11

I Am a Brain in a Vat (Or Perhaps a Pile of Sticks by the Side of the Road)

Jay L. Garfield

Introduction: Modern Western Vats and Ancient Indian Apparitions

There are many ways to think through the relationship between Madhyamaka and Yogācāra. Here I avoid all of the interesting historical and philological issues, and explore a way to bring to bear insights deriving from Madhyamaka and from Yogācāra on both questions about phenomenology and about the philosophy of mind and language. By using the Madhyamaka *catuṣkoṭi* to explore the interpretation of a metaphor from Vasubandhu's *Trisvabhāvanirdeśa* through the lens of a hypothesis discussed by Hilary Putnam, I hope to show that, at least in the context of their deployment in the service of contemporary metaphysics and phenomenology, Madhyamaka and Yogācāra are philosophical allies.

Hilary Putnam, in *Reason, Truth and History* (1981), famously argues that the Cartesian hypothesis that I am a brain in a vat is self-refuting, because inexpressible. The argument is disarmingly simple:

> . . . [T]he "sense data" produced by the automatic machinery do not represent trees (or anything external) even when they resemble our tree-images exactly. Just as a splash of paint might resemble a tree-picture without *being* a tree picture, so . . . a "sense datum" might be qualitatively identical with an image of a tree without being an image of a tree. How can the fact that, in the case of brains in a vat, the language is connected by the program with sensory inputs which do not intrinsically or extrinsically represent trees

> (or anything external) possibly bring it about that the whole system of representation, the langue-in-use, *does* refer to or represent trees or anything external?
>
> The answer is that it cannot. . . .[13]
>
> . . . "[V]at" refers to vats in the image in vat-English, or something related (electronic impulses or program features), but certainly not to real vats, since the use of "vat" in vat-English has no causal connection to real vats. . . . It follows that if their "possible world" is really the actual one, and we really are brains in a vat, then what we now mean by "we are brains in a vat" is that *we are brains in a vat in the image,* or something of the kind (if we mean anything at all). But part of the hypothesis that we are brains in a vat is that we aren't brains in a vat in the image. . . . So, if we are brains in a vat, then the sentence "We are brains in a vat" says something false (if it says anything). In short, if we are brains in a vat, then "We are brains in a vat" is false. So it is (necessarily) false.
>
> The supposition that such a possibility makes sense arises from a combination of two errors: (1) taking *physical possibility* too seriously; and (2) unconsciously operating with a magical theory of reference, a theory on which certain mental representations necessarily refer to certain kinds of things. [14–15]

For the statement "I am a brain in a vat" to be true in English, the word "I" must refer to me, the word "brain" must mean *brain,* and the phrase "a vat" must denote the vat I am in. But these words can only have these semantic values if they and I, as their user, bear the appropriate meaning-inducing relations to their referents. If I am a brain in a vat, these relations cannot obtain, and the words I utter (or take myself to utter) when I say "I am a brain in a vat" do not mean what they would in English. If the words are meaningful, they are false, since they can only mean that I am a brain in a vat if I am not a brain in a vat; were I a brain in a vat, these words would mean at most that I am an illusory brain in an illusory vat. I am, therefore, provably, not a brain in a vat. So argues Putnam.

I will sidestep the vast and interesting literature on this argument in order to juxtapose this argument with an equally famous argument from the Indian Buddhist tradition defending what at least appears to be a version of the brain-in-a-vat hypothesis, Vasubandhu's elephant simile at the end of *Trisvabhāvanirdeśa*.

In doing so, I would like to stir one more Indian ingredient into this philosophical masala, the *catuṣkoti,* or *tetralemma* of Buddhist logic. This form of analysis partitions the logical space determined by any proposition not only into truth and falsity, but into four possibilities, including the possibility of *both* truth and falsity and *neither* truth nor falsity. This rubric is made famous by Nāgārjuna, who uses it to great effect in *Mūlamadhyamakakārikā*. The juxtaposition of these two arguments through this Buddhist logical rubric will shed some light on the relationship between phenomenology, idealism, and semantics. I hope that this methodology suggests a way to deploy jointly texts from what might be seen to be rival Indian schools, and I hope that it reveals that there is enough common ground to justify this deployment.

The elephant simile in Vasubandhu is introduced at the end of *Trisvabhāvanirdeśa* in order to illustrate the relationship among the *parikalpita-svabhāva* (the imagined nature), the *paratantra-svabhāva* (the dependent nature), and the *pariniṣpanna-svabhāva* (the consummate nature). These are the three natures of phenomena articulated in the *Saṃdhimirmocana-sūtra* that form the basis of Yogācāra ontology.

The simile here refers to what appears to be a classical Indian roadside magic show in which (the details are hazy, and my attempts to replicate this feat have failed spectacularly) a magician uses a piece of wood or a pile of sticks as a prop, and somehow—allegedly by the use of a mantra that affects the minds of those in the audience, though it is important to the simile that only the magician really knows how the trick works—causes the audience to see these sticks as a real elephant. On a first, and by now fairly standard exegetical pass (one we will have good reason to reconsider below), in the opening verses of this section (27 and 28), Vasubandhu tells us that our perception of external objects is like the perception of the elephant by the naïve villagers in the show. There is no elephant.

