Form and Content in Mental Representation

Mark Simms

Philosophy, Adelaide University

June, 2004

Acknowledgements

Abstract

It is orthodoxy in contemporary philosophy of cognitive science to hold that the human brain processes information, both about the body in which the brain is located and about the world more generally. The internal states of the brain that encode this information are known as mental representations. Two matters concerning mental representation are interwoven here: the role of representational content in cognition and the format of mental representation. Robert Cummins, among others, argues that content is intrinsic to mental representation, rather than involving matters external to a representation, such as the use to which the representation is put. He also holds that resemblance accounts of representation best make sense of this fact. Thus, according to Cummins, the content of a mental representation is determined by its form.

This thesis argues that an account of representation requiring that representations possess resembling structure is unlikely to be correct given (a) the minimal requirements that something must meet in order to count as a mental representation, (b) the tasks required of representation in cognition, such as capturing abstract properties, combining with other representations, and tracking change, and (c) the possibility that content stands in a different relation to form and cognition from the one Cummins has in mind. In criticising Cummins, however, this thesis explores possible implementations of resemblance theories in connectionist representation. It also redraws his map of the psychosemantic field to suggest that classical theories of cognition, which posit concatenative schemes of symbolic representation, share some of the benefits of tying content to form. Finally, in exploring various notions of the role of form in representation, this thesis also advocates a pluralistic approach to the mental representations implicated in human cognition.

This thesis contains no material which has been accepted for the award of any other degree or diploma in any university and to the best of my knowledge and belief, the thesis contains no material previously published or written by another person, except where due reference is made in the text of the thesis.

I give consent to this copy of the thesis, when deposited in the Library, being available for photocopying and loan.

Mark Simms

Introduction: Form and Content

How do mental representations work?[1] What role do a mental representation's form and content play in cognition? How does form relate to content? This thesis considers various positions on these matters. In doing so, it deals with two important issues in the philosophy of mental representation: firstly, the matter of whether content is intrinsic to a representation or is, instead, dependent on the role that representation plays or how that representation is used by the cognitive system it is a part of. Secondly, if content is not a matter of role or use, this thesis asks if resemblance accounts of representational content are the only ones consistent with this fact. In effect, this thesis considers whether we should believe that a representation's content is necessarily exhausted by its form. This thesis's secondary consideration is how this idea plays out in the connectionist paradigm of representation and computation.

I consider three versions of the resemblance account of content: one general and based on Cummins (1996), and two 'implementational' and couched in the connectionist vision of representation and cognition. The latter two are a resemblance interpretation of State Space Semantics, and O'Brien and Opies' work on resemblance representation, which I describe using the term 'Weight State Semantics'. I contrast the explanatory power of these positions with the classicist's view of content and form. I take the classicist position on matters of resemblance to be that complex content is reflected only grossly in the structure of complex mental representations- for example, as in the way that the structure of a phrase like 'red house' might reflect the two-part structure of that concept (or, even the two-part nature of the thing in the world). I argue that, ironically, in State Space Semantics[2], structure in content is represented more obliquely than the kind of explicit representation of basic structure that we find in classical, concatenative representational schemes.

---

[1] For the purposes of this thesis I will more or less assume some version or other of the representational theory of mind is correct- that there are object-like states in the brain that have intentional and syntactic properties.

[2] I don't mean here that State Space Semantics is typically thought of as a resemblance theory, but that it can be.

This thesis's conclusion is that, contrary to Cummins' views, there is room in our best theories of cognition to include both resembling and non-resembling representation. In the first chapter, I attempt to characterise representation as being at least consistent with 'non-resemblance', and in the latter chapters, I attempt to show that non-resemblance might actually be a useful property for some representations. Against resemblance-only views of representation, this thesis also raises the possibility that representational content works in both ways I alluded to above. That is, that content can inhere in the form of a representation and, in this way, be realised as a 'physical player' in cognition; and that it can also emerge out of cognition in representations that do not 'use' their content to 'push' things around, but gain their content from their acts of 'pushing around' and being 'pushed around'.

