical events of a certain type must occur".

26 Or anyway, *part* of their identity depends on physical (neural) events *in the brain*; this qualification leaves room for externalist versions of materialism which say that thoughts are based upon facts about the brain *and* facts about its environment. My main point remains untouched by this complication. Statement (8') still misrepresents the mental situation of an envatted brain.

27 This has been pointed out to me by Christian Nimtz. The leading idea of my next paragraph is also credited to him.

28 In contrasting factual with conceptual versions of materialism I have to appeal to the notion of analyticity, which many philosophers believe to have been demolished by W.V.O. Quine. (The *locus classicus* is Quine [TDoE].) I have shown elsewhere how Quine's objections against the distinction between analytic and synthetic sentences can be met. (See Müller [DQDT] and [SA]:240ff, especially p. 275.) As I was able to explicate analyticity exclusively in naturalist terms, it is appropriate to use that notion in our present context, where we are moving on the ground of naturalism and materialism. — Even so, one might wonder whether there are non-conceptualist versions of materialism which would also make the soul's immortality inconceivable. Couldn't my opponent invoke non-conceptualist materialism in the form of an apriori truth, or as a necessity? I

do not think so. The claim that the idea of immortality lacks *meaning* is a *semantic* claim, not an epistemological or modal claim. To see this, notice that there can be necessities (like "Water is H_2O") as well as apriori truths (like "I am here now") whose falsity is conceivable rather than meaningless.

29 Of course, there may be many different arguments against materialism as an analytic theory about the mental. As we shall see at the end of this paper, the (new) anti-materialist line of reasoning adopted here invites a straightforward application concerning immortality. By contrast, most of the anti-materialist arguments from the literature are at best loosely connected to our main concern. This is my excuse for ignoring what other writers have argued against conceptualist versions of materialism.

30 Something close to this claim (including its motivation from the preceding paragraph) has been brought to my attention by Paolo Casalegno. (He did not subscribe to the claim, however; furthermore, he formulated it in terms of aprioricity rather than in terms of analyticity.) To prove the claim, more has to be done than is possible in this paper. The details of such a proof depend on the exact version of analyticity with which one wishes to work. The notion of analyticity that I have defined and defended elsewhere, for example, is built upon Quine's concept of sensory stimulation "at the surfaces of the speaker", such as visual stimuli at the retina of an ordinary human eye (see Quine [WO]:31f). Now an envatted brain does not have real eyes. Therefore, we are well advised to locate the stimuli in question still closer to the stimulated brain than in Quine's theory. I propose to substitute Quinean sensory stimulations by the neural input patterns as they arrive at the brain's interface. (This change does not affect much of the spirit of Quine's concept—nor of the usage I made of it in my definition of analyticity.) Having made this substitution, the proof of my claim is straightforward: *Per constructionem* an envatted brain enjoys the very same neural input (and the very same reactions on its output) that we enjoy; this is so even during the process of language acquisition. Therefore, the envatted brain develops the same (verbal) dispositions to react on input that we do. Now it is merely these dispositions which matter for questions of analyticity (says my definition of that notion, see [DQDT]:93 and [SA]:275). So there will be no difference between our analytic truths and those from the language of an envatted brain. (Similar proofs are possible for more traditional versions of analyticity; for example, if you define analyticity in terms of (confirmatory and disconfirmatory) evidence, see Grice / Strawson [iDoD]:210.)

Is my claim incompatible with Putnam's externalism? Doesn't externalism prevent us from internally knowing what our words mean and refer to, and hence, from internally knowing what is true by virtue of meaning alone? No. Externalism is at stake when we seek to interpret someone's language from outside; it infects all interlingistic characterization of meaning, synonymy, and reference, but not their *intra*linguistic characterization. (That's why externalism is compatible with disquotationalist claims about reference.) Now analyticity can be defined in terms of intralinguistic synonymy and *vice versa* (see Quine [WO]:65, [TDiR]:271 and Müller [DQDT]:97 note 12); thus externalism does not matter for questions of analyticity. (The worry concerning analyticity and externalism is credited to Christian Nimtz, oral communication.)

31 See previous endnote.

32 Shoemaker faces similar problems when he urges that we should resist the temptation to define "immaterial substance" as meaning simply "substance that is not a material substance" ([ID]:114/5).

33 If an envatted brain repeated Shoemaker's contemplations on "Non-Cartesian Dualism" ([ID]:113ff.), then it might refer to itself (to the brain, that is) by way of Shoemaker's expression "ghostly brain" ([ID]:113).

34 This is the envatted parallel to Kant's first antinomy concerning space, see Kant [KRV]:454-457* (A 426-433, B 454-461).