The dagger Macbeth *sees* in act II has an *imagined nature,* namely, as a mind-independent external object; nonetheless, there is no externality or mind-independence in the dagger Macbeth experiences. That imagined nature is purely projected, and is completely unreal, just like the elephant. Macbeth's dagger, however (like that more robust dagger wielded by his lovely wife), has a dependent nature. Its perception depends upon his mind and perceptual apparatus. That nature can either be misperceived as externality, or correctly perceived as ideality (of some kind, to be specified later). This is also the case with the sticks used by the magician: they

can be correctly perceived as a pile of sticks by the side of the road, as a prop for a cool trick, or misperceived as an elephant. The consummate nature of the dagger—that nature Macbeth would see from a fully awakened point of view—is the absence of the imagined in the dependent; that is, the fact that the dagger (again, whether his or hers) exists merely in dependence on mind; it is empty of externality and empty of a dualistic relation to my subjectivity. Just so, the fact that there is no real elephant in the pile of sticks is their actual nature, what is to be understood if one sees through the trick. Now, the (fairly standard) reading I am offering of these verses at this point is patently idealistic, and we shall have reason to complicate it, or even reject it, particularly as we situate these initial verses in the context of those that follow.

There is also a vast literature on Vasubandhu's views, much of which I will also sidestep. But we will need to wade into the complex controversy regarding whether Vasubandhu is best read as an idealist or as a phenomenologist and we will need to consider how to read this complex simile against the background of that dispute. It is by taking seriously his phenomenological side—even if we allow that he may also have been an idealist—that we are able to join the insights of this Yogācārin with those of Madhyamaka.

While I think that it is arguable that in at least some of his work—especially the *Vimśatikā* and the *Triṃsikākārikā*—Vasubandhu defends an idealistic position, it is also plausible that in other texts—particularly the *Trisvsabhāvanideśa*—he does *not,* and that in these texts he can be at least charitably, and maybe even most accurately, read as a phenomenologist. It is, at any rate, this latter reading that I will explore, in the effort to understand how this third-century Indian philosopher can help us to understand the sense in which I am indeed a brain in a vat; the sense in which I am *not* a brain in a vat; the sense in which I am *both* a brain in a vat and not a brain in a vat; and, finally, the sense in which I am *neither* a brain in a vat nor not a brain in a vat. I hope that these readings will provide a way to take Vasubandhu's arguments seriously on their own terms and in the context of contemporary philosophy of mind, epistemology and phenomenology. We will consider each *koti* in turn.

1. I Am a Brain in a Vat

I am a brain in a vat, and obviously so. Let us first consider a perfectly naturalistic reason for this conclusion. We will then extend this to a

phenomenological argument trading on Vasubandhu's analysis in the verses under consideration. What is it to be a brain in a vat? It is for one's brain to be located in a container, hooked up to input devices being controlled by external forces (in Putnam's case, of course, a mad scientist; in Descartes' an evil demon; in Vasubandhu's my karma) that generate my experiences; it is for all of my efferent activity to result in actions or their effects (*karma)* whose reality is only apparent to me through those same afferent pathways; it is to have no *unmediated* access to reality; and it is for all of the *mediation* to be through media *opaque* to my consciousness the veracity of which is impossible to verify independently—as opaque, one might say, as the operations of a magician at a good magic show.

Well, that's what I am, and what I guess you are, too. My brain, as far as I can tell, and as far as the best scientists (real or imagined) tell me, is housed in a vat, often called a *human body*. It is indeed—if I can believe my experience at all—hooked up to input devices (the afferent nerves and blood supply) that are indeed controlled by external forces, including—in Quine's felicitous phrase—sensory irritations, in turn perhaps caused by external objects, the chemistry of my blood, etcetera. And indeed it is only through these afferent pathways that I can have any knowledge of the effects or reality of my own apparent activity. I have no direct *unmediated* knowledge of any reality independent of these sensory inputs, and their actual nature and relation to whatever might lie beyond them is indeed opaque to me. I emphasize that this is not the nattering of nabobs of ontological negativism; it is just plain naturalistic description of the relation between our brains and the rest of the world. So, even on the *least* idealist reading of current science, I am obviously, and uncontroversially, a brain in a vat. And if that is the case, when I say so, I do so *truly*, and if truly, presumably *meaningfully*.

Note—and this is the first reason that Vasubandhu is an important partner in this conversation—that this conclusion is not necessarily idealist. It is neither to deny the materiality of the brain, nor of the vat, nor to deny the reality of the world to which I have only mediated access. There is a tempting way to take this in an idealistic direction: one could argue that the *objects of my experience*—the percepts in my sensory fields, for instance—inasmuch as they are only the inner effects of distal causes about which I know nothing, are purely mental. So, one would argue, in a somewhat Berkeleyan vein—albeit a vein that leads us directly to the more nuanced view articulated by Kant—that nothing I ever know exists externally to consciousness—including, for that matter, my brain, and the vat

that encloses it. And, sowhatever physicality might characterize whatever might exist in some other, unknown way, nothing I encounter is physical.

