The Sufficiency of Objective Representation

Robert D. Rupert

I. Introduction and methodology

Over the past half century, prevailing views about mental representation have undergone a series of drastic changes. Wittgensteinians and behaviorist psychologists made denial respectable, deriding the idea of mental representations as confusion borne of a category mistake or as unverifiable nonsense. The cognitivist revolution ushered in a realism about mental representations, eventually giving rise to dogged and ballyhooed attempts to "naturalize" the semantics of mental representations (by explicating the representation-determining relation between psychologically – and physically – real internal entities and the properties, kinds, or individuals in the environment represented by those entities) (see Dretske 1981, 1988, Fodor 1987, 1990, Millikan 1984, among many others). Over the past two decades, discussion of a subjective, fully internal form of representation has blossomed, driven by the assumption that we should take at face value direct introspective awareness of something that seems like meaning, content, or representation.

This whirlwind history runs roughshod over many distinctions, one of which is particularly relevant to position laid out below. Mid-century philosophers tended to dismiss talk of psychologically real mental representations on conceptual grounds: mental representations have meaning, meaning is partly constituted by normative constraints, and normative constraints are public; so, assuming that mental representations are internal by definition, there are no mental representations (Wittgenstein 1953). In contrast, behaviorist psychologists avoided talk of mental representation on methodological grounds; by their lights, there was no empirically legitimate way to study internal mental entities and, much more importantly, they thought they had a way of accounting for the data without invoking mental representations (Skinner 1957). A simple application of Ockham's razor cut mental representations out of the behaviorist tool-kit. The cognitivist revolution rose on similarly contrasting motivations. Some philosophers abandoned behaviorism because it did not countenance the first-person perspective (Putnam's super-Spartans). Others, however, put no special emphasis on the first-person perspective; they argued that one could not do justice to the empirical data – about language acquisition, for example – without positing internal representational units (Chomsky 1959, Fodor 1975).

Recent developments have a different flavor, however. The proliferation of books and articles about consciousness – about Mary the superscientist (Jackson 1982), the explanatory gap (Levine 1983), phenomenal consciousness (Block 1995), and the Zombie-motivated hard problem (Chalmers 1996) – aroused dissatisfaction with the cognitivist compromise; functionally defined mental representations, and their relations to external entities they tracked, left cold those who impressed by the apparently rich experiential contents of their own inner, mental lives. This most recent transition – to consciousness-based discussions of mental representation – lacks naturalistic motivation,[1] though, and, in my opinion, this is telling.

[1] Naturalism is not equivalent to empiricism. The former holds that the theoretical, analytical, and experimental methods of employed by our most successful sciences provide the best method for finding the truth or acquiring knowledge (or justified beliefs), but there is no commitment – quite the contrary – to the view that all concepts are constructs from observations, impressions, or sensory experiences or that all justification

The behaviorist rejection of introspectionism in psychology was meant to express a scientific urge, as was the introduction of internal mental representations by nativist linguists and memory scientists (e.g., Miller 1956). The cognitivists embraced Ockham's razor no less than the behaviorists; rather, they disagreed with behaviorists regarding what theoretical posits were necessary to account for the empirical data. This seems to me to be the right strategy, one that I pursue in the remainder. Ockham's razor – together with the long history of successful naturalistic theorizing – recommends that we make a serious attempt to account for the data that drive the subjective turn without positing any new form of mental representation. This data (which I refer to as the 'relevant data' or the 'data in question' in what follows) consists primarily of the ways that philosophers express their conviction that there is a distinctive category of representation or meaning: subjective or internalist content. I shall insist, however, that accounting for these reports is decidedly not the burden of an objective notion of mental content alone. In order to account, for example, for the judgments about possibility philosophers issue in the face of thought experiments, one must advert to theories of cognitive processing, broadly speaking; but this is no strike against the sufficiency of objective content, for, historically, theories of processing have been part of the theoretical package that includes mental representations with objective mental content. Appeals to objective content alone do not account for the relevant data, but they were never meant to. Objective theories of representational content attribute content to units that play a role in cognitive or mental processes; and the characteristics of these processes explain much of the relevant data:

rests solely on empirical observation; in other words, any sensible naturalist should reject empiricism.

the reports of introspective access to states with a special sort of subjective content or reports of intuitions (or judgments) – about thought experiments, for instance – that would seem to support a notion of internalist, subjective content. So, theories of objective content, as theories of content for mental representations, needn't be supplemented at all; they need only be placed in a package with the kinds of theoretical elements that normally accompany them in psychological modeling; that package contains only objective content, and thereby contains all the content needed.

