

A Modest Role for Content (Draft)

Frances Egan
Rutgers University

Many theorists of mind would describe themselves as ‘representationalists’.

Representationalism, in its most general form, is the view that the human mind is an *information-using* system, and that human cognitive capacities are to be understood as representational capacities. I shall use *strong representationalism* to refer to the view that representational mental states have a specific form, in particular, that they are functionally characterizable relations to internal representations. Proponents of strong representationalism typically endorse the view that the system of internal representations constitutes a language with a combinatorial syntax and semantics. This latter view is sometimes known as the *Computational-Representational Theory of Mind*, though perhaps it would be better if the view’s commitment to sentence-like representations was more explicit. So let me call it *strong linguistic representationalism* (SLR for short).

With this distinction in place I shall explore what a commitment to representationalism, within the context of computational cognitive science, amounts to. I shall discuss a package of assumptions underlying what I call the ‘standard’ view, and argue that these assumptions are mistaken. I shall argue that representational content is to be understood as a *gloss* on the computational characterization of a cognitive process.

1 Representationalism in computational cognitive science – the standard view

The notion of a ‘representational device’ is given a precise meaning in computer science by Alan Newell’s (1980) characterization of a *physical symbol system*. A *physical symbol system* (hereafter, *PSS*) is a device that manipulates symbols in accordance with

the instructions in its program. *Symbols* are objects with a dual character: they are both physically realized and have meaning or semantic content. A realization function f_R maps them to physical state-types of the system. A second mapping, the interpretation function f_I , specifies their meaning by pairing them with objects or properties in a particular domain. A given PSS is type-individuated by the two mappings f_R and f_I . By this I mean that if either f_R and f_I had been different, the device would be a different computational mechanism.

The concept of a PSS gives precise meaning to two notions central to mainstream cognitive science: *computation* and *representation*. A *computation* is a sequence of physical state transitions that, under the mappings f_R and f_I , executes some specified task. A *representation* is an object whose formal and semantic properties are specified by f_R and f_I respectively.¹

A PSS, Newell emphasizes, is a universal machine. Given sufficient, but finite, time and memory it is capable of computing any computable function. These systems have what Fodor and Pylyshyn (1988) have called a ‘classical’ architecture, an architecture that preserves a principled distinction between the system’s representations or data structures and the processes defined over them.

The *physical symbol systems hypothesis* is the idea that the mind is a specific sort of computer, namely, a device that manipulates (writes, retrieves, stores, etc.) strings of symbols. The PSS hypothesis is a version of *strong linguistic representationalism* (SLR), the idea that representational mental states – paradigmatically, beliefs, desires,

¹ Following Fodor’s 1980 usage, “formal” here will mean *non-semantic*.

and the other propositional attitudes – are functionally characterizable relations to internal representations with syntactic and semantic properties.

It is not hard to understand the attraction of the PSS hypothesis for philosophers of mind and psychology. Proponents of strong linguistic representationalism such as Fodor (1975, 1981, 1987) and Pylyshyn (1984), have hoped that computational models of cognitive processes will eventually mesh with and provide a scientific explanation of our commonsense explanatory practices. These practices appeal to content-specific beliefs and desires. For example, it is my belief that there is beer in the refrigerator, together with a content-appropriate desire (to drink a beer, or perhaps just to drink something cold), that explains my going to the kitchen and getting a beer. Appealing to my belief that there is beer at the local bar or my desire to win the lottery fails to provide any explanation of my beer-fetching behavior. Moreover, this behavior is rational just to the extent that it is caused by content-appropriate beliefs and desires. Similarly, according to PSS-inspired SLR, computational explanations of behavior will appeal to the contents of the symbol strings, or internal representations, the manipulations of which are the causes of our intelligent behavior. But these operations themselves respect what Fodor (1980) has dubbed the ‘formality condition’ – they are sensitive only to *formal* (i.e. *non-semantic*) properties of the representations over which they are defined, not to their content.

The formality condition is often glossed by strong linguistic representationalists as the idea that mental representations have their causal roles in virtue of their syntax. As Pylyshyn (1984) puts the point,

For every apparent, functionally relevant distinction there is a corresponding syntactic distinction. Thus, any semantic feature that can conceivably affect behavior must be syntactically encoded at the level of a formal symbol structure. By this means we arrange for a system's behavior to be describable as responding to the content of its representations – to what is being represented – in a manner compatible with materialism. (74)

The idea of syntax and semantics marching in lock-step, to produce mechanical reasoning, is of course the fundamental idea underlying theorem proving in logic.