But we should resist this idealistic temptation, for at least two reasons. First, it begs the ontological question in a subtle but important way. Consider the tulip before me. Even when I grant that the *experience* of the tulip is an inner event, caused proximally by the input to my brain, to argue that the tulip just *is* the experience of it presupposes its conclusion—that perception is not a causal interaction with a distal object but a mere conscious episode. Berkeley may be able to achieve a standoff—maybe—but certainly not victory on this terrain. Second, and more important—and we will have reason to reflect more carefully on this later—as Kant was to argue, what goes for the tulip goes for the percept, too. Just as we cannot treat the external object as a thing known as it is itself, in abstraction from the sensory and cognitive faculties that deliver it to us, we cannot treat our inner experiences as things in themselves known apart from our inner sense, or, as Vasubandhu would call it, our introspective consciousness—*manas-vijñāna*. This, of course, is the central and decisive point made by Kant in the "Refutation of Idealism" in the second edition of the *Critique of Pure Reason*. The asymmetry the idealist needs cannot be established.

The idealist needs a wedge that distinguishes the outer from the inner, giving privileged status to the latter; but all that is forthcoming is a distinction between experience and its object. This distinction is ontologically neutral. Note, for instance, that we can talk in English, as well as in Sanskrit or Tibetan, both of *real* and of *unreal* objects/*artha*/*don* of cognitive or physical acts. One can describe, wish for, or aim at the existent as well as the nonexistent. The act/object distinction instead distinguishes only the subjective from the objective aspects of a cognitive act, enabling an anatomy of experience, but not an investigation of reality. With this distinction between idealistic and phenomenological readings of our vat-confinement, let us return to the initial four verses of the text in question to see what it would be to read Vasubandhu in this way. Moving from plucked tulips to conjured elephants doesn't change much:

27. Like an elephant that appears
Through the power of a magician's mantra—
Only the percept appears;
The elephant is completely nonexistent.

28. The imagined nature is the elephant;
The other-dependent nature is the visual percept;
The non-existence of the elephant therein
Is explained to be the consummate.

29. Through the root consciousness
The nonexistent duality appears.
But since the duality is completely non-existent,
There is only a percept.

30. The root consciousness is like the mantra.
Reality can be compared to the wood.
Imagination is like the perception of the elephant.
Duality can be seen as the elephant.

The deployment of the example in stanza 27 certainly invites an idealistic interpretation of this passage, and I would not want to deny the cogency of that reading. But it does not *force* such a reading. For stanza 27 only sets out the example. In the example, the elephant is nonexistent, because the example is one of a conjuring trick. We need to look further to see how Vasubandhu articulates the analogy. In stanza 28 he tells us that the elephant is analogous to the imagined nature—the *parikalpita-svabhāva/kun brtag rang bzhin*—and so we are to conclude that *that nature* is what is unreal in the same sense that the elephant is unreal in the conjuring trick, and in stanza 30 he specifically identifies the elephant with *duality,* and hence, by transitivity, duality with the imagined nature. So, if we focus specifically on this set of verses, Vasubandhu is arguing only that subject-object *duality* is unreal, and that, just as the mantra causes the elephant to appear, that duality in our experience is caused to appear by our root-consciousness, what we might anachronistically call our neurocognitive processes.

Does Vasubandhu side here with Berkeley in denying that there is *no* reality beyond the imagined, or even beyond its intentional content? Not obviously. Let us parse stanza 30 with care. Part of the causal basis for experience is the root consciousness/our psychological processes, just as part of the basis of the audience's experience of the elephant is the mantra. So far, so good. Nothing idealistic, but nothing on the other side, either. But now we get to reality. To what is it compared? To the pile of sticks. They are certainly material in the analogy, and external to the minds of

the audience. Moreover, they are seen *as sticks* by the magician doing the conjuring—and even by the audience when the spell wears off. Once again, this does not *force* a nonidealist reading. After all, this is only the analogy. Just as the nonexistence of the elephant doesn't force idealism, the existence of the sticks doesn't force materialism.

Nonetheless, if we have not already taken the idealist side, it does open up another possibility, especially in conjunction with the claim that imagination (*rnam rtog/vikalpa)* is like the *perception* of the elephant. Perception has as its material condition, or *de re* object, (*ālambanā/dmigs rkyen) reality* or a pile of sticks, but delivers as its intentional *object (artha/don)* a subject-object duality *absent from reality itself,* or a hallucinated elephant. While the *intentional object* of perception is denied existence independent of the mind, neither perception nor the external world that occasions it is even *interrogated* ontologically. On this purely phenomenological reading, Vasubandhu argues that our ordinary experience involves a confusion of the nature of that experience with the fundamental nature of reality, caused by instinctive cognitive habits of which we are unaware, and leading us to ascribe the subject-object duality we superimpose in consciousness to reality itself as it is independent of that superimposition, thus confusing construction with discovery. This gains further support from a careful reading of stanzas 28 and 29, the verses that link those we have been examining so far:

28. The imagined nature is the elephant;
    The other-dependent nature is the visual percept;
    The non-existence of the elephant therein
    Is explained to be the consummate.

29. Through the root consciousness
    The nonexistent duality appears.
    But since the duality is completely nonexistent,
    There is only a percept.

Here the point to be realized (28) is that there is no elephant at all in reality—that subject-object duality is imaginary, and that it arises (29) through our cognitive processes, in which we confuse a real percept with the unreal structure of subject standing over against object.