II. Mental representation and objective content

As noted, the cognitivist revolution ushered in a new era of realism about mental representations, and it did so in conjunction with an emerging computer science. As a result, mental representations were frequently referred to using the language of 'data structures', 'symbol strings', 'information-bearing states', and 'machine tables'. Many of those who had functionalist leanings in philosophy of mind were inspired by computational cognitive science (Fodor 1975), and, as a result, this language appeared in philosophical as well as scientific discourse. What would render such a structure a representation, though? What makes it specifically representational? How, for example, should we understand the idea that it has content?

Think of this partly as a methodological puzzle (Stich 1983, Fodor 1987). Scientific procedure seems to speak in favor of a so-called narrow methodology, one that focuses solely on the causal processes that eventuate in intelligent behavior. Given a generally localist assumption about causation, one should expect the content of mental representations to be determined fully by internal processes, at least if such content is to play a causal role. The proximal cause of behavior had better be inside the organism doing the behaving!

An inferential- or conceptual-role semantics offers one objective notion of representational content for internal states, objective in the sense that the content of a given mental representation is determined entirely by causal and structural relations that can be specified fully, and can, in principle, be measured determinately, from the third-person, scientific perspective. On this view, the content of a mental representation is constituted entirely by some subset of the causal interactions it enters into (Block 1986). So far as I can tell, the many shortcomings of such a view (Fodor 1998) outweigh whatever benefits might accrue to the placement of content in a location that makes it a candidate causal contributor to the production of behavior.

Two such shortcomings strike me as particularly problematical. First, if part of what constitutes a given conceptual- or inferential-role content is that the vehicle bearing such content participates in certain, privileged inferences, then that vehicle's having that content cannot explain, causally, why those inferences occurs, on pain of circularity; one should not hold that the unit has its given content because the unit causes certain transitions *and* that it causes those transitions because it has the content in question (Fodor 1998, chapter 1). Second, the inferential-role view seems to rob mental representations of the sort of intentionality we take them to have. If their content supervenes only on the internal structure, then content isn't a matter of being related to the objects represented, the *actual things* that we think about – Sandy Koufax, zebras, charge, and so on – at least on the assumption that the internal states do not determine what's in the environment.

Moreover, many of the arguments taken to speak in favor of a competing externalist semantics for natural-language terms – Kripke's arguments (1980) from error and ignorance, for example – seem naturally to apply to mental representations, particularly if one adopts the view that linguistic units have the content they do partly because they express the content of the mental representations that produce them. Consider, too, certain realist intuitions in philosophy of science that seem best accommodated by a framework that includes external content: we tend to think that different scientists holding very different theories of, for instance, electricity are thinking about the same phenomenon – the one the nature of which they disagree about – and this thought might seem even more compelling as regards one scientist whose thinking evolves from one stage in her career to the next.

More scientifically oriented considerations seem to reinforce the need for externally oriented representational content. Psychology discovers laws stated in terms of content (Pylyshyn 1984, Fodor 1998, Burge 1986), but inferential- or conceptual-role content varies radically from subject to subject. It would seem that only external content can consistently provide a common aspect to various subjects' water-thoughts, for example; regardless of what idiosyncratic beliefs various subjects might have about water, they can all be about the same stuff in the environment: H2O. Moreover, regardless of what one thinks about intentional laws, cognitive science seems rife with some with explanations that presuppose externalist content; it is presupposed that stimuli activate internal units that control behavior distinctively oriented toward the kind of stimuli that those internal units track (cf. Ramsey 2007, who worries that tracking is garden-variety causal mediation). And, returning to a meta-scientific perspective, we might wonder how we could possibly make sense of the scientific endeavor itself if scientists weren't thinking *about* the subjects in the lab, the lab equipment, their co-authors, editors at the journals to which their results are to be sent, the NSF director who facilitates review of their grant applications, etc.

Here, then, are this section's takeaway messages. First, an objective notion of the content of mental representations is adequate to the phenomenon; no additional kind of specially subjective or consciousness-related kind of content is needed. Second, the most promising version of such an objective, or third-person, view takes the form of a semantic externalism, not an inferential- or conceptual-role theory. But, in the event of over-reaching, let me fall back to a watered-down line: since there seem to be good reasons to posit an objective, externalist (or tracking) notion of cognitive content, we should ask whether that content suffices to account for data that might suggest the need for additional forms of content, either conceptual- or inferential-role or some form of subjective content.

III. Concepts, conceptions, and architecture

In this section, I sketch the elements of a model of human cognition. The picture presented draws primarily on traditional computational modeling practices, although it can be adapted more or less easily to accommodate other approaches in cognitive science (e.g., dynamicist [Port and van Gelder 1995] or connectionist [Rumelhart, McClelland, and the PDP Research Group 1986] approaches). Although it is only a sketch, I hope it provides the reader with sufficient background to see how, in the section to follow, I mean to deploy this package of resources in order to account for the relevant data.