I associate this second possibility with Dennett's functionalist and interpretivist view of representation and content. I also argue that the first possibility, of content being manifest in form, while most intuitively associated with resemblance, also makes some sense in the classicist position where content can be reflected in both the coarse structure of concatenation, as I mentioned, but also in the identifying form of simple representations. I attempt to characterise this second kind of classical *content in form* as an 'implicit' encoding of content in the (identifying) form of simple representations, one that is not resembling in any way, but one that is still the potential basis of an intrinsicist view of content.

Thus, this thesis takes the logical geography of the debate over form/content relationship in mental representation to have resemblance theorists and classicists, potentially at least, allied as intrinsicists and in opposition to who Cummins' terms 'use' theorists of representation, such as Dennett. This is certainly not Cummins' view of the matter, according to which classicists are also 'use' theorists. In fact, Cummins has attempted to shift disputes between theorists such as Fodor and Dennett over the interpretative and holistic nature of content to one side and has contrasted both their notions of representational content with his own so called 'non-use' view. This thesis attempts to question that realignment.

Some preliminary distinctions:

1. I think it fair to say that the form and content distinction comes in two versions in cognitive science. In the field of psycho-semantics, a representation can be said to have physical and semantic properties. But, also, in what I call psycho-syntax, a representation, may have certain physical properties and a syntactic content, so to speak, namely, a formal type. To illustrate using a non-mental representation: the word 'tree' has certain semantic features such as referring to green woody things- a property it shares with the German and Indonesian words for tree. It also has the property of being a token of the word type 'tree'- a property it shares with 'tree' and '**TREE**'. So, in the same way, a mental counterpart to 'tree' might have a syntactic and a semantic type, or identity, that are both distinct from its 'basic' physical properties. That is, we can identify such a representation either by the property it relates to in the world or by the form it enjoys as a particular symbol type.

2. I also think that it is fair to say that representation is something that involves representations but, also, goes beyond them. Although this seems obvious, it is worth pointing out because it can be confusing as to whether the content of a representation is a property of the representation, itself, or the property of the relationship between the representation, its referent and a user of the representation. Since the act of representation requires a least a dyadic relationship between representation and referent, and possibly a triadic relationship between those two and a user (as Von Eckardt (1993), following Pierce, argues), representational content is thus dependent on some sort of relational fact for its existence. [3] On the other hand, it doesn't follow that representational content is a property of that entire relationship. That is, we

---

[3] We can note the strangeness of this relation, for example, the fact that the represented object may not exist, or that the relation between the parts seems to determine the nature of the parts, that is, it makes one of them a representation (Dennett and Haugeland, 'Intentionality' the Oxford companion to Mind, 1987).

can intelligibly claim that a representation's content is a property of the representing vehicle, even though it only enjoys that content in virtue of it's partaking in some external relationship or relationships. Resemblance theorists, in fact, rely on this possibility in that they take content to inhere in the actual form of a representation. 'Use' theorists, in contrast, must take any such resembling representations to possess content in virtue of the use any resembling structure is put to by the larger system as it deals with any usefully resembled features of the world. In effect then, 'use' theorists must take content to supervene on the representation-in-the-context-of-its-use. That is, more than the representation itself is required to determine the content of the representation[4].

3. Following on from this, the semantic identity of a representation can depend on the identity of the representational relationship it figures in and, at the same time, belong to the representation alone.

In general, this thesis wrestles with this object/relation dualism in representation- that representation requires both a vehicle and certain relationships existing between the vehicle, other representing vehicles, the larger cognitive system, and the world. At times, both syntactic and semantic properties of the representing vehicle seem to sit precariously as intrinsic properties. Situated intelligence theories, for example, by focusing on the coupling of brain and world push us towards giving up on isolating particular loci in the head as (asymmetrically) contentful. But less dramatically, as I will point out in more detail below, often times the syntactic and semantic structure of representations are portrayed as relational properties between representations.

The chapter-by-chapter structure of this thesis is as follows. Chapter 1 deals with what it is for something to be a mental representation and what types of things might qualify. This chapter is intended to sure up the notion of representation and outline some of the possibilities for its instantiation. The conclusions reached here bear on

---

[4] I'm not sure then, that a use theorist can even hold content or sematic type to be a property of a

subsequent disputes about the qualities a representation may or may not enjoy, for they may provide some separate support for the 'representation-hood' of objects not possessing certain properties, such as a structural basis for resemblance, for example.