Reading Vasubandhu this way, we see him arguing, and indeed arguing persuasively, that I am a brain in a vat. My experience (the dependent

nature, characterized as percepts) is the joint product of a reality that I never directly apprehend (the sticks) and a set of psychological processes that are opaque to me (the mantra, or root consciousness). To the extent that I take my experience to be a direct deliverance of reality, to exist as it appears to me, or to be, *qua* experience, *external to me*, I am simply deceived.

I am, however, to the extent that I am a brain in a vat, *also* a pile of sticks by the side of the road. For, as I have been emphasizing, in foregoing the idealist's distinction of outer versus inner, in turn mapped to real versus unreal, in favor of the phenomenologist's distinction between act and content, my own existence as subject is rendered as problematic as the existence of the object I confuse with an external cause of my experience. Where I seem to come upon a world neatly divided into *me,* the *experiencer,* and *it,* the *experienced,* all I find instead is *experience*. The division into subject and object, and the subsequent reification or deprecation of one with respect to the other, depending on how I take things, is my contribution, not my discovery. So, then, on this view, what am I? I am, independent of my experience, just what the elephant, or the tulip, is: a pile of sticks beside the road that I have never encountered directly, and probably never will.

2. *I Am* Not *a Brain in a Vat*

Putnam, as we have seen, argues that I cannot be a brain in a vat, and does so on semantic grounds, depending specifically on a causal theory of reference. But that argument is unsound, and indeed question-begging. It presupposes that the human body is unlike any vat, and so ignores the fundamental fact of embodiment. It presupposes that the English that Putnam and we speak is not Vat-English, thus taking its very conclusion as a premise. Only a philosophical community befogged by an uncritical realism, and by the view that the only alternative is a mystical idealism, could have been seduced by this particular mantra.

But bad arguments may have true conclusions, and this one does. I am *not* a brain in a vat, but since this conclusion is ontological, we can't get there semantically, a route that, to borrow a phrase from Russell, has all of the advantages of theft over honest toil. Let's first follow the sixth-century Candrakīrti and get there honestly, and then note how Wittgenstein suggests a similar path in *Philosophical Investigations* and *On Certainty*.

In *Madhyamakāvatāra-bhāṣya,* Candrakīrti begins where we left off, noting that any attempt to discredit the reality of external objects yields

arguments that, if cogent, discredit the reality of the self as well. Nonetheless, Candrakīrti does not concede the conclusion he takes the Yogācārins to defend viz., that the mind is *more real* than external objects—that while external objects are entirely imaginary, the mind is *real,* and that it must be, if it is to be that which experiences and imputes reality to an unreal external world. Candrakīrti argues instead that the same arguments that show external objects to be mind-dependent, impermanent, and without any ultimate entity show the mind, or the self, to be mind-dependent, impermanent, and with no entity of its own.

The subject, Candrakīrti argues, is not a unitary thing, but a composite of myriad functions that are themselves composite; not something that can be identified over time independent of our representation of it; and is dependent for its existence and character on innumerable causes and conditions. The self we experience and posit—the referent of the first-person pronoun—he argues, is merely a conceptual, verbal designation on the basis of that causal stream, not even that stream itself. We know ourselves not directly, but only imperfectly, using a conceptually mediated inner sense that is just as fallible as any outer sense. Candrakīrti's refutation of idealism, like that Kant was to write over a millennium later, proceeds, in Kant's words by "turning the game played by idealism against itself," (*Critique* B276) that is, by demonstrating first that idealism is essentially a contrastive doctrine, assigning the mind or the inner world a greater degree of reality than physical objects, or the external world; and, second, that it fails in its attempt to distinguish those degrees of reality.

What does this all have to do with brains in vats? Well, once we see that the essence of idealism is the ontological contrast it draws between mind and the material world, we see that the brain-in-a-vat hypothesis is an idealistic hypothesis par excellence. I, the brain, am real; the world I imagine is just that, purely imaginary. Candrakīrti's analysis bites here just as it does in the context of Buddhist dialectics. On the brain-in-a-vat hypothesis, my access to the external world is dubious, but my knowledge of my immediate cognitive state is secure. I know that I am experiencing a world containing trees and birds, but I do not whether there are indeed trees or birds.

None of this makes sense, though. For if the fact that my knowledge of the external world is mediated makes the epistemic status of the world dubious and renders coherent the claim that it is unreal, then the same goes for my self and my own experience. My knowledge of my own inner states and experience is mediated by my introspective processes. My

representation of myself as a continuing subject of experience requires a conceptual construction of a unity from a multiplicity of cognitive processes and states occurring over time. I have no better knowledge of my inner life than I do of the external world, and no greater assurance of my own reality—if that means the kind of reality that persons have—than I do of that of the external world. So if to be a brain in a vat means to be something assured of its own reality in intimate, veridical contact with its own experience—but with only dubious, mediated access to the external world, which may indeed be nonexistent—I am *not* a brain in a vat.