A. *Concepts (or mental representations), atomic and otherwise.*

The bearers of mental content – the things filling at least one slot in the representation-relation – are often referred to as 'concepts'. In what follows, I use the more neutral term 'mental representation' so as to avoid theoretical disputes over the requirements that a mental representation must meet in order to qualify as a concept. As bearers of content, mental representations can fruitfully be thought of as vehicles. Such vehicles should be individuated independently of their content – that is, nonsemantically (Rupert 1998) – which jibes nicely with computational theories of processing (Fodor 1994, chapter 1), as well as with other forms of mechanistic models in the cognitive sciences.

Mental representations can be either atomic or compound. Atomic mental representations, conceived of nonsemantically, are the smallest units that affect cognitive processing. Given a stock of atomic mental representations, cognitive operations can compound such units into strings or organized collections of other sorts. (I presuppose that, in our mechanistic models of cognition, cognitive operations are sensitive only to nonsemantic properties; this does not preclude a story according to which semantics also plays a role, but it will not be in the nuts and bolts of processing.) Atomic representations are thus the minimal content-bearing units – minimal relative to processing[2] – but (typically) they possess content, and the content of a compound representation is a

[2] The characterization of a representation's being minimal with respect to content is a tricky matter. If content is purely externalist, then one might think the only semantically minimal representations are representations of fundamental particles, forces, or relations; in all other cases, the thing represented is physically (or metaphysically) compound and thus, as a semantic value, not atomic.

function of the content of its atomic components, where that function may take the form, for example, of a typed grammar.

This view of mental representation provides at least three kinds of nonsemantic material to the causal-explanatory enterprise: nonsemantically individuated atomic units, processing operations that compound and otherwise operate on (by, for example, writing, rewriting, decomposing, or transforming) strings of those atomic units, and rules that determine the content of a compound string as a function of the content of component atoms. Bear in mind that these materials appear in standard theories concerning the role of content-laden units in psychological explanation (Pylyshyn 1984, Fodor 1994). As such, to invoke these when accounting for the relevant data is neither to supplement theories of objective content, qua theories of content, nor is it to supplement the theoretical framework that serves as the standard home for theories of objective content.

B. *Conceptions.*

The notion of a conception builds on the idea of a compound mental representation. Atomic mental representations are the building blocks for individual compound strings, which might be thought of, in the first instances, as analogous to simple sentences (e.g., “Cows are mammals”). Individually, such strings represent the world a certain way (by, e.g., having satisfaction conditions). Often it is thought that certain groups of such compound representations play privileged cognitive roles. Take a specific atomic mental representation in a given subject. Typically, this appears as a component of numerous stored or standing strings. So, we might characterize a subject’s conception of *x* (or *x*’s) as the entire collection of stored strings such that each string in the collection contains at least one instance of X. Typically there are *very* many of these, and, thus, to stand a chance of being theoretically useful, conceptions are typically limited to some proper subset of the collection – what is thought to be the subject's core knowledge concerning the individual, kind, or property represented by the atomic representation in question. An atomic mental representation might be COW, and the conception of cows might be a set of mental structures such as {COWS ARE MAMMALS, COWS ARE ALIVE, HUMANS KEEP COWS ON DAIRY FARMS, COWS ARE BIG, COWS ARE ANIMALS}; this set might be larger, and contain much of what the subject represents about cows, but it does not consist in everything the subject believes about cows. Various forms of conceptions of have been proposed, among them file folders (Forbes 1989), knowledge structures (Cummins 1996), and frames (Minsky 1974). It is a matter of some dispute what should go into this set. Putting too little into it creates versions of the frame problem, but putting too much into it creates the problem that no two people share the same conception of a given kind, property, or individual; more generally, the issue gives rise to much hand-wringing about the analytic-synthetic distinction (Fodor and LePore 1992) (among those who are inclined to think of conceptions as word meanings, which I am not).

For present purposes, I need not give a full account of conceptions. In fact, I'm inclined to think they play no role, as distinctive theoretical constructs, in the causal-explanatory enterprise. I discuss them here partly to warn against the conflating of intuitions about conceptions (of some grain or another) with intuitions about the content of atomic mental representations. To be sure, there could be some sort of content that attaches distinctively to conceptions (an inferential-role semantics seems to offer an obvious possibility). We should bear clearly in mind, however, the possibility that content attaches, in the first instance, to atomic mental representations only, and that contributions of other factors – such as intuitions about conceptions – account for erroneous intuitions about mental representations. Perhaps more to the point, interactions among strings of mental representations may account for the relevant data, regardless of whether there is, for any *x*, a privileged theoretical construct – the subject's conception of *x*; thus, we should keep in mind this particular aspect of the standard package, as a tool to be exploited without the introduction of any kind of content beyond externalist or tracking content.