Chapter 2 considers the notion of a syntactic property and makes the distinction between 'intrinsic' and 'relational' syntax. It shows how syntactic structure in language can be represented using both these types of syntax and concludes that cognitive systems can gain an advantage from using intrinsic syntax in typing representations by their form- both in isolation and in combination. The pertinent point is that they are able to represent important types of structure with structure, much as Cummins argues must be the case for all representation. My point is that, here, this applies to a classical account of the representation of syntax in language.

Chapter 3 deals with the relationship between physical and semantic structure and consider Cummins' claim that content must inhere in structure. I introduce some more concrete notions of resembling structure and sketch the landscape of possible replies to Cummins.

Chapter 4 expands on these replies and attempt to show that Cummins' treatment of symbolic theories of representation as 'use' theories may be in error. I also attempt to make a reserve 'safe haven' for non-resembling representation by making a case for the explanatory power of a clear 'use' theory of representation such as Dennett's. In doing so, I introduce the notion of a multiple content types existing within the one system. I also suggest that the 'use'/'non-use' distinction may not be as clear cut as Cummins thinks.

Chapters 5, 6 and 7 describe several related roles in cognition open to non-resembling representations. I argue that the roles of representing abstract content, combining with other representations, and tracking individuals or particulars- including, tracking other representations- are better filled by non-resembling representations. I argue that, in particular, one 'resemblance friendly' account of representation, namely, a 'second order isomorphism' treatment of connectionist representation, fails to meet the

representation itself.

representational requirements for playing these roles, including, ironically, Cummins'
own requirements of representational content being a causal player in cognition.

In chapter 8, I outline a recent account of how Cummins' account of *representation as
picture* might be approached in the connectionist paradigm. The main point of this
chapter is to contrast the merits of synchronic as opposed to diachronic
representational structure in the context of Cummins' account of representation and
resemblance. I also apply the distinctions made in previous chapters to the notion of
weight state representation, in general.

Chapter 1. What counts as a representation?

The high and low roads to representation

A mental representation, we might say, has properties that are more and less tangible. Content is less tangible than form; on the face of it, content is harder to understand and work into explanations of cognition. Fodor (1987) makes the point that cognitive psychology is more concerned with mental causation than the fact that brain states could have satisfaction conditions and 'the place of content in the natural order'. And perhaps this approach fits well with Dennett's general resistance to trying to find a precise content for representations and his belief that the attempt just serves to shift attention away from what matters, namely, what the representation does and how it does it (Dennett, 1987).

As well, some types of content, or intentional properties, may be easier for us to get a handle on than others. To put things metaphorically, perhaps we can take either a high or low road to representation. The high road comes at representation via the 'stronger' semantic properties such as truth-value, precise content, combination in propositional type representation, and, sense and reference. It also travels past a 'multiple use' component to representation[5]. The lower road comes at representation via the more basic, one-dimensional intentional functions such as indication, pointing, tagging, and labelling.

This choice can shape our ideas of what counts as a mental representation. Taking the high road produces a higher bar for qualification as a representation. Von Eckardt (2003), for example, sets the bar high when she sets out five features of representations. Representations are semantically: a) selective (definite in their referent) b) diverse (have a range of contents) c) complex (have a sense and a reference), d) are evaluable (true or false) and e) compositional. Von Eckhart considers representations without these properties merely formal objects, in fact,

---

[5] Multiple use and combinability seem to go hand in hand, since multiple combinatorial possibilities amount to multiple possible uses. Perhaps, though, representations can be consumed by different consumer systems in the absence of any further representational combination, and these 'consumptions' count as multiple uses, also.

barely representations at all. I contend she might be setting the bar too high and overlooking the importance of formal and low road properties in representation.

The highest point along the high road is probably Dennett's (2000) notion of 'florid' representing. This, as Clark (Clapin (ed) forthcoming) puts it, is not just the representation of other representations but the *witting* or knowing use of representations *qua* representations. In general, the high road seems to take us closer to conscious thought and ultimately to *conscious* conscious thought. In Dennett's case, it also takes us out on a crucial diversion into culture and language. I will argue, throughout this thesis, along classicist lines, that some features of public language seem to be present in mental representation, as well, namely a certain syntactic and semantic atomism.