Let us focus on the role of so-called 'representational content' in computational models of cognitive capacities. Representationalists (of all stripes) tend to endorse the following claims:

- (1) The internal states and structures posited in computational theories of cognition are *distally interpreted* in such theories; in other words, the domain of the interpretation function f_i is objects and properties of the external world.
- (2) The relation between the posited internal states and structures and the distal objects and properties to which they are mapped (by f_i) – what we might call the *Representation Relation* – is a substantive, naturalistically specifiable (perhaps a causal or teleological) relation.²
- (3) The distal objects and properties which determine the representational content of the posited internal states and structures serve to *type-individuate* a computationally

² See, for example, Dretske 1981, 1986, Fodor 1990, Millikan 1984, Papineau 1987, 1994, among many others.

characterized mechanism. In other words, if the states and structures had been assigned *different* distal contents then it would be a *different* computational mechanism.

I shall call this package of commitments the *Essential Distal Content View*. I shall argue that the Essential Distal Content View is false. It fundamentally misconstrues both the nature of the interpretation function f_I and the role of so-called ‘representational content’ in computational accounts of cognition.

2 The Chomskian challenge to representationalism

Noam Chomsky has argued in recent work that the so-called ‘representational’ states invoked in accounts of our cognitive capacities are not genuinely representational, that they are not about some represented distal objects or properties. Discussing Shimon Ullman’s (1979) work on visual perception he says,

There is no meaningful question about the “content” of the internal representations of a person seeing a cube under the conditions of the experiments... or about the content of a frog’s “representation of” a fly or of a moving dot in the standard experimental studies of frog vision. No notion like “content”, or “representation of”, figures within the theory, so there are no answers to be given as to their nature. The same is true when Marr writes that he is studying vision as “a mapping from one representation to another...” (Marr 1982, p.31) – where “representation” is not to be understood relationally, as “representation of”. (1995, 52-3)

The idea that “representation” should, in certain contexts, not be understood relationally, as in “representation of x”, but rather as specifying a monadic property, as in

“x-type representation”, can be traced to Goodman 1968.³ So understood, the individuating condition of a given internal structure is not its relation to an ‘intentional object’, there being no such thing according to Chomsky, but rather its role in cognitive processing. Reference to what looks to be an intentional object is simply a convenient way of type-identifying structures with the same role in computational processing.

The point applies as well to the study of the processes underlying linguistic capacities:

... here too we need not ponder what is represented, seeking some objective construction from sounds or things. The representations are postulated mental entities, to be understood in the manner of a mental image of a rotating cube, whether the consequence of tachistoscopic presentations or of a real rotating cube or of stimulation of the retina in some other way, or imagined, for that matter. Accessed by performance systems, the internal representations of language enter into interpretation, thought, and action, but there is no reason to seek any other relation to the world... (Chomsky 1995, 53)

³ According to Goodman,

Saying that a picture represents a so-and-so is thus highly ambiguous as between saying what the picture denotes and saying what kind of picture it is. Some confusion can be avoided if in the latter case we speak rather of ... a ‘Pickwick-picture’ or ‘unicorn-picture’ or ‘man-picture’. Obviously a picture cannot, barring equivocation, both represent Pickwick and represent nothing. But a picture may be of a certain kind – be a Pickwick-picture or a man-picture – without representing anything.” (1968, 22)

Goodman claims that the locution “representation of Pickwick” is syntactically ambiguous. On one reading it has the logical form of a one-place ‘fused’ predicate – ‘Pickwick-representation’ – where “Pickwick” is, in Quine’s (1960) terminology, ‘syncategorematic’. Chomsky is not committed to this syntactic thesis, but rather to the claim that locutions of the form “representation of x” are *semantically* ambiguous.

Chomsky rejects the idea that intentional attribution – the positing of a domain of objects or properties to which internal structures stand in a *meaning* or *reference* relation – plays any explanatory role in cognitive science. Intentional construals of David Marr’s (1982) theory of vision, such as Burge (1986), Chomsky claims, are simply a misreading, based on conflating the theory proper with its informal presentation. As Chomsky puts it, “The theory itself has no place for the [intentional] concepts that enter into the informal presentation, intended for general motivation.” (1995, 55)

Chomsky himself has not spelled the argument out explicitly, though the motivation for his recent anti-representationalism is not hard to find.⁴ As theories of our perceptual and linguistic capacities have become increasingly removed from commonsense, it becomes quite forced to say that the subject *knows* or *believes*, say, the *rigidity assumption* (Ullman 1979) or the *minimal link condition* (Chomsky 1995). Chomsky 1975 was willing to say that subjects ‘cognize’ the principles posited in cognitive theories, but these contents – *that objects are rigid in translation* or *that derivations with shorter links are preferred over derivations with longer links* – do not look like the sorts of things that subjects could plausibly be said to know, believe, etc. They are not inferentially promiscuous, not accessible to consciousness, and so on.