C. Architecture.

Models of cognition typically include an architecture. Cognitive architectures take many forms – classical, connectionist, dynamicist, subsumption, associationist – but, essentially, the architecture is the collection of basic elements and operations that constitute the cognitive system, together with any fixed structure or structure-related constraints on the execution of those operations; it is the collection of tools available to play a causal-explanatory role at the level of cognition. In the case of computational models, the cognitive architecture includes the stock of atomic mental representations and a description of their processing-related properties, the operations available (including such things as parameter settings relevant to the performing of those operations – say, decay rates in a short-term memory buffer – and rules for compounding those operations into more complex operations), and also the components that play a different role in the overall functioning of the cognitive system – what is distinctive of them and how they're connected to each other. For example, face recognition in humans might proceed by a series of operations that is relatively independent of the processing of the incoming speech stream, and it may be left to a third, downstream component to localize the source of the speech (thereby binding it to a face, if one is available). If so, these are architectural facts – about which components of the cognitive system transfer information to which others, to what extent they do, what limitations there are on such communication, and what forms of behavior they control as a result.

The preceding provides a sketch of the tools available for the construction of specific models of human cognition processing. Modeling that employs such tools has been productive (see, for example, various incarnations and applications of the ACT-R and SOAR architectures), although there is widespread debate among cognitive scientists as to whether this kind of modeling is on the right track or whether alternatives should be pursued more intensely (cf. Chemero 2009).

Consider, now, the dual role that modeling might play in the current context. On the one hand, modeling permeates the sciences. So, in describing the tools for modeling human cognition, I am providing no more or less than one would provide in connection with any other science. But, this is a model of human thought, and thus should, in principle, model the very cognitive processes involved in the formulation and use of models, including the formulation and use of cognitive models. This requirement might seem most pressing if one has a certain general view about human cognition. I contend that modeling manifests our fundamental cognitive urge: we model everything from the motions of objects in the heavenly bodies to the minds and behaviors of our conspecifics. We are, cognitively speaking, modelers in the first instance (and I am inclined to think that language-use is itself an act of model-application, which accounts for much of the context-specificity of language use). Everyday thought models everyday data and the systems giving rise to it; scientific thought models more carefully data that is systematically collected or experimentally produced. But, all human understanding is essentially an exercise in modeling, and this includes the understanding of how we formulate and use models in cognitive science or philosophy.

The preceding sketch of the tools available for the modeling of cognition also provides the materials for self-reflective modeling, for modeling the cognitive act of modeling. We are modelers, and thus, when we turn to understanding ourselves, we construct models of human cognition itself, models of how we model the world. Such modeling is vindicated by the results. (Presumably, the world is the sort of thing subject to modeling; the success of our various modeling enterprises itself is best explained by the assumption that the world is the sort of things with recurring elements and standing relations among them and is thus amenable to modeling.) In what follows, I will try explain away the relevant data concerning mental representations by applying the roughly computational model sketched above to show how we naturally model our own experiences and, to some extent our own thinking, erroneously; I will attempt to model how we naturally construct a model that includes a property of intrinsic, subjective internalist representation, in the absence of anything having that property. Furthermore, the model I suggest of the process of constructing an erroneous model includes only objective, tracking representation, at least so far as representation or content goes. In effect, then, I argue that there is only objective, tracking mental representation by invoking a model that includes only this one form of representation (together with other elements of the standard package) to explain why we produce the data that would seem to support the existence of subjective representation, by explaining how humans construct models of their own psychological processing that contain representations with empty reference, representations that nevertheless help to produce reports that include such terms as 'subjective representation'.

IV. Explaining the cases

In Uriah Kriegel's complementary piece (this volume), he lays out a range of kinds of intuition (or conceptual judgment) that seem to entail the existence of a distinctive form of subjective content. He expresses the first of these as follows: "There are conceptually possible scenarios where representation varies in the objective sense but remains invariant in the subjective sense" (ms, p. 3). As an example of such a scenario, Kriegel describes a color-inverted world, that is, a world in which human subjects have the same color experiences, but in which the colors in the world have been systematically swapped (for instance, subjects employ the actual-Earth internal color experience of red to track what are now blue things – e.g., red delicious apples – in the environment).

