

Two Notions of Mental Representation

Uriah Kriegel Arizona

Forthcoming in U. Kriegel (ed.), *Current Controversies in Philosophy of Mind*, Routledge

Introduction/Abstract

The main thesis of this paper is twofold. In the first half of the paper, (§§1-2), I argue that there are two notions of mental representation, which I call *objective* and *subjective*. In the second part (§§3-7), I argue that this casts familiar tracking theories of mental representation as incomplete: while it is clear how they might account for objective representation, they at least require supplementation to account for subjective representation.

1. A Parable

There is a possible world where, just as I was born, a brain neuroanatomically and neurophysiologically indistinguishable from mine was placed in a vat and fed random sensory stimulations by a machine suitably hooked to the vat. In fact, there are many, many such worlds. In one of them, the influx of sensory stimulation happens to be indistinguishable from the one my actual brain has enjoyed. Consequently, let us suppose, it is impossible to rule out from the inside that I am in fact such an envatted brain: the envatted brain's stream of consciousness is subjectively indistinguishable from mine. Thus whenever the stimulating machine is in state S1, the envatted brain undergoes an experience subjectively indistinguishable from the experience I normally undergo when I see an apple; when the machine is in state S2, it undergoes an experience indistinguishable from mine when I see a banana; when the machine is in S3, it undergoes an experience like mine when seeing a cherry; and so on and so forth.

An interesting question concerns what the envatted brain's S1-caused apple-ish experience represents. There are two conflicting views on this. A traditional view is that it represents exactly what my subjectively indistinguishable experience does: an apple. But following Putnam's (1981 Ch. 1) ingenious discussion, many philosophers have come to hold that it represents S1, the machine's state responsible for the apple-ish stimulation. The idea is, very roughly, that since the condition of the external world that stably covaries with and/or causes the brain's apple-ish experiences is S1, S1 is what experiences of that type represent in the envatted brain.

Which view is right? Does the brain's experience represent an apple or S1? My starting point is that this is really quite a silly question. There are two notions of representation, one on which the envatted brain's experience represents an apple and one on which it represents S1. Both are legitimate uses of the term 'representation': on the one hand, there is certainly a sense in which the experience represents S1, namely, the sense that the experience tracks the presence of S1; on the other hand, there is also a sense in which the experience represents an apple, namely, the sense that what the experience presents to the subject is apple-ish.

Thus the true moral of the brain-in-vat thought experiment, it seems to me, is that a distinction must be drawn between two notions of mental representation – two different senses in which a mental state may be said to represent. The experiment is needed because in ordinary circumstances it is hard to distinguish the two notions. In my own experience of the apple, for example, what the experience presents to me and what it tracks are the same: an apple. It is only in the fantastic circumstances of the thought experiment that an experience can be envisaged which presents to the subject one thing but tracks another.

2. A Distinction

There are no good terminological options for drawing the distinction brought out intuitively in the brain-in-vat thought experiment. I will use the labels 'the objective notion of representation' and 'the subjective notion of representation,' or *objective*

representation and *subjective representation* for short. These terms are in many ways sub-optimal, but they will have to do.¹

How to characterize the distinction in a more theoretically involved manner is, in a way, what this paper is about. For this reason, it would be unwise to prejudge certain issues by building commitments into the *definitions* of objective and subjective representation. Admittedly, we do need some way to fix our ideas regarding these two notions. But we can do so without unnecessary commitments through ostension of paradigmatic instances. One way to think of this is as offering a(n initial) model of the two notions not as necessary-and-sufficient-conditions notions, but as *prototype* or *exemplar* notions. We designate prototypical or exemplary objective and subjective representations – e.g., the envatted brain’s – and consider a given mental representation objective and/or subjective if it is sufficiently relevantly similar to the prototype or exemplar.²

The distinction between objective and subjective representation is closely related to other, more familiar ones. Distinctions between personal and sub-personal representations, narrow and wide representations, phenomenal and psychological representations, may all turn out to be co-extensive with the subjective/objective distinction.³ Even if they do, however, this should not be, and is not here, taken to be *definitionally* true. It is not *definitional* of subjective representation that it is personal, narrow, and phenomenal – nor of objective representation that it is sub-personal, wide, and psychological. My view is that the distinction between objective and subjective representation is deeper than all these, and underlies them, but I will not argue for this here.

Instead of arguing that the distinction is deep, I will now argue that it is *thorough*. In the brain-in-vat thought experiment, (a) representation varies in the objective sense but remains invariant in the subjective sense. We can also envisage circumstances in which the converse holds: (b) representation varies in the subjective but not objective sense. We can even envisage circumstances in which (c) representation occurs *at all* in the objective sense but not in the subjective sense, and ones in which (d) it occurs in the subjective but not objective sense. This is what I mean by the distinction being ‘thorough.’

The inverted earth thought experiment (Block 1990) exemplifies (b). There is a possible world just like ours except that (i) every surface's and volume's color is complementary relative to its color in the actual world and (ii) my counterpart's spectrum is inverted relative to mine. Because the complementation and the inversion cancel each other out, my counterpart's color experiences present to him exactly what mine present to me, even as they track different properties of surfaces and volumes. Therefore, Inverted-Earthlings' experiences represent one thing in the objective sense, another in the subjective sense.

There are familiar examples of (c) – representation in the objective sense in the absence of representation in the subjective sense. The number of rings on a tree trunk tracks the tree's age, but does not present the age *to* the tree; thermometers' internal states track the ambient temperature, but do not present the temperature *to* the thermometers; and so on.