It is particularly unclear what independent objects, if any, the structures posited in accounts of our linguistic capacities represent. Among the candidates are elements of the public language⁵, elements of the speaker’s idiolect, or, as Georges Rey (2003a, 2003b, 2005) has recently suggested, linguistic entities such as nouns, verb phrases, phonemes,

⁴ Though see Collins 2007 for the view that Chomsky has always been an anti-representationalist.

⁵ Chomsky himself is skeptical of the notion of a ‘shared public language’. See the papers in Chomsky 2000.

and so on – what Rey calls ‘standard linguistic entities’ (SLEs). SLEs, Rey argues, are to be understood as ‘intentional inexistent’, objects of thought, like Zeus or Hamlet, that don’t exist. Discussion of the merits and demerits of these various proposals is beyond the scope of the present paper. Chomsky, for his part, rejects them all, insisting that talk of represented objects is intended simply for informal exposition and plays no genuine explanatory role in the theory.

3 Rethinking the standard view

Chomsky is, in effect, a *methodological eliminativist* about representational content. He denies that the internal structures posited in computational theories are distally interpreted as representations of external objects and properties (i.e. Claim 1 of the Essential Distal Content View), and hence that computational mechanisms are type-individuated by a domain of external objects and properties (Claim 3).⁶ Any reference to such a domain in computational accounts, he claims, is merely “informal presentation, intended for general motivation.”

In what follows I shall spell out a view of representation in computational cognitive theories according to which Chomsky is correct in denying the Essential Distal Content View, but nonetheless wrong in denying to representational content a genuine explanatory role. Chomsky’s view fails to make clear the role played by the interpretation function f_I in computational accounts, and leaves mysterious how representational content could aid in the ‘general motivation’ of a computational theory.

⁶ It is consistent with Chomsky’s stated views that there *is* a substantive, naturalistically specifiable relation between posited structures and distal objects and properties (Claim 2), though Chomsky himself would reject such speculation as a manifestation of ‘methodological dualism’.

These points will be illustrated by reference to two computational theories drawn from different cognitive domains.⁷

David Marr's well-known explanatory hierarchy distinguishes three distinct levels at which a computational account of a cognitive capacity is articulated. Disputes about whether computational theories type-individuate the mechanisms they characterize by their representational content turn on how the level of description that Marr called the *theory of the computation* should be interpreted. The theory of the computation provides a *canonical description* of the function(s) computed by the mechanism. It specifies what the device does. By a "canonical description" I mean the characterization that is decisive for settling questions of type-individuation or taxonomy. The canonical description is given by the interpretation function f_i ; the canonical description is therefore a semantic characterization. But this is the important point: the canonical description of the function computed by a computationally characterized mechanism is a *mathematical* description. A couple of examples illustrate the point.

Marr (1982) describes a component of early visual processing responsible for the initial filtering of the retinal image. Although there are many ways to informally describe what the filter does, Marr is careful to point out that the theoretically important characterization, from a computational point of view, is a mathematical description: the device computes the Laplacean convolved with the Gaussian (1982, 337). As it happens, it takes as input light intensity values at points in the retinal image, and calculates the rate of change of intensity over the image. But this *distal* characterization of the task is, as Chomsky might put it, an 'informal' description, intended for general motivation. *Qua*

⁷ See Egan 1995b, 1999, 2003 for defense of this account of the role of content in David Marr's theory, in particular.

computational device, it does not matter that input values represent *light intensities* and output values the rate of change of *light intensity*. The computational theory characterizes the visual filter as a member of a well-understood class of mathematical devices that have nothing essentially to do with the transduction of light.

The second and third levels of Marr's explanatory hierarchy describe, respectively, a *representation and algorithm* for computing the specified functions, and the circuitry or neural hardware that implement the computation. Marr's account of early visual processing posits primitive symbol tokens – edges, bars, blobs, terminations, and discontinuities – and selection and grouping processes defined over them. It is at this second level that the theory posits symbol structures or representations, and processes defined over them. These symbol structures (edges, bars, blobs, etc.) and the processes that operate on them are type-individuated by the mapping f_R , which characterizes them at the level of physical states and processes, independent of the cognitive capacities that they subserve.

The second example, from a different cognitive domain, is Shadmehr and Wise's (2005) computational theory of motor control. Consider a simple task involving object manipulation. (See figure 1) A subject is seated at a table with eyes fixated ahead. The hand or *end effector* (ee) is located at X_{ee} , and the target object (t) at X_t . The problem is simply how to move the hand to grasp the object. There are an infinite number of trajectories from the hand's starting location X_{ee} to the target at X_t . But for most reaching and pointing movements, the hand moves along just one of these trajectories: specifically, it typically moves along a straight path with a smooth velocity. Shadmehr and Wise (2005) describe one way in which the task might be accomplished.