In my view, we can, and should, account for the judgment in question without invoking a kind of subjective representation that remains constant across subjects in the color-inverted and actual worlds. Insofar as the judgment itself involves a concept of such constant representation, the task in what follows is to model this erroneous application of the concept of representation. But, first, two preliminaries: Notice an element missing from the package described in section III: a self, beyond the architecture (see Rupert 2009, for further discussion). Among the enduring commitments of philosophy of mind is that there exists an entity, a person, to whom subjective, personal-level content is presented (McDowell 1994). I find such a view unmotivated, however. It is true that subjects learn to use such pronouns as ‘I’ and a valence of conviction colors many uses of them (in such sentences as “*I* am the one who sees; my visual cortex doesn’t see!”).[3] Moreover, it might be, for example, that certain forms of executive control work more efficiently if there is a set of compound mental representations (recall the discussion of conceptions) that is specially rigged to motor output and such that all of the compound strings in the set share a common atomic representational element that we would naturally describe as a way to refer to oneself. But none of this entails the existence of an entity or distinctive construct, the self, intuitions about which ground claims about subjective representation. Of course, the fact that some of us make the judgments in question (that, for example, the *person* sees, not the cortex) must be accounted for somehow. If the standard package can do so, however, without presupposing a distinctive

[3] Dualist philosophers sometimes reject the idea of an entity to which subjective content is presented and instead take the relevant relation to be something more like constitution; on the latter view, subjective content partly constitutes the states of the subject. This may be a promising path to pursue – whether within a dualist framework or not – but such pursuit should comprise the development of an adequate theory of processing, that is, an account of how something that constitutes one part of the self interacts with other things so as to give the erroneous intuition that the constitutive part is being presented to the whole; so far as I understand what it is for x to be presented to y, its holding entails that x is wholly distinct from y. I suspect that such a story, once told, will make reference to elements and relations structurally similar to those of the objective account. In doing so, it may lend itself to an eliminativist account (in the terms used by Kriegel) of subjective content, of the sort to be developed below.

person who, for instance, makes reliable judgments about what contents are presented to it, judgments that might be used to argue in favor of the existence of a distinctive form of subjective content, this may support an eliminativism about the self. I will not pursue this project in any detail, but the following discussion the causal efficacy of vehicles and of the illusion of internal content should provide the reader with a further sense of how I think one best accounts for intuitions about a distinctively personal-level.

Second, I propose to muddy the distinction made earlier between inferential- or conceptual-role theories of content, on the one hand, and externalist theories of content, on the other. From my perspective, what is essential to externalist theories are their tracking nature – the fact that some kind of causal, covariational, or informational relation holds between the representing vehicle and the individual, property, or kind represented. When conceived of as purely a matter of tracking, though, the issue of location becomes irrelevant; the thing being tracked can just as well be internal as it can be external to the human organism in which the tracking vehicle appears. There's nothing unusual about this idea from the standpoint of cognitive modeling; it's common enough to include "pointers" – that is, units that function to represent other units (e.g., the units stored at an address to which the pointer points [Newell and Simon 1997/1976]) – in computational models. Additionally, the complex pattern of neural connections one finds in the humans suggests an abundance of within-brain tracking relations (Goldman-Rakic 1987). To further muddy the waters, I hold that externalist contents in the cognitive system frequently piggy-back, exhibiting the kinds of relation found in cases of linguistic deference. For instance, one internal vehicle might borrow organismically external content from another internal vehicle by externalistically representing that second vehicle.

(In general, philosophers have, I think, tended to ignore such possibilities because of their privileging of a personal-level at which genuine content can be content only of a state of a complete subject, which is identified (roughly) with the organism. On such a view, the idea that there could be a vehicle with internal-externalist tracking content – which content is not self-revealing at the locus of that vehicle – doesn't make sense; any content is content of the entire organismically oriented subject, so any tracking relation between vehicles both of which are internal to the subject will determine a kind of content that must be accessible, or self-revealing, to the subject as a whole. On my view, all of this talk of entire subjects is misleading or at least puts the cart before the horse. Let's first model the data – intelligent behavior and the like – then see what sort of self that modeling yields, and whether it makes sense within that framework to talk about internal representational vehicles the content of which might be no more than another internal, tracked vehicle.)

Preliminaries out of the way, we can ask why philosophers would have the intuition that color-inverted earth is possible, if all representational content is externalist. The judgment in question involves the application of the concept of representation or of intentionality. Whether the judgment is correct depends on whether the property represented – that is, the property (or relation) of representing or being about – could be instantiated in a world that satisfies the description of color-inverted earth (or whether, say, given the nature of the intentional relation, there simply can't be sameness in internal intentionality when there is difference in external intentionality, and so no world satisfies the description in question).

How do we acquire the concept of intentionality? Elsewhere (Rupert 2008), I propose that the acquisition of REPRESENTS SOMETHING (as a one-placed representational vehicle) proceeds by the application of that vehicle to other internal vehicles, such as COW, DOG, MAMMA, HOUSE in contrast to its nonapplication to such vehicles as UNICORN, BOOGIE MAN, and SNIPE.[4] Grouping alone – using a method of samples and foils (Stanford and Kitcher 2000) – homes in on a tracking relation, although I suspect it does not do so without some feedback provided by interactions with the environment that help to guide the classification of different internal vehicles into samples and foils; some terms initially treated as samples may come to be treated as foils when the child's executive systems fail to discover the robust causal connections to, say, sensory experiences that executive systems detect in standard samples.