There are no familiar examples of (d), but consider the following thought experiment. We can envisage a world where the only concrete particular is a disembodied soul 'floating about' in otherwise empty space, undergoing a random string of conscious experiences. In fact, we can envisage many, many such worlds. In one of them, the space soul's sequence of experiences is subjectively indistinguishable from yours. In a sense, it is impossible to rule out from the inside that you *are* in fact that space soul.⁴ The space soul's experiences present to it exactly what yours present to you, but unlike your experiences, the soul's do not track anything.⁵ Therefore, the space soul's experiences represent in the subjective sense but not in the objective sense.

This thought experiment is similar to the brain-in-vat one, in that it too dissociates what an experience presents to its subject from what it tracks in the environment. But while in the brain-in-vat scenario the experiences still track *something*, in the space-soul scenario there is nothing for them to track. In a way, while the brain-in-vat thought experiment shows that the *identity conditions* of subjective representation are independent from those of objective representation, the space-soul thought experiment shows that the *existence conditions* of subjective representation are also independent from those of objective representation.

3. Familiar Theories of Mental Representation

There are, in fact, many different notions of representation. We say of a reflection in a mirror or a puddle that it is an *imagistic representation* of some object or surface; we say of a graph or diagram that it provides a *mathematical representation* of some pattern; we say of a rainbow metaphor that it constitutes a *literary representation* of hope; and so on and so forth. There may well be a feature common and peculiar to all these kinds of representation, in virtue of which they all deserve the appellation ‘representation,’ but there are deep differences among them as well.⁶ Most importantly, the nature of the representation relation implicated in each is very different. It is not as though literary and mathematical representations bear the same representation relation to their subject matters, with the former qualifying as literary simply because their subject matter is literary and the latter as mathematical because their subject matter is mathematical.⁷ Rather, the very representation relation they bear to their subject matters is plausibly different.

My contention is that even within the realm of *mental* representation, we can distinguish two different senses in which mental items may be said to represent. For what makes a mental state track what it does, and track at all, is very different from what makes it present to its subject what it does, and at all. The feature in virtue of which mental states represent in one sense is completely different from that in virtue of which they represent in the other – and this is not a matter of different subject matters, since the subject matter can and often is actually the same.

If this is right, then seeking ‘a theory of mental representation,’ in the sense of a unified framework that accounts for a single relation which determines what, and that, mental states represent, and then attempts to account for it, may be as misguided as seeking a unified theory of literary and mathematical representation. What invites this misguide, it seems to me, is the fact that in this case the two fundamentally different kinds of representation are exhibited by the same vehicles: mental states. Ultimately, however, each notion requires its own account.⁸

The problem is that the theories of mental representation we have pursued most vigorously over the past forty years – those that fall under the rubric of ‘naturalistic

semantics' or 'psycho-semantics' – seem geared to account for the objective notion of representation exclusively, disregarding the subjective notion. I will now illustrate this by going through some of the most prominent options, which tend to fall into two groups: causal-covariational theories and teleological theories.

In its simplest manifestation, the causal-covariational approach claims that a mental state *M* represents a property *F* just in case *F*s cause *M*s under the right conditions (Stampe 1977). This kind of causal relation is most natural to appeal to in accounting for the tracking of external conditions. Observe, however, that the envatted brain's apple experiences do not have apples as their causes *ever* – i.e., under *any* conditions. Yet in the subjective sense what they represent are apples. So this approach seems wrongheaded as an account of subjective representation. Note well: my present point is not quite that the causal-covariational approach lacks in principle the resources to account for subjective representation; rather that it is not geared to doing so, and appears to target instead (and quite plausibly) objective representation.

The best-known version of the causal-covariational approach is probably Fodor's (1990) 'asymmetric dependence' account: a mental state *M* represents a property *F* iff (i) it is a law of nature that *F*s cause *M*s, (ii) some *F*s actually cause *M*s, and (iii) if any non-*F*s cause *M*s, the fact that they do is asymmetrically dependent upon the fact that *F*s cause *M*s.⁹ Again, this may be a promising account of representation in the objective sense, but not so much in the subjective sense. At the very least, condition (ii) is flagrantly violated by the envatted brain's apple experiences: none of them are caused by apples.¹⁰

A different account in the same spirit is Dretske's (1981) early informational semantics. According to it, *M* represents *F* iff *M* is nomically dependent upon *F*, that is, iff the laws of nature are such that *M* is not tokened unless *F* is instantiated. This comes very close to the standard account of tracking in reliabilist theories of epistemic justification (indeed, see Dretske 1971), and is thus a good candidate for a theory of representation in the objective sense. But as a theory of representation in the subjective sense it seems utterly inadequate: the envatted brain has experiences as of apples even when the property of being an apple is not instantiated. So it is false that they are not tokened unless the properties they (subjectively) represent are instantiated.¹¹

The other familiar approach to mental representation is so-called teleosemantics. The idea, roughly speaking, is that a mental state represents in virtue of conferring the right kind of adaptive or reproductive advantage on the subject, or more accurately, in virtue of its ‘correspondence’ with external conditions conferring this kind of advantage.¹² Here again, there may well be much to recommend evolutionarily grounded tracking as an account of objective representation, but the prospects for such an account of subjective representation are on the face of it bleak. After all, if the envatted brain reproduces at all, it is certainly not in virtue of any correspondence between its apple experiences and apples, since there is none.¹³ Note well: here again, I am not presently concerned to argue that teleosemantics is in principle incapable of accounting for the subjective notion of representation; merely that clearly it is designed to account for the objective notion.