The overall problem can be broken down into a number of sub-problems. The first problem is *how does the brain compute the location of the hand?* *Forward kinematics* involves computing the location of the hand (X_{ee}) in visual coordinates from proprioceptive information from the arm, neck, and eye muscles, and information about the angles of the shoulder and elbow joints. Informally, this process coordinates the way the hand looks to the subject with the way it feels. The brain also has to compute the location of the target (X_t), using retinal information and information about eye and head orientation.

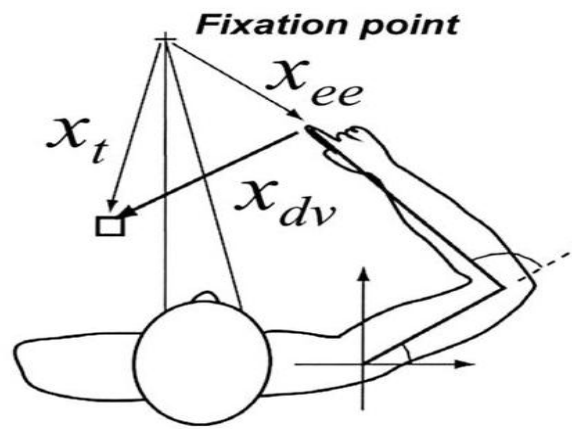


Figure 1

The second problem, computing a plan of movement, involves computing the *difference vector*, that is, the displacement of the hand from its current location to the target's location. But this 'high level' plan specifies a displacement of the hand in visual coordinates. This visually oriented plan has to be transformed into a specification of the joint rotations and muscle forces required to effect the displacement. So, the third problem, involving the computation of *inverse kinematics and dynamics*, is how the high

level motor plan, corresponding to a difference vector, is transformed into joint angle changes and force commands. Reaching and pointing movements involve continuous monitoring of target and hand location, with the goal of reducing the difference vector to zero. There are a number of complicating factors. Incidental eye and head movements require continuous updating of the situation. Deceleration of the hand should be smooth, to avoid knocking over the target.

Summarizing, the account decomposes the overall task into three computations, and specifies the function computed in each in precise mathematical terms:

- (1) $f(\theta) = X_{ee}$, *forward kinematics*, the computation of hand location, in eye-centered coordinates, from proprioceptive information and information about joint angles;
- (2) $X_t - X_{ee} = X_{dv}$, *the difference vector*, the difference between the target location and initial hand position in eye-centered coordinates; and
- (3) $f(X_{dv}) = \Delta\theta$, *inverse kinematics*, the computation from the high-level movement plan, in eye-centered coordinates to a required change of joint angles.

The motor control mechanism characterized by Shadmehr and Wise is not a physical symbol system; its operations are not interpreted in the account as manipulations of symbols. Nor does the account of the mechanism's implementation decompose neatly into representation and algorithm (Marr's level 2) and neural realization (Marr's level 3). Rather, the three computations that constitute the motor control mechanism are characterized as analog processes and realized in neural networks in the posterior parietal cortex, the premotor cortex, and the primary motor cortex respectively. The details need not concern us here.

The important point is that in both examples, the canonical description of the task executed by the device, the function(s) computed, is a mathematical description. As noted above, this description characterizes the mechanism as a member of a well-understood class of mathematical devices. A crucial feature of this characterization is that it is ‘environment neutral’: the task is characterized in terms that prescind from the environment in which the mechanism is normally deployed. The mechanism described by Marr computes the Laplacean of the Gaussian whether it is part of a visual system or an auditory system, in other words, independently of the environment – even the *internal* environment – in which it is normally embedded. In fact, it is not implausible to suppose that each sensory modality has one of these same computational mechanisms, since it just computes a curve-smoothing function. The same point holds for the motor control mechanism characterized by Shadmehr and Wise. A mariner who knew the distance and bearing from his home port to his present location and the distance and bearing from his home port to a buried treasure could perform the same computation to compute the course from his present location to the treasure. In both cases, it is the abstract mathematical description that type-individuates the mechanism or process, not what Chomsky would call the ‘informal’ description that characterizes the mechanism as computing *changes of light intensities* or the *displacement between target and hand location*.

To summarize: The characterization of a computational process or mechanism made available by the interpretation function f_I – the mapping that provides a canonical description of the function computed by the mechanism, and hence (along with the realization function f_R) serves to type-individuate it – is (*pace* claim 3 of the Essential

Distal Content View) an abstract mathematical description. This semantic interpretation does not provide a *distal* interpretation of the posited internal states and structures; the specified domain is not external objects and properties (as claim 1 of the Essential Distal Content View holds), but rather mathematical objects. The interpretation maps the states and structures to a domain of *abstracta*, hence the specified relation is not regarded, in the theory, as a Naturalistic relation (as claim 2 holds). It cannot be a causal relation, since abstracta have no causal powers.