My personal preference is not for philosophy that begins or ends with definitions; there is usually a case to be made against clear boundary lines in most things. So, I will try to describe a continuum of representation types. A tempting low road option is to be permissive and to follow Markman and Dietrich (1998, 2003) in treating any state that mediates between sensory information and goal directed behaviour as representational. In virtue of being connected to sensory input and goal-related behavioural output, a state counts as a representation.

> The basic idea is that mediating states are internal states of a *system* that carry *information* that is used by the system in the furtherance of its *goals* (Not all system states are information states: some are goal states) (1988, p140).

This, in effect, is merely to deny reflex driven cognition or behaviourism: *cognitive* systems contain *informational* mediating states as opposed to states that merely mediate stimulus-response pairings. 'Systems' and 'informational states' go together: systems, by Markman and Deitrichs' definition, use feedback loops to control their environment in some way, and feedback loops require internal states which give some indication of how the environment is. Being such an internal state is a low bar for a representation to pass over, but the characterisation captures the basic components of the notion of representation, namely, that representations are related to both sensory

input and non-random behaviour- or, that they are informational and used to some end.

What makes a system's movements gaol related behaviour, Markman and Dietrich don't explicitly say; they do point out that goals can be implicit. I suppose they are entitled to the claim that living organisms have some kind of 'telos' about them – related to either staying alive or protecting their genes within. Generally, an element of 'being for something' or 'being part of some problem solving device' seems essential to representation. This might be taken to mean that representation requires an interpretant, as Von Eckardt (1993) puts it. O'Brien and Opie (forthcoming, p 4) summarise her notion of interpretation as something that 'must modify a subject's behavioural dispositions towards the [representational] vehicle's represented object.' Alternatively, Millikan writes of 'consumer systems'. We might put it that some kind of design process, like evolution, has to have 'set up' a system of consumer systems or interpreters, or just 'uses', of representing objects.

What makes a state informational, according to Markman and Dietrich, is that it's presence in the brain allows us to infer that something about the world is either definitely true, or likely to be true. Markman and Dietrich use this two-tiered account of information in order to explain the possibility of misinformation. Low-level feature detectors are by definition successful, while higher level feature detectors, which depend on lower level feature detectors, are not necessarily always successful detectors (1988, p143-145).

Here is an example of the sort of thing Markman and Dietrich have in mind as lower level detectors. Suppose it is true (New Scientist, 2003) that some visual system neurons selectively respond to direction of movement. Furthermore, that it is also true that some illusory perception of movement is due, in part, to the exhaustion of such neurons, such that neurons with the opposite preferred direction contribute 'too much' to upstream cognition. Whatever the semantic properties of this neural activity (whether or not they are semantically selective, complex, evaluable and compositional), these neurons seem representational – and, on the sorts of grounds that Markman and Dietrich are suggesting. Such neural activity helps you to stay dry

because its informational content is used in the production of water directed behaviour.

Cummins (1996) would disagree that such activity is representational,

> Representation cannot be indication, as everyone knows, because there can be misrepresentation but not misindication: if *e* carries the information that there is a visual edge present, it follows that there is a visual edge present. But if *r* represents the presence of a visual edge, it does not follow that that there is a visual edge present (1996, p65).

Perhaps Markman and Dietrichs' notion of information enables them to make sense of 'misindication'- since 'inform' isn't necessarily a success verb, on their account. I deal with Cummins' treatment of indication in chapter 4. For now, I point out that requiring of a representation that it be possible for it to be used incorrectly seems a bit hard on states like our 'direction of movement' detectors and their representational aspirations. That these states are used in the detection/perception of moving objects seems to me to meet the teleological and informational requirements of representation.

Cummins would argue that if such neurons misfire, in doing so, they do not mis-indicate; a misfiring, of some kind, is not a kind of indication at all. Cummins' own solution to the problem of error is that misrepresentation is the misapplication of representational content- something like the 'mistargetting' of a representation at the world. Hence, in fact, it is *misapplications* that are false- and not the representation itself. Representations cannot be false, on his account, only misapplied.