Consider Dretske’s (1988) mature theory, which augments his original informational theory with a teleological component. According to the augmented theory, M represents F not iff M nomically depends upon F, but iff M is *supposed* to nomically depend upon F, where this means that (i) there is a motor response R, such that M has been recruited (through a process of ‘discrimination learning’) to have its present tokens cause R, (ii) past tokens of M nomically depended upon F, and (iii) it is the case that (i) because it is the case that (ii). This fails to accommodate the envatted brain’s apple experiences in several ways. First, the envatted brain does not *have* motor responses (at least if motor responses are construed as bodily states), though it may seem to itself to have them. Secondly, past tokens of the envatted brain’s apple experiences did *not* nomically depend upon any past instances of applehood, since there were no such instances. Finally, condition (iii) could certainly not be met, since there is no *reason* why the envatted brain has present token apple experiences – it is a pure accident.¹⁴

Another version of teleosemantics is Millikan’s (1984, 1993) ‘biosemantics.’ Millikan’s account is quite complex and not easily summarizable in a single cognitively surveyable biconditional, but a central *necessary condition* in the account appears to be this: M represents F only if there is a system S, such that (i) S consumes present tokens of M, (ii) past tokens of M occurred mostly when instances of F occurred, and (iii) S can perform its biological proper function because (i) and (ii) are the case. This necessary

condition is quite obviously not met by the envatted brain's apple experiences. For starters, it is unclear how we might attribute a biological proper function to the brain's consumer system, since the brain faces no selection pressures. But more obviously, the envatted brain's past token apple experiences did *not* occur mostly when instances of applehood did.

I conclude that both causal-covariational and teleological approaches to mental representation are geared towards its objective notion. The teleo-informational materials they employ are perfectly suited to account for the tracking of external conditions. (For this reason, I will often refer to them in what follows as 'tracking accounts.')

But to the extent that mental states can sometimes present to their subject something they do not track, it is not clear that these materials are well-suited to capture what mental states present to their subject. To repeat, my present point is *not* that no broadly teleo-informational story could ever accommodate subjective representation. It is rather that the familiar theories in that genre are so unnatural as accounts of subjective representation that it is most reasonable, and most charitable, to interpret them as not even *concerned* with subjective representation. This is doubly plausible given how genuinely promising they look as accounts of *objective* representation.

4. Notions and Properties

Agreeing that the familiar theories of mental representation in the teleo-informational genre are geared towards objective representation, but wishing to maintain that they fully account for all the relevant phenomena without need of supplementation, one might exploit the gap between notions and properties. After all, so far I have only argued for a *conceptual* distinction between two kinds of notions. I have not argued for a *real* distinction between two kinds of properties.

At the same time, there is a trivial sense in which a conceptual distinction creates a *prima facie* reason to believe in a real distinction. Consider again the envatted brain's apple-ish S1-caused experience. Suppose some interpreter claims that the experience represents an apple in the subjective sense, while another interpreter claims that it represents an elephant in that sense; it is natural to think that the first interpreter's claim

is true whereas the second's is false.¹⁵ Likewise, suppose one interpreter claims that the experience represents S1 in the objective sense while another claims it represents S27; here it is natural to say that the first claim is true and the second false. Thus we have here two different true ascriptions and two different false ones. However, if two different ascriptions are true here, then there are two different properties picked out by the notions deployed in those ascriptions; each putative property constitutes (or is a constituent of) the truthmaker of a different true ascription.¹⁶ In this way, the conceptual distinction between notions creates a defeasible presumption in favor of a real distinction between properties, at least against the background of intuitive applicability of truth values to claims made using both notions.

Given a presumptive distinction between two putative properties, several views on the ultimate relation between the two are possible. Three stand out: eliminativist, reductivist, and nonreductivist. On the eliminativist view I have in mind, ultimately there is no such property as subjective representation. The concept may be useful, but nothing in reality matches it. Accordingly, subjective representational ascriptions are forsooth never strictly true except in a minimalist sense.¹⁷ On the reductivist view I have in mind, there *is* such a property as subjective representation, but ultimately it reduces to objective representation, or some complex objective-representational structure.¹⁸ Accordingly, subjective representation is an 'ontological free lunch,' one fully grounded in objective representation. On the nonreductivist view, by contrast, subjective representation is a real and genuine ontological addition over and above objective representation. (This is not to say, of course, that it is irreducible to *physical* properties; merely that if it does, it is not by reducing first to objective representation.) The choice can be appreciated by considering the following inconsistent triad:

- 1) There exists subjective representation.
- 2) Subjective representation does not reduce to objective representation.
- 3) A theory of objective representation accounts for all the phenomena of mental representation.

Each of these is individually plausible, but they cannot all be true. Accordingly, any stable position on the matter must reject one of them. The eliminativist rejects 1, the reductivist rejects 2, and the nonreductivist rejects 3.

The proponent of familiar tracking theories of mental representation should argue for the eliminative or reductive view. If either turns out right, such familiar theories may be in good shape. The nonreductive view is the least conservative of the three, and would straightforwardly require that familiar theories be supplemented with a distinct account of subjective representation.