If this account is correct, then what should we make of the idea that visual states represent such visible distal properties as *depth* and *surface orientation*, and motor control states represent *hand location* and *shoulder angle*? Are such distal contents explanatorily idle, as Chomsky claims? And if they aid in “general motivation”, how precisely do they do that?

Ordinary, distal, representational contents serve several important explanatory functions. The questions that antecedently define a psychological theory’s domain are usually couched in intentional terms. For example, we want a theory of vision to tell us, among other things, how the visual system can detect three-dimensional distal structure from information contained in two-dimensional images. A characterization of the postulated computational processes in terms of distal objects and properties enables the theory to answer these questions. This characterization tells us that states of the system co-vary, in the normal environment, with changes in depth and surface orientation. It is only under an interpretation of some of the states of the system as representations of depth and surface orientation that the processes given a formal, mathematical, characterization by a computational theory are revealed as *vision*. Thus, content

ascription plays a crucial *explanatory* role: it is necessary to explain how the operation of a formally characterized, mathematical, process constitutes the exercise of a cognitive capacity in the environment in which the process is normally deployed. The device would compute the same mathematical function in any environment, but only in some environments would its doing so enable the organism to see.

This is the most important function of representational content. Because the ascription of distal contents is necessary to explain how a computational process constitutes the exercise of a cognitive capacity in a particular context, I shall call the interpretation that enables the assignment of such distal contents the *cognitive interpretation*. The cognitive interpretation is to be sharply distinguished from the interpretation specified by f_j . Only the latter plays an individuating role.

To recap: When the computational characterization is accompanied by an appropriate cognitive interpretation, in terms of distal objects and properties, we can see how the mechanism that computes a certain mathematical function can, in a particular context, subserve a cognitive function such as vision or reaching and pointing. So when the input states of the Marrian filter are described as representing *light intensities* and the output states *changes of light intensity* over the image, we can see how this mechanism enables the subject to detect significant boundaries in the scene. When the input states of the mechanism that computes inverse kinematics are described as representing *displacement in visual space* and the output states *changes in joint angles* we can see the role that the mechanism plays in the subject's successfully grasping the target object.

The account presented here draws a sharp distinction between the computational theory proper – the mathematical description made available by the mapping f_L , which (along with f_R) type-individuates the mechanism – and the distal characterization that accompanies it and explains the contribution of the abstractly characterized mechanism to the larger cognitive life of the organism. We can also understand how representational content, while not type-individuating computational mechanisms, can, as Chomsky puts it, provide “general motivation” for the theory.

The cognitive characterization is essentially a *gloss* on the more precise account of the mechanism provided by the computational theory. It forms a bridge between the abstract, mathematical characterization that constitutes the explanatory core of the theory and the intentionally characterized pre-theoretic explananda that define the theory’s cognitive domain. Unless the processes and/or structures given a precise mathematical specification in the theory are construed, under interpretation, as representations of such distal properties as *edges*, or *joint angles*, the account will be unable to address the questions that motivated the search for a computational theory in the first place, such questions as *how are we able to see the three-dimensional structure of the scene from two dimensional images?*, or *how are we able to move our hand to grasp an object in sight?*

To call the cognitive characterization a ‘gloss’ is not to suggest that the ascription of representational content is unprincipled. The posited states and structures are not interpretable as representations of distal visible properties (as, say, *object boundaries*, or *depth* or *surface orientation*) unless they co-vary with tokenings of these properties in the subject’s immediate environment. It would be a mistake, though, to conclude that the structures posited in computational vision theories must (even in the gloss) represent their

normal distal cause, and to find in these accounts support for a *causal* or *information-theoretic* theory of content.⁸ Some structures – zero-crossings in Marr’s account, for example – may be interpreted as representations of proximal features, in particular, as *discontinuities in the image*. The ascription of content is sometimes driven by purely expository considerations, such as allowing us to keep track of what the process is doing at points in the processing where the theory posits structures that do not correlate neatly with a salient distal property tokening. Even within a single content assignment (cognitive interpretation), no single, privileged relation is assumed to hold between posited structures and the elements to which they are mapped. The choice of a cognitive gloss is governed by explanatory considerations, which we can, following Chomsky, characterize as “informal motivation.”

An implication of the foregoing account of the role of representational content in computational models is that cognitive science has no need for a Naturalistic Semantics – the specification of non-intentional and non-semantic sufficient conditions for a mental state’s having the meaning it does.⁹ Whatever the *philosophical* interest of such an account, it would hold little interest for the computational cognitive theorist. As the example in the next section illustrates, there is no reason to think that there are naturalistic sufficient conditions for the assignment of representational content. Rather, the choice of a cognitive interpretation will typically depend on a variety of pragmatic considerations.

⁸ See Dretske 1981 and Fodor 1990 for examples of information-theoretic accounts of content.

⁹ See fn. 2.