Accordingly, I find it hard to understand why the fact that an indicating state cannot be false is enough to rule it out as a representation- when being false isn't possible of representations anyway, on Cummins' account. If 'indicate' is a success verb, why can't we just use another name, like 'suggestor'? Even the earliest indicators/suggestors will only be correct a percentage of the time. If this isn't carrying 'information', then so be it- but getting things right most of the time still seems 'indicational', in another sense of the term. In any case, there are other

candidates for making sense of 'mis-indication' as misrepresentation- namely, 'asymmetric dependence' accounts of misfirings and 'improper role' accounts of misfirings. Cummins rules these out in virtue of being part of use theories of representation, which he rules out as theories of representation, in general. I try to meet his argument in chapter 4.

Actually, Cummins (in Clapin (ed) forthcoming, p123) puts it that states like indicating states are *intentional* without being *representational*. His approach to representation seems 'high road', then. That something can have intentional properties and not be representational suggests that perhaps there must be a descriptive component to representation. That a representation must say more about something than that it is merely present. And, it must say it using its own resources; as Smith (ibid) points out, representation for Cummins is about the sort of content that cannot be cashed out in any non-intrinsic properties of a representation. A representation possesses content by itself. Smith is sympathetic,

> Intentionality and content are achievements, in my view, that allow a system to be existentially oriented towards that which they are not causally engaged (p.108).

Maybe so, but is this true of all representations- that they can represent in the absence of a referent- or even that they should have the function of doing so? Bechtel (1998), in contrast to Cummins, takes a similarly generous position to Markman and Dietrich. He depicts representation as a coordinating device - as something that, in virtue of standing in for something, enables the brain to itself coordinate an organism and its environment. The 'standing in' relation requires some connection between the 'stand in' and the world, but also, and more importantly, that the 'stand in' have the function of standing in- that its use be part of a larger system. Bechtal is persuaded of this by Millikan's consideration that something might be a stand in for something *in waiting,* as it were- that is, before the thing it stands in for has ever arisen. This leaves its status as a 'stand in' resting on its *role* as one.

So, like Markman and Dietrich, Bechtal sees representation as requiring a (potential) object, a vehicle, and a larger system. Bechtal's position is similarly *low roadish* since 'goal directed coordination', like 'goal directed mediation', doesn't carry much

semantic baggage. For example, Bechtal considers the famous Watt Governor as a representational system. In particular, the spindle arm represents in virtue of the fact that it is used by the governor to stand in for the flywheel.

> Watt devised the whole device so that the steam valve could use the information encoded in the angle of the spindle arms as an indicator of the speed of the flywheel (1998, p303).

Similarly, we might claim that Mother Nature devised the whole visual/motor cortex so that it could use the information encoded in various indicator states of the visual system. Presumably, Bechtel is not inclined to worry about the fact that the spindle arms were not designed to represent in the absence of a flywheel. One point Bechtel does want to make about representation is that it can be found in such tightly closed feedback loops as in the Watt governor. In this way, Bechtal hopes to cut the anti representationalists off at the pass by setting the bar low for representation. We noted, above, that Markman and Dietrich also require a feedback loop of some kind between the world and a representation-using system, albeit one not necessarily so tightly closed as the Watt Governor.

A possible fly in the ointment, here, for our low road notion of representation, might be that information extraction requires more than the kind of relationship we see between the Watt Governor's fly wheel and spindle arm. This is just a straightforward physical relationship such that the spindle arms rise and fall, according to centrifugal force, as the flywheel increases and decreases in speed. Consider this version of the role of representation in cognition, put forward by Wheeler,

> Adaptive richness and flexibility of intelligent behaviour requires that behaviour be sensitive to the information carried by environmental stimuli and not simply to the *physical form of the stimuli* [my emphasis]; ... such informational sensitivity is impossible without representation... (2001 p 211).

Here, representation is supposedly informational in the sense that it extracts information out of the form of sensory input using a process that is more than 'brute

causal'. Representation is something that makes behaviour more than just brute response to the causal powers of input. We might think of brute causal processes as something like 'one step' processes. On this account, it seems that the Watt governor does not contain representations given that the spindle arm seems to be merely responding to the form of the flywheel. This view of representation seems to require more encoding to be involved in the relationship between stimulus and representation- or, at least the storage of either the mediating state or something to match against the mediating state. If this is required of representation in general, then the bar has been raised a little higher. And, we can note that the bar has been raised in much the same direction as Cummins seems to want it raised, namely to require a 'disconnect' of some kind between a representation and the physical presence of its referent.