Importantly, there are also more radical epistemic possibilities in the area. One is eliminative in the opposite direction, denying the existence of a property of objective representation. Another is reductive in that opposite direction, claiming that objective representation ultimately reduces to, or is somehow grounded in, subjective representation. If either of these more radical options prevails, familiar tracking theories of mental representation would be cast not just as inadequate, but as in an important sense wrongheaded.

My primary goal here is to argue that there is a notion of mental representation which familiar tracking theories are ill-equipped to account for. With this in mind, I turn to address the eliminative and reductive views of subjective representation in the next two sections. I will argue that eliminativism is implausible, and that any reductivism that might accommodate subjective representation would involve *something* importantly *unfamiliar*.

5. Eliminativism

Eliminative positions are typically motivated by considerations of explanatory dispensability. Thus our eliminativist may argue that subjective representation would not explain anything, so there is no need to posit it.

This eliminativist tack may be resisted by denying the explanatory impotence of subjective representation.¹⁹ But more deeply, argumentation from explanatory dispensability presupposes a description of what needs to be explained. And the very description of the explananda is typically ontologically committal. Thus *explanatory* dispensability can support eliminativism about *x* only when combined with *descriptive* dispensability – the claim that there is no need to invoke *x* in describing what needs explaining. Eliminating subjective representation would thus require such a claim. Yet it

is unclear how we can *describe* the scenario presented in the opening parable without mentioning a kind of representation shared by subjective duplicates. What needs to be explained in that scenario is precisely how an envatted brain stimulated identically to me is experientially presented with an apple. The very description of the explanandum thus invokes the subjective notion of representation.

In light of this consideration, the eliminativist may wish to call into question the possibility of the scenario presented in the opening parable. In conceiving the scenario, we conceive of a mental state whose representational properties in the subjective sense are different from its representational properties in the ‘objective’ sense. It is to describe this conceived scenario that we introduce the notion of subjective representation. But the eliminativist about subjective representation can resist the introduction of the notion by claiming that although the scenario is conceivable, it is not genuinely possible – it is not *metaphysically* possible. The view may be that, as a matter of Kripkean a posteriori necessity, the envatted brain’s S1-caused apple-ish experience represents only S1; there is no sense in which it represents an apple.

Regardless of whether conceivability is generally a good guide to metaphysical possibility, however, the present objection is completely implausible. For an identically stimulated envatted brain ought to be not only metaphysically but *nomologically* possible. Consider that conscious experience is widely acknowledged to have *neural realizers*, and it is certainly possible, consistently with the actual laws of nature, to duplicate the neural realizers of my stream of consciousness by duplicating exactly (i) the neural state of my brain at the beginning of my biological life and (ii) the sensory inputs I have enjoyed since. Everything we know about brain function suggests that this is nomologically possible. The quality of the link between conceivability and metaphysical possibility is irrelevant.²⁰

Conceding that the scenario from the opening parable is metaphysically, indeed nomologically, possible, the eliminativist may insist that the property it requires us to posit does not qualify as representational *in the philosophically relevant sense*. On this version of eliminativism, the envatted brain’s state does present an apple to it, but it does not *represent* an apple in any philosophically relevant sense. (This certainly smacks of a verbal issue, but set that aside.)

Whether this kind of objection is plausible depends, of course, on what we require from a property to qualify as representational in the ‘philosophically relevant’ sense. If there is any substantive answer to the question at hand, I think it would have to appeal to the traditionally central idea of *intentionality*. The thought would have to be that ‘the philosophically relevant sense of representation’ is the *intentional* sense, that is, the sense in which representation exhibits the features definitive of intentionality. Two of these stand out: (i) the feature that underlies failure of truth-preserving existential generalization; (ii) the feature that underlies failure of truth-preserving substitution of co-referential terms (Chisholm 1957).

It seems clear, however, that representation in the subjective sense exhibits both features. Thus, the envatted brain’s apple experience presents an apple to the brain even though there is no apple, and its morning-star experiences are ‘presentationally different’ from its evening-star experiences even though the heavenly body is one and the same.²¹ Accordingly, from ‘the envatted brain’s experience presents an apple to the brain’ we cannot truth-preservingly infer ‘there is something that the envatted brain’s experience presents to it’ (failure of existential generalization), and from ‘the envatted brain’s experience presents the morning star to the brain’ and ‘the morning star is the evening star’ we cannot truth-preservingly infer ‘the envatted brain’s experience presents the evening star to it’ (substitution failure). So, subjective representation does exhibit the definitive features of intentionality, and thus qualifies as representation in the philosophically relevant sense.²² In other words, there is such a thing as *subjective intentionality*.²³

I conclude that eliminativism is *prima facie* highly implausible. It is true that the combination of familiar tracking theories of mental representation and a compelling argument for eliminativism about subjective representation would protect these theories’ status as fully adequate to the phenomena. But it is far from obvious what such an argument would look like. Arguments to the effect that subjective representation is explanatorily and descriptively dispensable, metaphysically impossible, or philosophically irrelevant do not appear to work. Another argument might, but it is the burden of the tracking theorist to produce it. Observe that this is, essentially, the burden

to supplement her tracking theory with a compelling argument for eliminativism about subjective representation.

6. Reductivism

The reductivist gambit in this area is to develop a broadly causal-covariational or teleological account of subjective representation. In doing so, she would show that the familiar tracking theories of mental representation, even if not *geared* toward subjective representation, have the *resources* to account for it.