One other type of case is worth mentioning. Computational mechanisms are sometimes characterized in distal terms even though no states of the mechanism are themselves interpreted. Shimon Ullman's (1979) *structure from motion mechanism* is described as assuming that objects are rigid in translation even though no state or structure in the mechanism is actually assigned this content. The mechanism simply computes the so-called 'structure from motion' function: given three distinct views of four non-coplanar points it computes the unique rigid structure compatible with the points. In a world where objects are rigid in translation the output of the mechanism is interpretable as veridical. In a world populated with gelatinous objects that deform in translation its outputs would often not be veridical. Here, a cognitive characterization of the mechanism provides a useful, though informal, way to understand what it does.

4 Cognitive interpretation as gloss: an illustration

That a cognitive interpretation *is* a gloss may not be obvious if we consider only the normal case, so I will use a fanciful example to illustrate the point.

Whether a computationally characterized device succeeds in computing, say, the depth of objects and surfaces in the scene from information about the disparity of points in the retinal image depends on whether its internal states covary with changes of depth in the environment. This requires a certain *fit* between the mechanism and the world. In the actual world, the fit is a product of natural selection. Let us call a computational visual mechanism adapted to the terrestrial environment 'Visua' and the distal property it tracks – changes of depth – 'C1'.¹⁰

¹⁰ This is an adaptation of an example from Segal (1989).

A given computational mechanism would not enhance fitness in every environment. Being an adaptation is a contingent property of a computationally characterized system. We can imagine a duplicate of Visua, Twin Visua, appearing spontaneously (perhaps by random mutation) in a world to which it is not adapted. Imagine that the counterfactual world, E2, is different enough from the actual world that Twin Visua's states track some distal property other than changes in depth. Call this property 'C2'. Since Visua and Twin Visua are physical duplicates, the two mechanisms have the same discriminative and recognitional capacities. Visua would track C2 if it were transported to E2. Twin Visua will contribute to the fitness of the organism containing it only if C2 is a useful property to track or represent in E2. C2 is some function of surfaces, local light, and local optical laws, but tracking C2 might not allow the organism containing Twin Visua to recover what is where in the scene. If it does not, we might wonder whether it is even appropriate to call Twin Visua a *visual* mechanism.¹¹

The important question for present purposes is 'what do Visua's and Twin Visua's internal states represent?' It is natural to say that Visua represents C1 – changes of depth. It is in virtue of tracking changes in depth in the scene that Visua contributes to the organism's successful interaction with its environment. Perhaps it is also natural to say that Twin Visua represents the distinct property C2. In any case, it would be odd to say that Visua represents some more general property C3 that subsumes both changes of

¹¹ If Twin Visua is not a visual mechanism, then Visua is a visual mechanism only contingently. Since the cognitive interpretation of a computational mechanism is not an essential component of a computational account (in the sense that a different cognitive interpretation may appropriately describe the mechanism in a different context) computational mechanisms are cognitive mechanisms – they subserve particular cognitive functions – only contingently.

depth (C1) and the strange and (from the terrestrial perspective) unnatural property C2.¹² In other words, there is a clear rationale for attributing to Visua and Twin Visua distinct, environment-specific contents that make apparent the contribution of the mechanism to the success of the organism in its normal environment, rather than an unobtrusive, more general content that does not.

To summarize: Visua and Twin-Visua are type-identical computational mechanisms. The computational characterization that specifies the mechanism's basic causal operations (i.e. the mapping f_R) and characterizes in precise terms the mathematical function(s) it computes (i.e. the mapping f_I) subsumes both of them. This environment-general characterization allows us to predict and explain how a mechanism would behave in counterfactual worlds. But it doesn't tell us what, if anything, the mechanism would *represent* in other environments. The content-determining correlations are those that obtain between states of the device and property tokenings in the local environment. The correlations that obtain in counterfactual environments, where the objects, properties, and laws may be quite different, are not relevant to what I am calling the 'cognitive interpretation' of the mechanism.

I claimed above that it is *natural* to ascribe environment-specific contents to cognitive mechanisms. The reason is not hard to find. The questions that antecedently or pre-theoretically define a cognitive theory's domain are typically framed in terms that presuppose the organism's success in its normal environment. We want to know how the organism can recover the depth and orientation of the objects in the scene from

¹² A content internalist might claim that Twin Visua represents C1, but this seems unmotivated, since Twin Visua's states may never co-vary with C1 in E2. E2 is not a world where Twin Visua (or Visua, for that matter) sees depth.

information in two-dimensional retinal images. The theorist of cognition typically sets out to answer questions that are already framed in environment-specific terms. If the mechanism's states are interpreted as *representing* depth and orientation, rather than more general properties determined by correlations that obtain in counterfactual environments, then the theory is poised to answer these questions.