However, perhaps a case can be made for information extraction in the operation of the Watt Governor. The system is sensitive to the speed of the flywheel, which is the information it requires, in virtue of a component (the spindle arm) being sensitive to the physical form of the wheel. In this case, the informational content of the spindle arms and the form of the stimulus, the speed of the flywheel, are tightly connected. There are no causal intermediaries between the two. However, although there is no causal 'disconnect' between the referent and the representation, they are not the same thing, so there is at least an ontological disconnect of some kind. We might say the angle of the spindle arms is very directly extracted information about the speed of the flywheel.

We might have on our hands, here, another fuzzy divide in nature. Things might just get blurry when we try to show where reflex ends and representation begins. Take what Sterelny (forthcoming) terms *detector systems.* Detector systems link a sensory state, broadly conceived, with a behaviour. They are 'single-cued discriminatory systems'. For example, corn counts as a detector system when it releases, on being eaten, something harmful to its predators. Insects that detect conspecifics by chemical signal alone also count as detector systems. It is obviously contentious whether this kind of 'tight' cue detection involves information capture – and, hence, also if it requires representation. Moreover, this may be the case, even though a detector

system may have to work via *intermediaries* between the cue or proximal stimulus and the changes in the organism it causes.

What seems to bother Sterelny about such proto-representations, such as certain chemical intermediaries between stimulus and response, is that a detection system is a 'one cue-one response' system. A detection system only requires an invariant signal or stimulus to trigger a fitness enhancing change in the organism. (In the case of ants detecting conspecifics by chemical signal, for example, the 'change' is suppressed aggression.) Even if the triggering comes in degrees, the effects on the detector system are still one dimensional, eg, as in the Watt governor with its limited dimensions of response to changes in flywheel speed. Even if the tracking is more robust than via this kind of detection, that is, if the tracking is via the detection of multiple cues, we may still have the problem of there being too tight a connection between the cue detection and a single use of that detection (ibid, Chapter 3).

On the face of it, demanding multiple uses of a representation is setting the bar higher for representation than a Watt Governor's spindle arm might reach. 'Multiple uses' seem to imply a 'consideration' or selection process that determines the use of a 'detection' – and that the representation be capable of 'hanging around' for that to happen. Sterelny writes that

> Intentional systems, if there are any, guide their action at least in part by decoupled representation: registrations of the environment that are relevant to many possible actions but functionally specific to none (chapter 3).

Yet, in defence of low road representation, perhaps this potential multiplicity of use just amounts to, either, other detecting states sharing causal influence on behaviour; or, maybe, the detecting state being responded to in different ways depending on subtle differences in its form. Either way, we just seem to have a more complex *mediating* story- perhaps one involving the combination of multiple mediating states, or context sensitive consumer systems. However, the basic feature of purposeful mediation, we might argue, remains the basis of any representation. The opposing view, which is not entirely unattractive, is that a necessary feature of representation is a 'sophisticated' control system that intervenes between informational mediation and

behaviour. This would rule out the Watt Governor and cue detection systems as representational. It would also sit with Cummins' demand for the possibility of misapplication for a representation.

That Sterelny sets the bar reasonably high for representation is suggested by his attitude towards generalisation. He accepts that any one-dimensional response to cue detection may involve generalisation of a kind. For example, for a chicken, retinal patterns caused by hawks vary with angle of view and with the hawk. Presumably the chicken visual/motor cue detection system generalises in some way across these differences. 'So even a cue-bound organism must and will generalise, treating physically different stimuli as functionally equivalent' (chapter 2). Yet, if the generalisation is not multiply used, it won't be representational. In contrast, once we get to generalisation, most cognitive psychologists are prepared to admit representation. Generalisation suggests classification and classification suggests a coding of some sorts. Different instances of a class are understood to be coded as being similar, if not belonging to that class.

To make the case for the existence of (a plethora of) mental representations, Pinker, for example, uses the case of recognition of a word in different scripts (or pronounced in different accents). Take the word *elk* and your ability to read sentences containing that word written in different fonts.

> In reality, your knowledge [about elks] must have been connected to a node, a number, an address in memory, or an entry in a mental dictionary representing the abstract word *elk*, and that entry must be neutral with respect to how it is printed or pronounced. …This is how we know that your mind contains mental representations specific to abstract entries for words, not just for the shapes of the words when they are printed (1997 p86).