Of course the resonances of this argument to those of Kant in the *Refutation of Idealism* are strong. But there are also intriguing affinities to important insights of Wittgenstein both in his treatment of self-knowledge in *Philosophical Investigations* and in his discussion of idealism and certainty about the external world in *On Certainty*. In §§305–308 of *Philosophical Investigations,* Wittgenstein notes the ways in which we use our conception of external phenomena as models for understanding the mind, leading us to posit inner mechanisms—mechanisms we neither observe nor whose nature we really understand. The critique of behaviorism and of mechanism in the context of which this observation occurs need not concern us here. But the insight that our self-knowledge is not *immediate,* but is given by an infallible inner sense, is important. We can join this with the Sellarsian insight (1963) that, to the extent that we think of our inner episodes as *significant, as meaningful,* we understand them on the analogy of language. And language, in turn, can only be a public phenomenon, inasmuch as meaning emerges from rule-governed behavior, and rules require communities to constitute them. Therefore, to the extent that we think *anything at all,* or think that we do, we do so in virtue of being members of actual linguistic and epistemic communities. This entails that if I am a mind at all, and if I know myself at all, I am *not* a brain in a vat.

In his consideration of Moore's refutation of idealism in *On Certainty,* Wittgenstein returns to the theme of the social dimension of knowledge. He argues persuasively that, since knowledge is *justified* true belief, and since justification is a social practice that must be learned from others and that is responsible to evidentiary practices and arguments that get their warrant from their reliability and their acceptance by others, knowledge is possible if, and only if, we participate in epistemic communities in the context of a world against which our claims are tested. Moreover, Wittgenstein argues, *doubt* is an epistemic activity that must be learned, and whose felicity conditions are socially and pragmatically determined.

Doubt, moreover, presupposes a background of true beliefs. To doubt a proposition requires one to know how to doubt, what justifies doubt, what it would be for the proposition to be true, and so forth. Genuine doubt is impossible in the context of massive Cartesian error. These epistemic attitudes, like all others, are not individualistically characterized psychological states, but are norm-governed social epistemic practices. Therefore, even to *doubt* that there is an external world presupposes that there is one; and to *know* that I am a brain in a vat presupposes that I am in fact a person among persons whose beliefs are, by in large, true.

So, even to ask whether I am a brain in a vat—an individual nervous system massively deceived into thinking that it is a person among persons, embedded in an external world—presupposes that I am *not* a brain in a vat, that I *am* a person among persons, embedded in an external world. The transcendental *epistemic* conditions of asking the question guarantee that the answer is negative. I am *not* a brain in a vat.

3. I Am and Am Not a Brain in a Vat

So far I have argued soundly that I *am* a brain in a vat and that I am *not* a brain in a vat. So much for the first two *koṭis*. It would seem to follow from this that I *both am and am not* a brain in a vat, and I will now argue that that is so. This conclusion is in fact part of Śāntarakṣita's synthesis of Madhyamaka and Yogācāra in *Madhyamakālaṁkāra*. In that text, Śāntarakṣita argues that Yogācāra provides the best analysis of conventional reality, and Madhyamaka the best analysis of ultimate reality. This conclusion is prima facie surprising. It would appear odd to claim that, *conventionally*, the external world does not exist, and that subject-object duality is an illusion. But this is not so odd if we read Yogācāra not as idealism, but as phenomenology. And it might appear odd to contrast this view of conventional reality with a view of *ultimate* reality according to which all phenomena are *conventionally* real, though ultimately empty. Śāntarakṣita writes:

92. On the basis of Cittamātra,
One should understand the absence of external objects.
On the basis of our system,
One should understand that there is also a complete absence of self.

93. Whoever rides the chariot of these two systems,
Guiding them with the reins of logic,
Will thereby attain the goal,
The realization of the Mahāyāna itself.

rGyal tshab rje comments incisively:

> Consider all phenomena comprised under causes and effects. They are not substantially different from consciousness. This is because they exist in virtue of being experienced through authoritative perception. This entailment is valid because given this premise, they necessarily exist substantially as consciousness. These phenomena should be understood conventionally in this way as merely mind, in virtue of lacking any external realty. But ultimately, even mind does not exist. For ultimately, it has neither a singular nor a manifold nature. [599]

rGyal tshab rejects the inconsistency of *cittamātra*—at least as it is deployed by Śāntarakṣita—with Madhyamaka. The former, he indicates, gives us an analysis of our experience of the natural world (all phenomena comprised under causes and effects) as known to us only through consciousness. The latter shows us that neither object nor subject exists ultimately; there is no contrast possible between their ontological status. This is an apposite development of Śāntarakṣita's insight. Inasmuch as the world we experience is only a world delivered by our consciousness, nothing we immediately experience can be substantially different from that consciousness. But that nondifference from consciousness does not in the end give consciousness a privileged position; both the subject and object side are ultimately known in the same way—through perceptual and conceptual mediation—and exist in the same way—as empty of intrinsic identity.

Śāntarakṣita's synthesis would make no sense if we read Yogācāra as idealism, and indeed his analysis might be the best textual case for the claim that Yogācāra was read in India phenomenologically. It is hard to see how one could either take idealism seriously as an analysis of our ordinary view of reality, or join a doctrine according to which the external world is nonexistent and the mind is substantially existent with one according to which they have an identical status. If we read Yogācāra phenomenologically, on the other hand, Śāntarakṣita's project makes much more sense. A lot of sense, in fact. It is important for any Madhyamaka

account of the two truths that both are *truths,* and that they are consistent with one another. And Śāntarakṣita is a Madhyamaka, after all, even if he is also Yogācārin. Let's take Candrakīrti's account of the conventional nature of things seriously—a thing's conventional nature is the way it appears to ordinary people. Let us, with Śāntarakṣita, take that to be the way things are *experienced.* As we have seen, they are experienced, just as is the elephant in the magic show, only as they appear as delivered by our senses, through input channels opaque to us, shot through with subject-object duality. As far as our experience goes, we are brains in vats; the conventional truth—the everyday world—is the world that Yogācāra phenomenology characterizes in such detail.