The cognitive interpretation of a computational mechanism can be responsive to these explanatory and pragmatic considerations, addressing the questions that initially motivated the search for and development of the theory, only because the computational characterization given by f_R and f_I , provides an environment-independent characterization of the device. As noted above, this characterization provides the basis for predicting and explaining its behavior in any environment. The cognitive interpretation can serve our more parochial interests, most importantly, explaining the cognitive tasks that are typically characterized in terms that presuppose the device's success in its local environment.

Let us return to Visua and Twin Visua. Suppose now that the property tracked by Twin Visua in E2 – property C2 – is a useful property for an organism in E2 to detect. Twin Visua therefore contributes to the fitness of the organism containing it. Imagine that an enthusiastic editor on earth (E1), always on the lookout for new markets, asks the theorist responsible for characterizing Visua to produce a textbook that could be marketed and sold in both E1 and E2. Since Visua and Twin Visua are computationally identical mechanisms – the mappings f_R and f_I that characterize Visua will apply to Twin Visua as well – the theorist needs only to produce a single cognitive interpretation that specifies what this formally characterized mechanism will represent in E1 and E2. Since

the mechanism does not track C1 in E2 or C2 in E1, neither *C1* nor *C2* are plausible candidates for the content. Rather, a cognitive interpretation appropriate to both worlds would take the mechanism to represent some more general property *C3* that subsumes both *C1* and *C2*.¹³

Notice, first, that the content *C3* is still a wide content. The new cognitive interpretation specifies what the mechanism represents in E1 and E2, but not what a physically indistinguishable mechanism might represent in some third environment E3. (This follows by an iteration of the reasoning above.) While nonetheless wide, *C3* is, in a sense, *narrower* than either *C1* or *C2*. *C3* prescind from the environmental differences between E1 and E2. The explanatory interests served by the new interpretation are less local, less *parochial*, than those served by the original interpretation, which was designed to address questions posed in vocabulary appropriate to earth. Whereas the old cognitive interpretation enabled the theory to address such pre-theoretic questions as “how is the organism able to recover the depth of objects in the scene” by positing representations of *depth*, the new interpretation provides the basis for answering this question and an analogous question framed in vocabulary appropriate to E2 – “how is the organism able to recover information about *C2*?” – by positing representations of the more general distal property *C3*, and supplying auxiliary assumptions about how *C3* is related to the locally instantiated properties *C1* and *C2*.

As it happened, the over-eager editor was somewhat surprised that sales on earth of the new inter-planetary textbook fell off rather sharply from the first edition, designed solely for the local market. Besides introducing a new vocabulary containing such

¹³ *C3* may be understood as the disjunction of *C1* and *C2*, or as a determinable that has *C1* and *C2* as determinates. In any case, it is a distal content.

unfamiliar predicates as “C3”, the new edition required cumbersome appendices appropriate to each world, explaining how to recover answers to questions about the organism’s capacities in its local environment, questions that motivated the search for an explanatory theory in the first place. Readers complained that the new edition was much less “user-friendly”.

The editor was therefore dissuaded from her original idea of commissioning an intergalactic version of the text, which would provide a genuinely narrow cognitive interpretation that would specify what *Visua* would represent in *any* environment.¹⁴ She came to realize that a distal interpretation of computationally characterized processes is primarily a *gloss* that allows a theory to address local explanatory interests. Any gloss that shows that the theory is doing its job will be couched in a vocabulary that is perspicuous for the local audience with these interests. An important moral here is that a truly intergalactic computational cognitive science would not be *representational* in the following sense: it is not likely to assign anything that looks remotely like ordinary representational content.

5 Concluding remarks on representationalism

My exposition of the role of representational content in computational cognitive theories relies heavily on Newell’s notion of a physical symbol system; though the account is intended to be completely general. *Strong representationalism*, recall, construes human cognitive states as relations to internal representations, thereby positing a sharp

¹⁴ Instead the editor commissioned an environment-specific cognitive interpretation for each world, to accompany the environment-neutral account of the mechanism provided by f_R and f_I .

distinction – inherent in the notion of a physical symbol system – between data structures or representations on the one hand, and the processes defined over them on the other.

This view certainly has its attractions, not the least of which is that it purports to explain *how* thinkers can be information-using systems. They can use information by using representations of that information. This idea requires a distinction between the part of the system that uses representations and the representations themselves, which is exactly what the data structure/process distinction enforces.