Whatever it is that different tokens of the word 'elk' activate in our heads, and assuming it is representational, that thing seems to be involved in our reading in a way that is more than brute causally connected to the tokens of 'elk' themselves and our behaviour. It seems part of a sophisticated process- our extraction of the information in a sentence containing the word 'elk' is not as brute causal as the extraction of

information about flywheel speed in the Watt Governor. So perhaps another way of putting the case against Watt Governor representations is that there is no generalisation involved. But again, in defence of Bechtel's notion of representation, we can ask if generalisation is a necessary feature of representation using systems. Isn't purposeful information extraction possible without it? Also, to return to Sterelny's downplaying of generalisation, given the validity of low road representation, can Sterelny be sure that generalisation is not sufficient for representation - that whatever mediates generalisation isn't automatically representational? What other purpose can generalisation have, apart from purposeful information extraction? Hence, the mediation of generalisation should be representational activity.

Following Markman and Deitrich, why shouldn't we think of representations that mediate generalisation partly as indicator states, or as the results of a detector system at work? Although their indirect causal linkages with stimuli make them less than perfect indicators, they do seem to be in the business of feature detection. Pinker refers to some neat evidence from the lab for levels of such detections lying behind recognition. In a famous experiment, subjects were faced with a 'same or different?' task for two letters. Choices involving one upper and one lower case letter took one tenth of a second longer than 'same case' decisions. By contrast, if there was a break of a second between the presentations of the letters, then case made no difference. Presumably, in that second, a visual representation of the first letter is converted into an alphabetic one. Surely this is just an 'alphabetic' detector system having time to work. Hence, we might argue generalisation, purposeful mediation and co-ordination, and representation are closely bound up with one another.

To anticipate subsequent chapters, it might be possible to reason from representation's role in generalisation to the format of representation. Markman and Deitrich argue that abstract representations must be discrete (see chapter 2). Given that we seem to possess abstract concepts, they feel we must possess abstract representations. However,

continuous representations are not abstract ... . They are typically highly specific, because they correspond to some specific continuously varying input. In contrast discrete representations are often abstract. In fact, creating abstractions produces discrete representations (2003, p 112).

But, should recognition (of abstract properties) assure us of the presence of representation, in the first place, and in particular, of discrete, explicit representation? We can imagine a neo-behaviourist like Dennett asking, 'Why should every achievement in cognition be explicitly represented for another part of the mind to 'read'?' In other words, why can't an ability to generalise not be implicit in our wiring or functional organisation? Similar inputs might, somehow, be channelled into similar outputs without any intervening coding taking place. This would leave only an *implicit* representation of the unity of different tokens of the same abstract type.

However, I think Dennett's point is more likely to be, here, that coding for coding's sake is a problem, rather than the existence of any explicit coding whatsoever. In particular, it is a mistake to think that recoding or re-representation is how and where 'understanding' happens; the myth of double transduction is a powerfully distorting idea, according to Dennett ('Things about things', internet version, p3). Dennett's (1978b, 1998) *giant robot* thought experiment, I think, tries to show the pointlessness of recoding for recoding's sake. We are asked to imagine ourselves inside a giant robot and faced with the task of guiding the robot by 'connecting' sensory information, in the form of flashing lights on the consol, with the right buttons for motor commands. Our intuition is that we need to translate each input light and output button into a language we understand. However, Dennett argues 'this can hardly be the brains solution', on the grounds (I think) of a homunculus/ regress objection. In any case,

[t]he job of getting the input information *interpreted* correctly is …not a matter of getting the information translated or transcribed into a particular internal code unless getting the information into that code is ipso facto getting it into functional position to govern the behavioural repertoire of the whole organism (1978, p 305).

So, it seems, by Dennett's lights anyway, there is no reason, here, to be anti (explicit) representation, in general. Representation can be part of interpretation, but we should remember that interpretation is not to be understood as something achieved in representation itself. Similarly, 'understanding' is not the capturing of the content of representation in re-representation. We can note, here, signs of the centrality of 'use' in Dennett's account of representation.