But there is no ontology there, no account of the nature of the *objects* of our experience, only of the experience itself. Ontology, for a Mādhyamika, comes at the ultimate level—even if that ontology is itself a recusal from the project of ontology itself. And for Śāntarakṣita, it is at the ultimate level that we find that things have no intrinsic nature—not because they are nonexistent, but because they exist only in dependence on causes and conditions, parts and wholes, and conceptual imputation in the familiar Madhyamaka way. That is, we, and the objects and others with whom we interact, are interdependent realities. We are not, at the ultimate level, brains in a vat.

If I take Śāntarakṣita's synthesis seriously, and if I do so in the context of a phenomenological understanding of the Yogācāra of Vasubandhu and a realistic reading of Candrakīrti's Madhyamaka, I see that I am and am not a brain in a vat. It is true that, from a phenomenological standpoint, I am a brain in a vat. The self I experience, and the objects I experience, are nondually related, but dualistically experienced; they are mere appearances caused in ways I can never know. From an ontological standpoint, however, this view is untenable. I can only make sense of the truth even of the claim that I am a brain in a vat if I am not a brain in a vat. The very fact that I *know* that I am a brain in a vat, or even the fact that I *doubt* that I am not a brain in a vat, shows that I am not.

These facts, moreover, are not reducible to one another. Despite their apparent inconsistency, they are both true. Despite being both true, neither is reducible to the other. From the mere fact of emptiness and conventional reality, one cannot deduce the phenomenological character of experience. Moreover, it is not through an analysis of our experience that we gain an understanding of the fundamental nature of reality, but through ontological analysis. Even if, *per impossibile,* we had substantial

selves, and lived in a world of things with essences, our access to them would be mediated; we would still be brains in vats from that point of view.

The fact that phenomenology and ontology are so independent and yet constitute two indispensible levels of analysis appears to me to be one of Śāntarakṣita's deepest and most original insights. It is not only the basis for his own synthesis, but a promising basis for the project of joining a Madhyamaka metaphysics to a Yogācāra ontology as we bring Buddhism to bear on contemporary discourse.

4. I Neither Am Nor Am Not a Brain in a Vat

One *koṭi* to go. I neither am nor am not a brain in a vat. And this one is simple. To accept the brain-in-a-vat hypothesis as Putnam develops it is to accept a radical idealism or Cartesian skepticism with regard to the external world. And this makes no sense at all. The hypothesis cannot, as we have seen, even be asserted coherently.

But I am also not *not* a brain in a vat. And this is so precisely because I am *not* a brain in a vat. How can this be so? Well, as we have seen, the fact that I am not a brain in a vat amounts to a kind of robust realism about other persons and about the world we inhabit. That robust realism also entails accepting a naturalistic and realistic understanding of my own sensory and cognitive apparatus, and so of the fact that my perceptual and cognitive states—including even my apperceptive and reflexively cognitive states—arise in the familiar opaque way adumbrated in contemporary cognitive science. Given that they do, as Śāntarakṣita and rGyal tshab rJe point out, it can't be that I am not *not* a brain in a vat.

To say *only* that I am *not* a brain in a vat would be to deny my embodiment—to deny that my body is indeed a vat containing my brain, and to deny the disjunction between reality as it is experienced and what reality is independent of experience, to succumb, in other words, to the strongest possible version of the Myth of the Given—the view that the world is given directly to consciousness as it is in itself. To say *only* that I *am* a brain in a vat would be to deny the role that my social and natural context plays in my cognitive life. I am therefore neither a brain in a vat nor am I not a brain in a vat. I am a pile of sticks by the side of the road, experiencing itself as a brain in a vat believing that it is not one, and a non-brain-in-a-vat knowing itself to be one.

5. Mu

Anyone familiar with Madhyamaka dialectics knows that the *catuṣkoṭi* comes in two positive and negative flavors: whatever can be asserted in each limb of the positive *catuṣkoti* can be denied in each of the negative *catuṣkoṭi,* underlying the Madhyamaka claim to thesislessness. So it is with the brain-in-a-vat thesis.

It is not the case that I am a brain in a vat. Why not? Because if I was, I couldn't say so. The conditions for the meaningfulness of my own thought and speech would not obtain. It is not the case that I am not a brain in a vat. Were I not a brain in a vat, I would not be embodied; I would not be a perceiver or a conceiver; I would not be connected to the world or to my own experience via the only channels that can connect me to the objects of my cognition and speech. It is not even the case that I am both a brain in a vat and not a brain in a vat. After all, if I am not a brain in a vat, nor am I a brain in a vat, I certainly can't be both.

Moreover, it is not the case that I am neither a brain in a vat nor not a brain in a vat, and this is both the hardest *koṭi* to grasp and the most profound. At any moment in the dialectic of reflection and experience it is impossible either for me to deny that I am a brain in a vat (after all, reason tells me that I am), or that I am not a brain in a vat (after all, reason tells me that I am not).