It might be suggested, for example, that a theory of subjective representation can be obtained simply by taking the best among the familiar theories of objective representation and adding to it a condition of subjective accessibility. Thus, one might hold that a mental state *M* presents a feature *F* to its subject iff (i) *M* tracks *F* and (ii) *M*'s tracking of *F* is somehow introspectively accessible to the subject. If this is right, then familiar theories of mental representation only need to be supplemented with an account of introspective access to accommodate subjective representation.

The problem with this suggestion is twofold. First, for the reductivist gambit to succeed, it would need not just any account of introspective access, but one in terms of objective representation. Secondly, however we account for introspective access, the emerging account appears to be falsified by the brain-in-vat scenario: the envatted brain has a mental state that presents an apple to it, even though it is not the case that (i) it tracks an apple and (ii) its tracking of the apple is introspectively accessible to the brain. Condition (i) is not met.

It might be suggested that both problems could be overcome by a higher-order representation account of introspective access combined with a tracking account of higher-order representation. This would result in a picture whereby first-order trackers of the environment are themselves tracked by second-order trackers, and the subjective representational content of a mental state is determined by the content of these higher-order trackers. Suppose *M* objectively represents *F*, but is higher-order objectively represented to represent *G*. Then on this view, although *M* objectively represents *F*, it subjectively represents *G*. This accounts for the independence of objective and subjective

representation. At the same time, M's subjective representational content is fully accounted for in terms of objective-representation relations: since higher-order tracking is itself a form of objective representation, subjective representation is fully grounded here in objective representation.

This reductive approach to subjective representation is much more promising, but it does face two important challenges. First, the extant literature on naturalistic theories of mental representation includes, to my knowledge, no higher-order tracking account.²⁴ In this respect at least, such an account would be somewhat *unfamiliar*. Thus in an important sense producing such an account to accommodate the subjective notion of representation would effectively constitute *supplementation* of the existing familiar theories. More pressingly, it is far from clear how higher-order tracking could deliver the right results in brain-in-vat scenarios. To account for the brain's experience presenting an apple to it, this reductivist would have to say that the brain harbors a higher-order state that tracks an apple-tracker. But it is not at all clear how the brain might *acquire* this first-order apple-tracker. After all, the acquisition of apple-trackers depends on causal interaction with apples.²⁵ On the face of it, the only features of relevance that the brain's internal states track are sensory stimulations. Recall, moreover, that this brain is disembodied – unlike our brains, it is not 'coated' with a genuine sensorium. Thus the sensory stimulations of relevance would have to be construed as stimulations of entry points to the brain itself, e.g., the lateral geniculate nucleus (LGN) for visual stimulation. The brain may well develop internal states that systematically track certain apple-appropriate states of its LGN, thus acquiring a first-order LGN-tracker. But then any higher-order state that would track these would *not* be tracking an apple-tracker; it would be tracking an LGN-tracker. The account of subjective representation in terms of higher-order objective representation would then return the result that the envatted brain represents, in the subjective sense, its own LGN. Yet the needed result is that it represents an apple.

There may, of course, be some other, more plausible reductive account that would cast subjective representation an ontological free lunch given the existence of objective representation. But as with eliminativism, it is the reductivist's burden to provide such an account. A reductivist defense of familiar, tracking-based theories of mental representation as fully adequate would, in any case, require two parts: (i) an account of

objective representation in terms of tracking relations and (ii) a reductive account of subjective representation in terms of objective representation. The extant literature contains admirable work toward (i), but virtually no work toward (ii). I contend that the phenomena of subjective representation cannot be dealt with simply through disregard; some positive account of them – if only a reductive account – is called for.

7. Nonreductivism and More Radical Options

A subjective representation enthusiast may turn the table on the proponent of familiar tracking theories by arguing for eliminativism or reductivism about *objective* representation. Recall that one of the eliminativist tacks above involved denying that the property shared by my envatted subjective duplicate qualifies as representational in the ‘philosophically interesting’ sense, presumably the sense of intentionality. A parallel claim could be made by the subjective representation enthusiast. She can claim that so-called objective representation does not exhibit the features underlying failure of substitution and existential generalization.²⁶

The notion that objective representation does not exhibit the feature underlying substitution failure has already been foreshadowed in the literature. Some arguments due to Searle (1991, 1992 Ch.7) and Loar (1995) could certainly be adapted in this direction. The thought is that any mental state that tracked Phosphorus would *eo ipso* be tracking Hesperus. Tracking relations, even teleologically augmented, simply cannot discriminate between coextensive entities – for reasons explored already a generation ago (see esp. Fodor 1984). In consequence, ‘M tracks F’ and ‘F=G’ *do* entail ‘M tracks G’ – contrary to substitution failure.²⁷

There are also arguments to the effect that objective representation does not exhibit the feature underlying failure of existential generalization (see Kriegel 2007, 2011 Ch.3). Here the idea is that tracking relations, even when teleologically augmented, cannot obtain in the absence of their relata. Imagine that one of the worlds in which the envatted brain happens to be a subjective duplicate of me there simply do not exist tables, chairs, trees, and flowers; then it would seem that my envatted duplicate’s internal states cannot track – and hence, cannot objectively represent – tables, chairs, trees, and flowers

there. Thus tracking relations require the existence of the tracked. In consequence, ‘M tracks F’ *does* entail ‘There is an x , such that M tracks x ’ – contrary to failure of existential generalization.