This thought brings two essential elements to the fore: vehicles and their interconnections. When we think about our own thoughts, we activate vehicles that track other vehicles. We don't know this *a priori* because the vehicles tracking other internal vehicles do so via a causal relation; thus, what's on one side of the relation, the tracking vehicle, may quite successfully track what's on the other side of the relation, without the tracking vehicle's controlling accurate reports on the various properties of the thing

[4] The use of, for example, 'UNICORN' refers to a certain vehicle individuated nonsemantically – say, in terms of its computational role or some of its neural characteristics. Which vehicle? It is easiest to designate it as the one that systematically controls utterances of 'unicorn'.

tracked – that is, the represented vehicle.[5] It is possible, then, that when we have the intuition that subjects in color-inverted earth share subjective representations with earthly subjects, the things actually shared are vehicles. I say to myself, "I could be in *that* same state, even if colors were inverted," which is true, but what I may not be able to report on or reason about very accurately is the nature of "*that* state"; I claim that what is demonstrated by the vehicle controlling judgments and reports is a vehicle, not a content.

[5] The treatment of the attempt at *a priori* knowledge in this case does not place it on par with other attempts at *a priori* knowledge. We may be able to achieve more reliable *a priori* mathematical knowledge by applying structural operations to vehicles that represent number properties (in my view, via a causal semantics: TWO tracks *two-ness* in the environment, so far as I can tell). It is one thing to track structural relations and perform structurally sensitive operations that preserve truth; this ability may be built into the architecture and may facilitate the acquisition of mathematical knowledge. It is another thing to think that when one vehicle tracks the activation of another vehicle, the former thereby can produce accurate reports about the various properties of the tracked vehicle; thus, there's plenty of room for, and reason for, skepticism in the case of supposed *a priori* reasoning about the workings of our own minds that does not automatically bleed over to other domains in which we think we have *a priori* knowledge. Thanks to David Chalmers for pressing me on this issue.

(Thus, this yields a very thin notion of sameness of subjective representation: to have the same vehicles active across contexts).[6]

What role is played by the interconnect-ness of vehicles? As suggested above, such interconnections play a role in the acquisition of the notion of intentionality or representation, even if these represent a pure tracking relation; patterns of interconnections (for instance, DOG's being activated in a variety of contexts in which sensory representations – such as FURRY FELLING ON MY BODILY SURFACE – are also active) help to determine which vehicles activate the further vehicle REPRESENTS SOMETHING. Such patterns of activation help the subject to home in on vehicles that represent; moreover, because of their tight connection to the application of REPRESENTS, these associations create an illusion of inferential-role content: INTENTIONAL and REPRESENTS are the ur-semantic mental representations (their activation tracked by SEMANTIC) and as a consequence things closely associated with their application – such as causally interconnected sets of vehicles – get treated semantic as well, even when they are not of the same natural kind or do not instantiate the same natural properties as those represented by the other terms to which SEMANTIC applies. So, subjects treat these interconnections as somehow content-constitutive, even though they are mere causal contributors to content-determination (Rupert 2008). When we consider color-inverted earth, then, we are inclined to think that such networks of interconnected vehicles remain in place (although we couldn't produce this description

[6] This thought can be extended to the case of phenomenal experience: my thinking about what it's like for me to see red is just to think about the sensory vehicle that plays the red-detection role in my actual life!

on simple reflection), and this contributes to the judgment that subjects on earth and on color-inverted earth share representations; and this generates the illusion that there is some kind of content that is non-tracking.

Notice that this deflating explanation is not built from materials assembled *ad hoc*. Two of the most influential tracking theories (Fodor 1987, Dretske 1988) propose, for independent reasons, that the content-fixing, tracking relation can be causally mediated by other representations.

What about "conceptually possible scenarios where representation varies in the subjective but not objective sense" (Kriegel, ms), an illustration of which is the traditional inverted-spectrum case? A straightforward explaining away of this intuition – that an inverted internal spectrum is possible – runs as follows: one can imagine that a very different network of internally tracked and internally tracking vehicles gets attached to the environment in just the way that one's current network is – at least, this is how one should articulate what one is imagining if the conceptual possibility is a genuine metaphysical possibility. Such appeals to difference in vehicle across sameness in external content help to explain other phenomena as well (including substitution failures – see Fodor 1990, chapter 4).