A real attraction of cognitive models that conform to strong representationalism is their *explanatory transparency*. Symbols are structures ready-made for semantic interpretation – they just *are* objects with both formal and semantic properties, characterized by f_R and f_I respectively. Symbol structures – representations – are, in effect, ‘hooks’ on which interpretations can be hung. Moreover, the information in physical symbol systems is *accessible for use*, encoded in exactly the features of the structures to which computational processes are sensitive. Thus, physical symbol systems are said to *explicitly represent* the information that they encode. And we can track the flow of information in the system by keeping track of the operations on the encoding structures.¹⁵ This is another important explanatory role served by what I have been calling a ‘cognitive interpretation’

The structure/process distinction inherent in Strong Representationalism is not inevitable. There are ways to eschew the distinction while preserving the central idea of (regular strength) representationalism – that the mind is an *information-using device*.

The Shadmehr and Wise motor control mechanism described above is an example. It is

¹⁵ This is the point of positing what Egan and Matthews 2006 call ‘intentional internals’.

not a physical symbol system; its computations are characterized as analog processes and realized in neural networks. Parallel distributed processing (PDP) systems, analog relaxation systems, massive cellular automata, and other kinds of computational mechanisms for which the structure/process distinction is not preserved are *not* ready-made for interpretation. Often, in PDP systems, no distinct state or part of the network serves to represent any particular object, property, or proposition. Rather, the encoding of information is distributed over many units, connection strengths, and biases, with the result that the representation of any given object, property, or proposition is widely scattered throughout the network. It becomes quite forced to talk of ‘representations’ in such systems, given that our paradigm of a representation is the printed word, a discrete object that is spatially compact, movable, and, most important, meaningful (but only under interpretation). These systems are far from explanatorily transparent. One often cannot tell by looking at the computational and engineering details of the system which of its spatiotemporal parts are candidates for interpretation. It can be quite difficult to track the flow of information in these systems. But the point of cognitive interpretation is the same as for PSS systems – to make the computational processes perspicuous.

References

- Burge, T. (1986). Individualism and psychology. *The Philosophical Review*, 95.
- Chomsky, N. (1975). *Reflections on Language*. Pantheon Books.
- Chomsky, N. (1995). Language and nature. *Mind*, 104.
- Chomsky, N. (2000). *New Horizons in the Study of Language and Mind*. Cambridge University Press.
- Collins, J. (2007), Meta-scientific eliminativism: a reconsideration of Chomsky's review of Skinner's *Verbal Behavior*. *British Journal for the Philosophy of Science*, 58: 625-58.
- Dretske, F. (1981). *Knowledge and the Flow of Information*. MIT Press.
- Dretske, F. (1986). Misrepresentation. In R. Bogdan (ed.), *Belief: Form, Content and Function*. Oxford University Press.
- Egan, F. (1995). Computation and content. *The Philosophical Review*, 104.
- Egan, F. (1999). In defense of narrow mindedness. *Mind and Language*, 14.
- Egan, F. (2003). Naturalistic inquiry: where does mental representation fit in?. In L. Antony and N. Hornstein (eds.), *Chomsky and his Critics*, Blackwell.
- Egan, F. and Matthews, R. (2006). Doing Cognitive Neuroscience: A Third Way. *Synthese*, 153: 377-91
- Fodor, J. (1975). *The Language of Thought*. Thomas Y. Crowell.
- Fodor, J. (1980). Methodological solipsism considered as a research strategy in cognitive science. In *Behavioral and Brain Sciences*, 3.
- Fodor, J. (1981). *Re:Presentations: Philosophical Essays on the Foundations of Cognitive Science*. MIT Press.

- Fodor, J. (1987). *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*. MIT Press.
- Fodor, J. and Pylyshyn, Z. (1988). Connectionism and cognitive architecture: a critical analysis. *Cognition*, 28.
- Goodman, N. (1968). *Languages of Art*. Bobbs-Merrill.
- Marr, D. (1982). *Vision*. Freeman.
- Millikan, R. (1984). *Language, Thought, and other Biological Categories*. MIT Press.
- Newell, A. (1980). Physical symbol systems. *Cognitive Science*, 4.
- Papineau, D. (1987). *Reality and Representation*. Oxford University Press.
- Papineau, D. (1993). *Philosophical Naturalism*. Blackwell.
- Pylyshyn, Z. (1984). *Computation and Cognition: Toward a Foundation for Cognitive Science*. MIT Press.
- Quine, W. (1960). *Word and Object*. MIT Press.
- Rey, G. (2003a). Chomsky, intentionality, and a CRTT. In L. Antony and N. Hornstein (eds.), *Chomsky and his Critics*, Blackwell.
- Rey, G. (2003b). Intentional content and a Chomskian linguistics. In A. Barber (ed.), *Epistemology of Language*, Oxford University Press.
- Rey, G. (2005). Mind, intentionality, and inexistence: an overview of my work. In *The Croatian Journal of Philosophy*, 5.
- Shadmehr, R. and Wise, S. (2005). *The Computational Neurobiology of Reaching and Pointing: A Foundation for Motor Learning*. MIT Press.
- Ullman, S. (1979). *The Interpretation of Visual Motion*. MIT Press.