I find that on reflection, then, I neither can affirm nor deny that I am a brain in a vat. I can't say anything at all on the matter. My relation to this hypothesis, when considered from this standpoint of *Mu,* of absolute Madhyamaka negation, is inexpressible. And there is a reason for this, which is itself a lesson, and part of the lesson Vasubandhu intends by the analogy with which he closes *Trisvabhāvanirdeśa*: Both the claim that I am a brain in a vat and the claim that I am not a brain in a vat suffer from profound presupposition failure. They begin by presupposing a referent for the first-person pronoun, and then ask about its status.

Who, or what, is this thing that might either be or not be a brain in a vat? It is at least a metaphysical or epistemic *subject,* posited as distinct from, and related somehow to, its object (*ālambana).* Having taken for granted its identity, reality, and distinctness from its objects, we can then ask about its precise status, and use the law of the excluded middle to assert that it either is or is not vat-bound. The set of presuppositions should remind Mādhyamikas of the negative tetralemmas in *Mūlamadhyamakakārikā* regarding causation, emptiness, and the Tathāgata. In these contexts Nāgārjuna argues

that none of the four *kotis* makes sense, precisely because of such presupposition failure.

The same presupposition failure obtains in this case. In posing the vat question, we begin with the presupposition of an unproblematic subject, a subject necessarily distinct from its object, and then pose our question. But that presupposition is equivalent to the presupposition of the reality of the self, and of subject-object duality, and these are the very targets of both Yogācāra and Madhyamaka analysis, each of which is aimed at establishing selflessness and nonduality. The very posing of the question, "Am I a brain in a vat?" then *begs* the question against Vasubandhu and, for that matter, Nāgārjuna (and, for that matter, the entire Buddhist tradition). It does so by presupposing an unproblematic unified subject of experience, precisely the subject that is the target of the dialectic of each of these schools. Despite the fact that Candrakīrti tars Yogācāra with the brush of reification of the self, we see that when we read Vasubandhu—at least in the present text—phenomenologically, they are very much in agreement in this analysis. The negative tetralemma constitutes the reductio on the hypothesis they both reject.

The ensuing thesislessness with regard to the self, issuing from the insight that the self in question is merely a nominal posit and has no independent existence, brings us back to *Trisvabhāvanirdeśa*. Let us return to the final verses:

The imagined nature is subject-object duality. We take the world to be given to us primordially as structured by this duality, but that duality is a superimposition. Indeed, at one level we experience the world in that way, but a deeper phenomenological analysis of that experience shows it to be an illusion of consciousness, not unlike an optical illusion. Just as at one level we see the two lines of the Müller-Lyer illusion as unequal, we know that at another level we see them as equal in length. The duality, just like the inequality, is illusory, and is nothing more than a percept, a way something is taken, not anything given. Subject-object duality, and with it the kind of subjectivity we posit as the presupposition of the brain-in-a-vat question, is, like the elephant, nonexistent, although experienced.

In the magic show, when the spell is broken, the elephant vanishes, and we just see a pile of sticks; in ordinary life, when the spell of self-grasping and the reification of subjectivity is broken, subject-object duality vanishes, and we inhabit a world of impermanent, interdependent phenomena. The dissolution of the experience of objects as dually related

to the self is also the dissolution of the experience of the self, as it is posited only in relation to its objects. To paraphrase (or to torture) Dōgen:

> Being in the world.
> To what can it be compared?
> Sticks, piled by the road.

Is Vasubandhu's purpose, as most exegetes, following Candrakīrti, would have it, the reduction of the external world to the status of a hallucination and the establishment of the hallucinating consciousness as the only reality? It need not be. On this reading, the external world is not rejected; only its *externality* is rejected. That is, it is not the *world* that is nonexistent, but the duality between mind and world. Our *lebenswelt*—the only world we ever inhabit—emerges in full reality not in spite of, but in virtue of, it emptiness of independence, and in virtue of, not in spite of, its constitution through the operation of our sensory and cognitive apparatus.

Is this realistic reading, as some exegetes of cognitive science would have it (Churchland 1976, Metzinger 2003), the reduction of mind and subjectivity to the status of a hallucination, and the establishment of the physical world as the only reality? Again, not necessarily. On this view, psychological phenomena—all of them, including both hallucination and perception—are real natural phenomena. Just like any real phenomena, we experience them only subject to the conditions of our own *real* cognitive processes. But cognitive processes need not be thought of as any different with respect to their degree of reality from any other natural phenomena. So the *reality* of our inner life is not denied, only its *internality*. Just as the world emerges only from mind, mind emerges only from the world.

6. What the Brain-in-the-Vat Argument Does and Doesn't Show

Consideration of the brain-in-a-vat scenario shows us something, but not as much as one might have thought. In particular, it shows us that we need carefully to distinguish a number of perspectives from which the question can be asked, and to nuance and qualify the question accordingly. And Vasubandhu's approach is helpful here. There is a perfectly good sense in which we are brains in vats, and another sense in which we

can't be. These senses are perfectly consistent with one another, and must be integrated if we are to understand the relation between subjectivity and objectivity, and the relationship between our experience and the world we experience—including, I emphasize, our experience of ourselves, and the objective subjects who do the experiencing.