35 Cf. Quine [WO]:59 and Davidson [oVIo]:196.

36 This has been shown by Tetens in [FWMR].

37 When dualists are tempted to speak about "other material than physical matter" (Hick [TV]:265), or about "immaterial stuff" (Swinburne [PI]:27), they risk being charged with self-contradiction: "[...] speaking of immaterial stuff seems dangerously close to speaking of immaterial matter", cf. Shoemaker [SSR]:125. (The charge is rather old; it can be traced back to Hobbes who directed it against the expression "incorporeall substance", see [L]:17.) In his own, positive account, Shoemaker considers "immaterial substances" that can be individuated not spatially but only quasi-spatially ([ID]:114-118). He admits, however, that "[...] it is in a rather thin sense that it is 'conceivable' that there should be immaterial substances", because to say that some of the relationships between immaterial substances are quasi-spatial "is not to say what they would be" ([ID]:130, note 8). It seems that my line of thought (as regards the realm beyond nature) is directed towards a goal comparable to Shoemaker's. If I am not mistaken, our understanding of metaphorical expressions such as "quasi-spatial" becomes less thin than in Shoemaker's discussion, once we consider them as expressions from the vat language and contemplate their *non-metaphorical* translations into our language (see the preceding section).

38 Notice that I have said nothing to defend the conceivability of the idea that after my physical death I will be the same *person* that I was while biologically alive. The question of personal survival raises additional problems which are beyond the scope of this paper; for a discussion see Shoemaker [ID] and H. Lewis (et al.) [LaD].

39 This is a modified version of a paper originally presented at the conference on *Epistemology and Metaphysics of the Mental* (October 1st—3rd, 2000, Internationales Wissenschaftszentrum, University Heidelberg). I am grateful to my commentator on this occasion, the late Paolo Casalegno, whose brilliant and sharp remarks have helped me see more clearly which version of materialism should have been my target. I also want to thank Wolfgang Carl, Andreas Kemmerling, Christian Nimtz, Sven Rosenkranz, and Hans-Peter Schütt for valuable criticism of the paper's first version. I am indebted to Alison Laywine and Cherilyn Keall for stylistic advice. I also wish to express my gratitude to both Hilary Putnam and Holm Tetens, who convinced me in 1996/7 that what I am trying to do here is metaphysics proper. (In Putnam's view, this is an objection, in Tetens' view, it is not.) Indeed, their reactions to my first attempts at applying the notion of semantic stability to envatted brains have led me to embark on a metaphysical adventure. I presented the resulting paper to Putnam in April 1997 (Müller [AZ]), and gave presentations of it in Berlin (FU, January 9th, 1997), Munich (LMU, September 17th, 1997), Göttingen (November 25th, 1997), Bonn (April 30th, 1998), Konstanz (July 2nd, 1998), Heidelberg (July 11th, 2000), Hannover (January 9th, 2001), and Bayreuth (December 18th 2001). English versions were delivered in the *Philosophy Club* at St. Andrews (May 1st, 2002) and at the conference *The Limits of Knowledge*, Berlin (FU, July 9th, 2002). Many of the critical remarks from the audiences on these occasions have helped me gain clarity about the nature of my project and the best way to argue for its central

claims. To my surprise, nobody seemed to regard the project's overall goals tenable.

Literature

Ayer, Alfred Jules [LTL]: *Language, truth and logic.* (London: Victor Gollancz, second edition, 1946).

Bealer, George [APKS]: "A priori knowledge and the scope of philosophy". *Philosophical Studies* 81 (1996), pp. 121-142. [Entry added in 2012.]

Bealer, George [IAoP]: "Intuition and the autonomy of philosophy". In DePaul et al (eds) [RT]:201-239. [Entry added in 2012.]

Bealer, George [MERR]: "Modal epistemology and the rationalist renaissance". In Gendler et al (eds) [CP]:71-125. [Entry added in 2012.]

Bealer, George [MP]: "Mental properties". *Journal of Philosophy* 91 No 4 (1994), pp. 185-208. [Entry added in 2012.]

Bealer, George [oPoP]: "On the possibility of philosophical knowledge". *Nous* 30 (1996), Supplement: *Philosophical Perspectives*, 10, *Metaphysics*, pp. 1-34. [Entry added in 2012.]

Bealer, George [PLoS]: "The philosophical limits of scientific essentialism". *Philosophical Perspectives* 1, *Metaphysics* (1987), pp. 289-365. [Entry added in 2012.]