Such arguments (which admittedly would require more sustained development) might inspire some to go eliminativist with respect to objective representation. The claim would be that objective tracking relations do not qualify as representational in the philosophically interesting sense – they are not *intentional*. Others may take the same considerations to suggest a *reductivist* account of objective representation. Here the idea would be that tracking relations qualify as representational (in the philosophical sense) only in virtue of bearing the right relation to subjective representation. Again following Searle’s (Ibid.) lead, one might suggest that M objectively represents F iff (i) M tracks F and (ii) M *potentially* subjectively represents F (or M is *disposed* to subjectively represent F). More circuitously, and now following Loar’s (2003) lead, one might hold that M objectively represents F iff (i) M tracks F and (ii) M is functionally/inferentially integrated with subjective representations. Other reductive accounts could also be suggested (see Kriegel 2011 Ch.4 for relevant discussion). What they would all have in common is the claim that tracking relations, however sophisticated, only qualify as representational in the relevant (read: intentional) sense if they bear the right relation to subjective representations. Objective representations thus inherit their status as representations *from* subjective representations. In this way, objective representation is grounded in subjective representation – without the latter there cannot be the former.

The most antecedently neutral approach would be nonreductivist, rejecting eliminativism and reductivism about either objective or subjective representation. On this view, the conceptual distinction between two notions of representation is paralleled by a real distinction between two mutually irreducible kinds of representational property. Importantly, such nonreductivism does constitute a departure from familiar tracking theories of mental representation, as it invites supplementation of such theories by some account of subjective representation. If we cannot reduce subjective representation to objective representation, a comprehensive account of all the phenomena of mental representation would require a theory of subjective representation.²⁸

Such a theory would tell us what it is in virtue of which a mental state M subjectively represents a feature F. Just as familiar theories of mental representation designate tracking relations as that in virtue of which a state objectively represents a feature, so some relation (or perhaps a monadic property) would have to be designated as similarly underlying subjective representation. One way to understand the recent flurry of work surrounding the so-called phenomenal intentionality research program (Kriegel forthcoming) is in this context: mental states represent subjectively in virtue of their phenomenal character. This is not the place to recapitulate work within this research program.²⁹ But note that insofar as the notion of subjective representation is motivated by consideration of environmentally insulated phenomenal duplicates (i.e., brains in vats), it is *prima facie* plausible that phenomenal character would underlie subjective representation. Within a nonreductive framework, this is not taken to be due to the fact that tracking relations underlie phenomenal character; rather, tracking and phenomenology are taken to be two different sources of two different kinds of representation.³⁰

8. Conclusion and Future Work

It has not been my goal, in this paper, to make a case for this sort of nonreductivism, casting objective and subjective representation as mutually irreducible. My goal has been more modest, and may be described as follows. Once we draw a *conceptual* distinction between objective and subjective representation, and rule out eliminativism about subjective representation, a certain structure emerges for a general theory of mental representation – a theory that accounts for *all* phenomena of mental representation. Such a general theory would comprise three chapters: (a) a theory of objective representation, (b) a theory of subjective representation, and (c) an account of the relationship between them. My goal in this paper has been to argue that familiar theories of mental representation in terms of tracking relations are inadequate as *general* theories. For they offer us is only (a), and are silent on (b) and (c). They thus fail to account for all the phenomena of mental representation. To do so, they would have to add either (i) a

reductive account of subjective representation in terms of objective representation or (ii) an independent account of subjective representation.

To repeat, I am open to the possibility that a reductive account of subjective representation in terms of objective representation might turn out to be right. Such an account would effectively constitute an approach to (c), and would pave the way for an approach to (b). Still, none of the familiar tracking theories of mental representation discussed in §3 actually offers such a reductive account of subjective representation (nor, of course, sketches the resulting tracking account of subjective representation). This type of work remains an outstanding intellectual debt on the part of tracking theories.³¹

References

- Block, N.J. 1990. 'Inverted Earth.' *Philosophical Perspective* 4: 52-79.
- Chalmers, D.J. 1996. *The Conscious Mind*. New York: Oxford UP.
- Chisholm, R. 1957. *Perceiving: A Philosophical Study*. Ithaca: Cornell UP.
- Dennett, D.C. 1969. *Content and Consciousness*. London: Routledge.
- Dretske, F.I. 1971. 'Conclusive Reasons.' *Australasian Journal of Philosophy* 49: 1-22.
- Dretske, F.I. 1981. *Knowledge and the Flow of Information*. Oxford: Clarendon.
- Dretske, F.I. 1988. *Explaining Behavior*. Cambridge MA: MIT Press.
- Dretske, F.I. 1995. *Naturalizing the Mind*. Cambridge MA: MIT Press.
- Fodor, J.A. 1984. 'Semantics, Wisconsin Style.' *Synthese* 59: 231-250.
- Fodor, J.A. 1990. *A Theory of Content and Other Essays*. Cambridge MA: MIT Press.
- Horgan, T. and J. Tienson 2002. 'The Intentionality of Phenomenology and the Phenomenology of Intentionality.' In D. J. Chalmers (ed.), *Philosophy of Mind: Classical and Contemporary Readings*. Oxford and New York: Oxford University Press.
- Horwich, P. 1998. *Truth*. Oxford: Oxford University Press.
- Kriegel, U. 2003. 'Is Intentionality Dependent upon Consciousness?' *Philosophical Studies* 116: 271-307.
- Kriegel, U. 2007. 'Intentional Inexistence and Phenomenal Intentionality.' *Philosophical Perspectives* 21: 307-340.
- Kriegel, U. 2010. 'Intentionality and Normativity.' *Philosophical Issues* 20: 185-208.
- Kriegel, U. 2011. *The Sources of Intentionality*. Oxford and New York: Oxford University Press.
- Kriegel, U. Forthcoming. 'The Phenomenal Intentionality Research Program.' In U. Kriegel (ed.), *Phenomenal Intentionality: New Essays*. Oxford and New York: Oxford University Press.