But no version of the argument can lead to the conclusion that there are no real objects of experience. Nor can any version of the argument get to the conclusion that the world we inhabit and with which our thought and language engage exists as we experience it independent of our minds, thought, and language. The argument has nothing to do with reference, and nothing to do with the reality of subjectivity, but everything to do with the complex reciprocal relationship between experience, reality, and being. Consideration of the possibility that we are brains in vats shows us that we are and are not: were we not brains in vats, we could not know anything; were we only brains in vats, there would be nothing to know.

7. Being in the World: What the Elephant Simile Does and Doesn't Show

Vasubandhu's elephant simile also shows us something, but also less than one might have thought, and Putnam's approach is helpful here. The simile does not show us that there is *no* external world; nor does it show us that there *is* one. It is not even aimed in that direction. Instead it illustrates the complex nature of our subjectivity, a subjectivity in which at the most basic level we inhabit a world in which the distinctions between subject and object, internal and external, are entirely absent; but a subjectivity that also systematically mis-takes that world to be saturated with that very duality in virtue of cognitive processes that, in a kind of cognitive reflex, superimpose that structure at a higher level on an experience that does not present it at a more primordial level. That is the conjuring trick. We systematically deceive ourselves about the nature of our own experience, and hence about the world in which we live. But it is a deception through which we can learn to see.

The point is not that when we see elephants in the street (at least when we do so in *India*) we are really hallucinating; that there are only piles of sticks beside the road and that only our experience—the hallucination—is real. That is the idealistic reading of the simile I urge us to reject. Vasubandhu's point is that when we see elephants in the road, that

experience is multilayered. In naïve introspection, we take both our own subjective state and the objectively presented pachyderm to be presented to us as they exist, related to one another as experiencing subject and experienced object. But this is a mis-taking not only of the elephant, but also of ourselves, and of the structure of the experience at a more primordial level. At that more fundamental level, the elephant we perceive on the road is a conceptual-perceptual construction wrought by our sensory and cognitive apparatus in response to stimulation; our subjectivity is constructed by apperceptual processes, and the duality we project in which we take ourselves simultaneously to be aware of self and other as distinct entities in this experience is itself constructed. That is the *conjured elephant*. We, the elephant, and the moment of experience are all sticks in a pile by the side of the road. This is a conclusion in which Mādhyamikas and Yogācārins happily concur, and on which their analyses nicely converge.

Vasubandhu and Putnam each invite us to interrogate the structure of experience and the structure of Dasein. The figure of the brain in the vat is meant by Putnam to demonstrate the necessity of the existence of the external world as we perceive it; that of the illusory elephant is often taken to deny the external world's existence at all. Instead, I have argued, each calls upon us to challenge neither the reality nor the illusory character of the objects we perceive, but rather our instinctive view that they, we, and our experience of our own being are given to us just in the way that they exist, or that anything ever could be.

Appendix: Final Verses of Vasubandhu's Trisvabhāvanirdeśa (Treatise on the Three Natures)

27. Like an elephant that appears
Through the power of a magician's mantra—
Only the percept appears;
The elephant is completely nonexistent.

28. The imagined nature is the elephant;
The other-dependent nature is the visual percept;
The non-existence of the elephant therein
Is explained to be the consummate.

29. Through the root consciousness
The nonexistent duality appears.
But since the duality is completely non-existent,
There is only a percept.

30. The root consciousness is like the mantra.
Reality can be compared to the wood.
Imagination is like the perception of the elephant.
Duality can be seen as the elephant.

31. When one understands how things are,
Perfect knowledge, abandonment,
And accomplishment—
These three characteristics are simultaneously achieved.

32. Knowledge is non-perception;
Abandonment is non-appearance;
Attainment is accomplished through non-dual perception.
That is direct manifestation.

33. Through the non-perception of the elephant,
The vanishing of its percept occurs;
And so does the perception of the piece of wood.
This is how it is in the magic show.

34. In the same way, through the non-perception of duality
There is the vanishing of duality.
When it vanishes completely,
Non-dual awareness arises.

35. Through perceiving correctly,
Through seeing the non-referentiality of mental states,
Through following the three wisdoms,
One will effortlessly attain liberation.

36. Through the perception of mind-only
One achieves the non-perception of objects;
Through the non-perception of Objects
There is also the non-perception of mind. (Garfield 2002)

References

Churchland, P. (1976). *Scientific Realism and the Plasticity of Mind*. Cambridge: Cambridge University Press.

Garfield, J. 1997/2002. "*Trisvabhāvanirdeśa*: A Translation and Commentary," *Asian Philosophy* 7: 2, 133–154, 1997, reprinted in *Empty Words*, New York: Oxford University Press, 2002, pp. 128–151.

Lusthaus, D. 2002. *Buddhist Phenomenology: A Philosophical Investigation of* Yogacara *Buddhism and the Ch'eng Wei-shih Lun. Curzon Critical Studies in Buddhism Series*. London: Routledge.

Metzinger, T. (2003). *Being No One*. Cambridge: MIT Press.

Putnam, H. 1981. *Reason, Truth and History*. Cambridge, MA: Harvard University Press.