Carnap, Rudolf [PiPS]: "Psychologie in physikalischer Sprache". In *Erkenntnis* 3 (1932-33), pp. 107-142.

Carnap, Rudolf [SiP]: *Scheinprobleme in der Philosophie. Das Fremdpsychische und der Realismusstreit.* (Günther Patzig (ed); Frankfurt / Main: Suhrkamp, 1966).

Chalmers, David J. [CoC]: *The character of consciousness.* (Oxford: Oxford University Press, 2010). [Entry added in 2012.]

Chalmers, David J.: "The matrix as metaphysics". In Chalmers [CoC]:455-494. [Appeared first 2003 in the *philosophy section* of the official *matrix* homepage under http://whatisthematrix.warnerbros.com/rl_cmp/new_phil_main.html; on September 10th, 2012, this link was deactivated]. [Entry added in 2012.]

Clark, Peter / Hale, Bob (ed) [RP]: *Reading Putnam.* (Cambridge / Mass.: Blackwell, 1994).

Crawford, Sean (ed) [PoM]/IV: *Philosophy of mind. Critical concepts in philosophy. Volume IV. Consciousness.* (London: Routledge, 2011). [Entry added in 2012.]

Davidson, Donald [IiTI]: *Inquiries into truth and interpretation.* (Oxford: Clarendon Press, 1984).

Davidson, Donald [oVIo]: "On the very idea of a conceptual scheme". In Davidson [IiTI]:183-198.

DePaul, Michael Raymond / Ramsey, William (eds) [RT]: *Rethinking intuition. The psychology of intuition and its role in philosophical inquiry*. (Lanham: Rowman & Littlefield, 1998). [Entry added in 2012.]

Ducasse, Curt John [ECfP]: "The empirical case for personal survival". In Flew (ed) [BMD]:221-230.

Flew, Antony (ed) [BMD]: *Body, mind, and death*. (New York: Macmillan, 1964).

Flew, Antony [CMWH]: "Can a man witness his own funeral?" *Hibbert Journal*. Vol. LIV (April 1956), pp. 242-250.

Flew, Antony [LoM]: *The logic of mortality.* (Oxford: Blackwell, 1987).

Geach, Peter [GS]: *God and the soul.* (Bristol: Thoemmes, 1994).

Geach, Peter [MA]: *Mental acts. Their content and their objects.* (London: Routledge and Kegan Paul, 1957).

Gendler, Tamar Szabó / Hawthorne, John (eds) [CP]: *Conceivability and possibility*. (Oxford: Clarendon Press, 2002). [Entry added in 2012.]

Gendler, Tamar Szabó / Hawthorne, John [I]: "Introduction: Conceivability and possibility". In Gendler et al (eds) [CP]:1-70. [Entry added in 2012.]

Grice, Paul / Strawson, Peter F. [iDoD]: "In defense of a dogma". In Grice [SiWo]:196-212.

Grice, Paul [SiWo]: *Studies in the way of words.* (Cambridge / Mass.: Harvard University Press, 1989).

Grünbaum, Adolf [IINP]: "Is it never possible to falsify a hypothesis irrevocably?" In Harding (ed) [CTBR]:260-288. [Entry added in 2012.]

Hales, Steven D. (ed) [CtR]: *A companion to relativism.* (Malden, Mass.: Wiley-Blackwell, 2011). [Entry added in 2012.]

Harding, Sandra G. (ed) [CTBR]: *Can theories be refuted?* (Dordrecht: Reidel, 1976). [Entry added in 2012.]

Hick, John (ed) [EoG]: *The existence of god*. (New York: Macmillan, 1964).

Hick, John [PoR]: *Philosophy of religion*. (Englewood Cliffs: Prentice-Hall, 1963).

Hick, John [TV]: "Theology and verification". In Hick (ed) [EoG]:253-274.

Hirsch, Eli [MNCT]: "Metaphysical necessity and conceptual truth". *Midwest Studies in Philosophy* 11 (1986), pp. 243-256. [Entry added in 2012.]

Hobbes, Thomas [L]: *Leviathan, or The matter, forme, & power of a common-wealth eclesiasticall and civill.* (Oxford: Clarendon 1909).

Hoerster, Norbert (ed) [KPD]/1: *Klassiker des philosophischen Denkens. Band 1.* (München: dtv, 1982).

Kant, Immanuel [KRV]: *Kritik der reinen Vernunft.* Nach der ersten und zweiten Original-Ausgabe neu herausgegeben von Raymund Schmidt. (Hamburg: Meiner, 1926; durchgesehener Nachdruck 1976).