- Loar, B. 1987. 'Subjective Intentionality.' *Philosophical Topics* 15: 89-124.
- Loar, B. 1995. 'Reference from the First-Person Perspective.' *Philosophical Issues* 6: 53-72.
- Loar, B. 2003. 'Phenomenal Intentionality as the Basis for Mental Content.' In M. Hahn and B. Ramberg (eds.), *Reflections and Replies: Essays on the Philosophy of Tyler Burge*. Cambridge MA: MIT Press.
- McGinn, C. 1988. 'Consciousness and Content.' In *Proceedings of the British Academy* 76: 219-239.
- Millikan, R.G. 1984. *Language, Thought, and Other Biological Categories*. Cambridge MA: MIT Press.
- Millikan, R.G. 1993. *White Queen Psychology and Other Essays for Alice*. Cambridge MA: MIT Press.
- Putnam, H. 1975. 'The Meaning of "Meaning".' In his *Mind, Language, and Reality*. Cambridge: Cambridge UP.
- Putnam, H. 1981. *Reason, Truth, and History*. Cambridge: Cambridge UP.
- Searle, J.R. 1991. 'Consciousness, Unconsciousness, and Intentionality.' *Philosophical Issues* 1: 45-66.
- Searle, J.R. 1992. *The Rediscovery of Mind*. Cambridge MA: MIT Press.
- Siewert, C.P. 1998. *The Significance of Consciousness*. Princeton NJ: Princeton University Press.
- Stampe D. 1977. 'Towards a Causal Theory of Linguistic Representation.' *Midwest Studies in Philosophy* 2: 42-63.
- Strawson, G. 2008. 'Real Intentionality 3: Why Intentionality Entails Consciousness.' In his *Real Materialism and Other Essays*. Oxford: OUP.

¹ We could, of course, use 'representation-as-tracking' and 'representation-as-presentation.' But while these terms applied intuitively in the brain-in-vat scenario, words undergo subtle changes in connotative profile across contexts, and so these terms may not apply intuitively in other scenarios, even though there are no important substantial differences. (This is especially true of the 'presentation.') In addition, the expressions 'representation-as-tracking' and 'representation-as-presentation' are clumsy.

² What makes similarity relevant or sufficient is of course a complicated matter. If we had a full account of this, we could offer something like necessary and sufficient conditions for objective and subjective representation. Objective representation: 'x is an objective representation iff x is similar to the envatted brain's state of tracking S1 in respect R and to degree D.' Subjective representation: 'x is a subjective representation iff x is similar to the envatted brain's state of presenting an apple to the brain in respect R* and to degree D*.' For an initial grasp of the two notions, however, we can rely on intuitive takes on relevance and sufficiency.

³ Dennett (1969) distinguishes between personal and sub-personal mental states: the former are states of us, the latter are not. This distinction applies to representational states: sometimes we do the representing, sometimes sub-systems within us do it. The distinction between wide and narrow representations is due to

Putnam (1975): the former are not shared by intrinsic duplicates, the latter are. And Chalmers (1996 Ch.1) distinguishes between psychological and phenomenal conceptions of mental phenomena: the former characterizes mental phenomena in terms of their causal relations to each other and to the environment (their ‘long-armed functional role’), the latter characterizes them in terms of their experiential feels (their ‘phenomenal character’). This *conceptual* distinction, too, applies to mental representations as well (see Kriegel 2010).

⁴ Here too, it may well be possible to justifiably believe, and know, that you are not the space soul, but again this would be non-demonstrative knowledge, and being non-demonstrative, it would not *rule out* possible alternatives.

⁵ It is worth stressing that it matters not whether the situation the thought experiment enjoins us to envisage is genuinely possible or merely conceivable. For our distinction is in the first instance between two *notions*, not two *properties*. I will address the issue of whether there is a different property corresponding to each notion later on.

⁶ It was pointed out to me, in this connection, that the Oxford English Dictionary’s entry on representation is parceled out into seven different meanings, each sub-divided into distinct usages, in a way that suggests a heterogeneous domain of phenomena. Although the entry’s authors’ taxonomy and organization of the domain leaves much to be desired, the multitude and variation of meaning seems to be a genuine phenomenon.

⁷ Likewise, it is not as though literary and mathematical representations represent due to the same representation relation, with the former qualifying as literary simply because they employ literary vehicles of representation and the latter as mathematical because they employ mathematical vehicles. We can imagine a single item functioning as a diagram in one context and a metaphor in another.

⁸ It is important to note, however, that there is also an intimate relation between the two notions, inasmuch as the following non-trivial connection holds: whenever I have a conscious experience which represents veridically in both the objective sense and the subjective sense, what it tracks and what it presents to me are one and the same. This congruence can hardly be an accident. There is thus a non-accidental tie between what an experience represents in the objective and subjective senses when everything goes the way it should. So while we need separate theories for mental representation in each sense, there needs to be sufficient contact or overlap between them to explain this non-accidental tie.

⁹ This last condition requires that the non-Fs cause Ms because Fs cause Ms whereas the Fs cause Ms not because non-Fs cause Ms.

¹⁰ Arguably, conditions (i) and (iii) are not satisfied either, but we do not have to worry about that, as the dissatisfaction of (ii) is sufficient to generate the problem.