The third case is "representation in the objective sense in the absence of representation in the subjective sense." Kriegel offers the example of tree rings, and there are many others that have been discussed in the literature, from thermostats and fuel gauges to magnetosomes used by certain bacteria. Take the example of tree rings. Depending on one's theory of content, a number of tree rings may not qualify as a representation; it is one thing to label a theory 'tracking' to get across a core element of it, but theories of the tracking variety generally involve a complex set of necessary and sufficient conditions; no serious theory in the field holds that x represents y if x naturally means (in Grice's sense) y or x was simply caused by y. For the sake of argument, though, let us pursue the matter further, as if tree rings do represent the age of the tree, in keeping with our best tracking theory. Here it's important to distinguish between cognition and representation. Representations are part of the cognitive scientist's tool kit, but no one in the field thinks that the activation of representations alone accounts for intelligent behavior (the *explananda* that the standard package was assembled to explain). Trees don't use language, plan, remember, build buildings, construct scientific theories, etc., and they don't partly because they have only (objective) representations and none of the other components that contribute to the explanation of intelligent behavior. Recall, too, the contribution of interacting components to the creation of the illusion of subjective content. In this case, we judge that something is missing, relative to the human case, but it is a mistake to take that something to be a form of content (subjective content); it's everything else (architecture, interaction between representations, etc.) that's missing.

The fourth case involves subjective representation in the absence of objective representation. The clearest case would be that of a conscious being in a universe containing nothing else (Kriegel, ms.). Again, I think an account of the possibility-intuition falls out of my framework. In this world, the subject has within her all of the standard elements – including cognitive vehicles and their causal interrelations.

V. Epilogue

Have I eliminated subjective content, or rather provided a (perhaps boring) reduction of it? Readers might suspect that it's the latter, for why not take SUBJECTIVE CONTENT itself to represent – that is, to internal-externalistically track – a natural kind or property, the property had by collections of appropriately interrelated strings of mental representations (roughly, those related by inferential roles in the way the elements of a conception are supposed to be – perhaps with a special emphasis on diagnostic roles of certain connections relative to the determination that a given vehicle actually represents something)?

Although this seems like a reasonable reading of the situation, matters are not so straightforward. In section II, I reviewed various reasons to be skeptical about the value of inferential- or conceptual-role content. If my concerns about such content are well-founded, it plays no role in the causal-explanatory enterprise. By some lights then, although interactions between various vehicles might be genuine aspects of reality – ones that supervene on natural processes that are part of the causal order or are covered by natural laws – they nevertheless fail to provide an appropriate target for tracking (cf. Kriegel 2011, 96). At least on one conception of the naturalization of content, content-determining relations should hold between natural kinds or properties (Rupert 1999); that would seem to be what it amounts to for intentionality or content to “really be something else” (Fodor 1990), where that something else is part of the natural, causal order. If conceptions aren’t natural kinds or properties, how can they enter into the causal relations that they must in order to be tracked, and thus represented?

Perhaps, though, I’m mining an excessively narrow-minded vein here, with regard to the relata of the tracking relation. Perhaps, SUBJECTIVE CONTENT does genuinely track something along the lines of conceptions. In that case, I have offered a reduction of sorts, but one that vindicates only a very thin notion of subjective content, relative to how subjective content is often understood. This reduction provides no support, for example, for the view that the perceptual states of which we're immediately aware have intrinsic qualitative character or that the mind has immediate awareness of a rich sort of content that makes its theoretically interesting properties available directly to the cognitive processes that generate responses to thought experiments or produce verbal reports of philosophical intuitions. Moreover, on this view, the reduced notion of subjective content is a structured collection of interrelated vehicle strings, not something that attaches to an individual mental representation, atomic or compound (of the form of a simple sentence). It may be something real – there to be picked out by tracking vehicles – but may play no role in cognition or the production of behavior, beyond their of the activation of the vehicles doing the tracking.

Works cited

Block, N. (1986). Advertisement for a semantics for psychology. In P. French, T. Uehling, and H. Wettstein (eds.), *Midwest Studies in Philosophy*: *Studies in the Philosophy of Mind*, vol. 10 (pp. 615–78). Minneapolis: University of Minnesota Press.

Block, N. (1995). On a confusion about a function of consciousness. *Behavioral and Brain Sciences* 18, 227–287.

Burge, T. (1986). Individualism and psychology. *Philosophical Review* 95, 3–45.

Chalmers, D. (1996). *The conscious mind: In search of a fundamental th*eory. Oxford: Oxford University Press.

Chemero, A. (2009). *Radical embodied cognitive science*. Cambridge, MA.: MIT Press.

Chomsky, N. (1959). A review of B. F. Skinner's *Verbal Behavior*. *Language* 35, 26–58.