Lewis, Hywel D. (et al.) [LaD]: "Life after death. A discussion. (Anthony Quinton, Hywel D. Lewis, Bernard Williams)". In Hywel Lewis [PLaD]:49-74.

Lewis, Hywel D. [PLaD]: *Persons and life after death. Essays by Hywel D. Lewis and some of his critics.* (London: Macmillan, 1978).

Müller, Olaf [AZ]: "Anders Zweifeln. Eine metaphysische Provokation" [= "The different doubt. A metaphysical provocation"]. Unpublished working paper, Cambridge / Mass., 1997. (For an online version see http://nbn-resolving.de/urn:nbn:de:kobv:11-100204450.)[Entry added in 2012.]

Müller, Olaf [DQDT]: "Does the Quine/Duhem thesis prevent us from defining analyticity? On fallacy in Quine". *Erkenntnis* 48 No.1 (January 1998), pp. 81-99. (For an online version see http://nbn-resolving.de/urn:nbn:de:kobv:11-10066558.)

Müller, Olaf [MSS]: *Metaphysik und semantische Stabilität oder Was es heisst, nach höheren Wirklichkeiten zu fragen. Wirklichkeit ohne Illusionen, Band 2.* [= *Metaphysics and semantic stability, or, What it means to discuss higher levels of being. Reality without illusions, volume 2*). (Paderborn: Mentis, 2003; second part of the habilitation thesis, submitted to Goettingen University in April 2001). [Entry added in 2012.]

Müller, Olaf [SA]: *Synonymie und Analytizität: Zwei sinnvolle Begriffe. Eine Auseinandersetzung mit W.V.O. Quines Bedeutungsskepsis. [= Synonymy and analyticity. A critique of W.V.O. Quine's meaning skepticism].* (Paderborn: Schöningh, 1998).

Papineau, David [KPTW]: "Kripke's proof that we are all intuitive dualists". In Crawford (ed) [PoM]/IV:80-103. See http://www.kcl.ac.uk/ip/davidpapineau /Staff/Papineau/OnlinePapers/Kripke%27s%20Proof.htm (downloaded on August, 7th 2012). [Entry added in 2012.]

Patzig, Günther [P]: "Platon". In Hoerster (ed) [KPD]/1:9-52.

Putnam, Hilary [MLR]: *Mind, language and reality. Philosophical papers, volume 2.* (Cambridge: Cambridge University Press, 1975).

Putnam, Hilary [MoM]: "The meaning of 'meaning'". In Putnam [MLR]:215-271.

Putnam, Hilary [RTH]: *Reason, truth and history*. (Cambridge: Cambridge UP, 1981).

Quine, Willard Van Orman [fLPo]: *From a logical point of view*. (Cambridge / Mass.: Harvard UP, second edition, revised, 1961).

Quine, Willard Van Orman [TDiR]: "Two dogmas in retrospect". *Canadian Journal of Philosophy* 21, No. 3 (September 1991), pp.265-274.

Quine, Willard Van Orman [TDoE]: "Two dogmas of empiricism". In Quine [fLPo]:20-46.

Quine, Willard Van Orman [WO]: *Word and object.* (Cambridge / Mass.: MIT Press, 1960).

Sankey, Howard [ITC]: "Incommensurability and theory change". In Hales (ed) [CtR]:456-474. [Entry added in 2012.]

Schlick, Moritz [MV]: "Meaning and verification". *The Philosophical Review* Vol. XLV, No. 4 (July, 1936), pp. 339-369.

Shoemaker, Sydney / Swinburne, Richard [PI]: *Personal identity.* (Oxford: Blackwell, 1984).

Shoemaker, Sydney [ID]: "Immortality and dualism". In Hywel Lewis [PLaD]:110-131.

Shoemaker, Sydney [SSR]: "Sydney Shoemaker's reply". In Shoemaker / Swinburne [PI]:139-152.

Strawson, Peter [I]: *Individuals. An essay in descriptive metaphysics*. (London: Methuen, 1959).

Swinburne, Richard [PI]: "Personal identity: the dualist theory". In Shoemaker / Swinburne [PI]:1-66

Tetens, Holm [FWMR]: "Folgen die Wissenschaften methodologischen Regeln?" ["Do the sciences follow methodological rules?"]. (Unpublished manuscript 1996).

Williams, Bernard [IS]: "Imagination and the self". In Williams [PoS]:26-45.

Williams, Bernard [PoS]: *Problems of the self. Philosophical papers 1956-1972.* (Cambridge: Cambridge University Press, 1973).

Wright, Crispin [oPPT]: "On Putnam's proof that we are not brains in a vat". In Clark et al. (eds) [RP]:216-241.