¹¹ Moreover, in most cases (though not quite all), the nomic dependence conditions means that Fs are the only lawful causes of Ms. It is clear that this more specific condition will not be satisfied by the envatted brain’s apple experiences, since they have no apples as causes, lawful or not.

¹² The term ‘correspondence’ is Millikan’s (1984, 1993), and is left unexplained. One would be warranted to suspect that the relevant correspondence relation is ultimately to be accounted for in terms of the sort of tracking relation appealed to by causal-convariational theories. If so, teleosemantics is in fact just a teleological augmentation of causal-convariational theories.

¹³ This line of argumentation against teleosemantics parallels closely Strawson’s (2008), and is in fact adapted from it. Strawson argues that teleosemantics cannot accommodate the possibility of *pure observers*, because such observers are incapable of adaptive behavior. As example of pure observers, he cites the ‘weather watchers’: intelligent creatures stuck to the ground and unable to move but sensitive to and intensely interested in the weather. These creatures do not behave at all, let alone adaptively, but

clearly have mental representations, argues Strawson. We can stipulate that something very similar is true of the envatted brain: it does not genuinely interact with its environment, though it seems to itself to do so.

¹⁴ Recall that in our version of the tale the machine stimulates the brain randomly, it is not controlled by an intelligent and purposeful ‘evil scientist.’

¹⁵ The term ‘interpreter’ is used somewhat technically here, to designate any agent engaged in ascription of representational states.

¹⁶ We can think of it this way: if objective and subjective representation were one and the same property, given that S1 is not an apple one of the two ascriptions would have to be false. Conversely, since both ascriptions are true, objective and subjective representation must be two different properties.

¹⁷ What I mean by ‘true in the minimalist sense’ is something like: true in the sense of satisfying the platitudes used to formulate the so-called truth schema in minimalist or deflationary theories of truth, such as Horwich’s (1998).

¹⁸ We may also define a looser version of reductivism, allowing for reduction of subjective representation to objective representation plus other familiar and recognizably ‘kosher’ materials, such as functional role or other causal/mechanistic notions. I will ignore this possibility in what follows, because everything I say about the more rigorous type of reductivism should apply *mutatis mutandis* to this looser variety as well.

¹⁹ Thus, there are ways of construing action so that the envatted brain acts in a way that calls for positing an experience that represents an apple. Suppose we can read off the monitor that controls the vat that our envatted brain initiates motion of ‘its’ apparent arm in the direction of the apparent apple, performs an apparent grasping motion, and brings the apple of ‘its’ apparent mouth. This pattern of information we see on the monitor invites an explanation that includes the claim that the brain has a representation as of an apple.

²⁰ In addition, even if conceivability does not under any circumstances entail metaphysical possibility, surely it provides defeasible evidence of metaphysical possibility. If so, in the absence of defeaters we would be epistemically obliged to adopt the hypothesis that the brain-in-vat scenario is metaphysically possible.

²¹ In saying that the experiences are ‘presentationally different,’ I mean to suggest that – at least for an envatted brain unaware of the identity of Phosphorus and Hesperus – what is presented in the case of one experience feels different from what is presented in the case of the other. I say more about this in Kriegel 2011 Ch.3.

²² Interestingly, it has sometimes been argued that *objective* representation does not exhibit those features. Searle (1991, 1992 Ch.7) argues that any mental state that tracked the morning star would *eo ipso* be tracking the evening star, and I have argued (Kriegel 2007, 2011 Ch.3) that if a mental state tracks an apple then there is an apple.

²³ Loar (1987) uses the term ‘subjective intentionality’ to pick out something which ends up being more or less the same as our subjective intentionality. But my own way of introducing the notion is much more theoretically neutral: as a label for one sense of representation we find in contemplating brain-in-vat scenarios.

²⁴ I develop an account of this sort in Kriegel 2011 Ch.2.

²⁵ Or at least, familiar stories about tracker acquisition appear to require this – an alternative story would have to be devised if the idea is to get around such a causal-interactive requirement.

²⁶ A claim in the general vicinity is made by Strawson (2008).

²⁷ Indeed, one might reasonably suggest that even the combination of ‘M tracks F’ and ‘F coextends with G’ entails ‘M tracks G.’

²⁸ It is worth noting that avoiding reduction of subjective representation to objective representation, while it institutes a kind of dualism about representation, does not constitute the sort of ontological dualism that challenges physicalism. It is perfectly consistent with subjective representation being irreducible to objective representation that it is reducible to, say, 32 Hz oscillations in the hypothalamus.

²⁹ For relevant work, see *inter alia* Loar (1987, 2003), McGinn (1988), Searle (1991, 1992), Siewert (1998), Horgan and Tienson (2002), Strawson (2008), and Kriegel (2003, 2007, 2011).

³⁰ The view that tracking relations underlie phenomenal character, in that a mental state has its phenomenal character in virtue of bearing the right tracking relations to the environment, has been defended by a number of proponents of familiar theories of mental representation that also endorse a representational account of phenomenal character (e.g., Dretske 1995).

³¹ For very useful comments on a previous draft, I would like to thank Stephen Biggs, Davide Bordini, David Chalmers, Allan Hazlett, and Farid Masrour. I have also benefited from presenting the paper at the University of Wisconsin – Madison. I am grateful to the audience there, in particular Matthew Kopec, Larry Shapiro, Allan Sidelle, and Peter Vranas.