Cummins, R. (1996). *Representations, targets, and attitudes*. Cambridge, MA: MIT Press, 1996.

Dretske, F. (1981). *Knowledge and the flow of information*. Cambridge, MA: MIT Press.

Dretske, F. (1988). *Explaining behavior: Reasons in a world of causes*. Cambridge, MA: MIT Press.

Fodor, J. (1974). Special sciences. *Synthese* 28, 77–115.

Fodor, J. (1975). *The language of thought*. Cambridge, MA: Harvard University Press.

Fodor, J. (1980). Methodological Solipsism Considered as a Research Strategy in Cognitive Psychology. *Behavioral and Brain Sciences* 3, 63–73.

Fodor, J. (1987). *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*. Cambridge, MA: MIT Press.

Fodor, J. (1990). *A Theory of Content and Other Essays*. Cambridge, MA: MIT Press.

Fodor J. (1994). *The Elm and the Expert: Mentalese and Its Semantics*. Cambridge, MA: MIT Press.

Fodor, J. (1998). *Concepts: Where Cognitive Science Went Wrong*. Oxford: Oxford University Press.

Fodor, J., and LePore E. (1992). *Holism: A shopper's guide*. Oxford: Blackwell.

Forbes, G. (1989). Cognitive architecture and the semantics of belief. In P. French, T. Uehling, and H. Wettstein (Eds.), *Midwest studies in philosophy* XIV (pp. 84–100). Notre Dame: Notre Dame University Press.

Goldman-Rakic, P. (1987). Circuitry of Primate Prefrontal Cortex and Regulation of Behavior by Representational Memory. In F. Plum and V. Mountcastle (eds.), *Handbook of Physiology* (pp. 373-417) Vol. 5. Bethesda, MD: American Physiological Society.

Jackson, F. (1982). Epiphenomenal qualia. *Philosophical Quarterly* 32, 127–136.

Kriegel, U. (2011). *The sources of intentionality.* Oxford: Oxford University Press.

Kriegel, U. ms. Two notions of Mental Representation. Forthcoming in U. Kriegel (ed.), *Current Controversies in Philosophy of Mind* (New York: Routledge).

Kripke, S. (1980). *Naming and Necessity*. Cambridge, MA: Harvard University Press.

Levine, J. (1983). Materialism and qualia: The explanatory gap. *Pacific Philosophical Quarterly* 64, 354–61.

McDowell, J. (1994). The content of perceptual experience. *Philosophical Quarterly* 44, 190–205.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review* 63, 81–97.

Millikan, R. G. (1984). *Language, thought, and other biological categories*. Cambridge, MA: MIT Press.

Minsky, M. (1974). A framework for representing knowledge. MIT-AI Laboratory Memo 306.

Newell, A., and Simon, H. (1997). Computer science as empirical inquiry: Symbols and search. In J. Haugeland (ed.), *Mind design II: Philosophy, psychology, and artificial intelligence* (pp. 81–110). Cambridge, MA: MIT Press. Reprinted from the *Communication of the association for computing machinery*, 19 (March 1976), 113–26.

Port, R., and van Gelder, T. (Eds.). (1995). *Mind as motion*. Cambridge: MIT Press.

Pylyshyn, Z. (1984). *Computation and cognition: Toward a foundation for cognitive science*. Cambridge, MA: MIT Press.

Ramsey, W. (2007). *Representation reconsidered*. Cambridge: Cambridge University Press.

Rumelhart, D., McClelland, J., and the PDP Research Group. (1986). *Parallel distributed processing: Explorations in the microstructure of cognition,* Vol. 1, *Foundations*. Cambridge, MA: MIT Press.

Rupert, R. (1998). On the relationship between naturalistic semantics and individuation criteria for terms in a language of thought. *Synthese* 117, 95–131.

Rupert, R. (1999). The best test theory of extension: First principle(s). *Mind & Language* 14, 321–55.

Rupert, R. (2008). Frege's Puzzle and Frege Cases: Defending a Quasi-syntactic Solution.

Cognitive Systems Research 9, 76–91.

Rupert, R. (2009). *Cognitive systems and the extended mind*. Oxford: Oxford University Press.

Skinner, B. F. (1957). *Verbal behavior.* Acton, MA: Copley Publishing Group.

Stanford, P. K., and Kitcher, P. (2000). Refining the Causal Theory of Reference for Natural Kind Terms. *Philosophical Studies* 97, 99–129.

Stich, S. (1983). *From Folk Psychology to Cognitive Science: The Case against Belief*. Cambridge, MA: MIT Press.

Wittgenstein, L. (1953). *Philosophical investigations*. G. E. M. Anscombe and R. Rhees (eds.), G. E. M. Anscombe (trans.), Oxford: Blackwell.