However, the moral for us seems to be that recognition of types of things is probably still good reason to posit representation; there is nothing here to make us shy away from representations with abstract content. Such representations are perfectly compatible with functionality; we do though, when positing representations, need to have some idea how they might be functional. Perhaps Dennett is right about the limits of the notion of (explicit) representation in explaining the experience of understanding something, or explaining intelligence. However, this doesn't mean that it is a mistake to treat, for example, brain states that mediate judgements of letter type as detectors or 'representers' of abstract properties. The question for this thesis is, is this representation likely to be resemblance representation?

I am going to assume, then, that the more sophisticated the recognition, i.e., the more that groupings of stimuli into types rest on sophisticated similarities, the more representation, either of the nature of those similarities or of the different groupings themselves, is likely to be in play. This leaves me in opposition to the Dynamical Systems Theory approach to cognition, which does not view coordination with the environment as being reliant on information bearing intermediaries, or intentional states in general. I outline now, very briefly, where representation, as I have characterised it, can be best defended against this view of cognition.

The DST school and the *situated* or *embedded* approach to representation (see Clark 1997 for a survey) wish to emphasise fluid, circular, and self-organising interaction, or coupling, between brain and environment (including the body). This has been motivated by, and motivates, consideration of the role of extra-brain factors in shaping behaviour; factors which force the brain into a corner and, hence, coordinate behaviour as much as being the 'object' of coordination. Gibson's ecological optics, with its emphasis on proximal stimuli as opposed to internal representation of distal

stimuli, is a related idea. We have, here, the notion of 'causal spread' for flexible behaviour, where the causes of flexible behaviour are spread outside the brain, and work at the same time in tight causal interaction with the brain in bringing about intelligence. In considering the significance of causal spread, Clark and Wheeler (1999) ask, 'Can we isolate some factors as informational? And, do these drive behaviour more than other non-informational factors?

Perhaps if organisms were just complex collections of detector systems, then behaviour *would* in large part depend on what happened on the outside surface of the animal, i.e., the animal would be in large part stimulus driven. This would be true even if the integration of detection systems required some subsumption architecture and, perhaps, some unifying stand-ins, as in the hawk example, above.[6] However, it would still be the case that, taking a low road approach, we should be able to defend, within or between these detector systems, informational states that intervene between sensory input and behaviour as necessary representational components of intelligent behaviour.

This leaves the presence of high road representation to defend. Representational accounts of mind often posit a de-coupling, de-reflexing and stimulus-escaping process that delivers more autonomous coordination with the world. This de-reflexing is thought to be what gives an organism *real* flexibility. As we noted, Sterelny uses the term 'decoupled' representations for informational or registration states about the world which are decoupled from any particular response or use- that is, for states of the brain that are available for various uses, depending on the informational and motivational context of their instantiation. Accordingly, flexible use of intentional resources lies behind flexible behaviour- and, internal flexibility is the cause of external flexibility, to this way of thinking.

The pinnacle of such internal intentional flexibility is the combinatorial possibility of a language of thought; our thoughts know no bounds in virtue of the limitless 'combinability' their components enjoy. In general, at the top end of the decoupled market, human planning and imagination (offline cognition) seem to, more or less,

obviously require mental representations, given their extra sensory nature. They are transparently unreflex-like merely in virtue of being about non-present states of affairs- something we noted Cummins thinks is central to representation.

Retreating from folk psychology for the moment, this de- reflexing story is the mainstay of cognitive psychology. For instance, in its search model of problem solving and learning, 'search' is another word for de-reflexing. In a system like Newell's SOAR (Newell, 1990), for example, solutions to new problems are composed using searching techniques reliant on discrete and structured representations of new goal states and old solutions. (Goal states (Sterelny calls them 'target representations') are broken down into representations of sub goal states; old solutions to old problems, themselves representations, are mixed and matched and retried so that alternative decompositions of a problem can be tested.)

Of course, merely mentioning *goal states*, for example, proves nothing. Returning to the fringes of representation, it seems that in some cases of reflex-like cognition or detection no explicit goal is ever represented, even if it is possible to describe a goal state from outside the system. For example, in a phonotaxis case discussed by Clark and Wheeler, the female cricket seems driven by the very male mating call itself, given her physical and neural organisation. There is no need for a representation meaning 'locate mate' or 'move towards mate'. Such things just fall out from the nature of the male's call and the way the female's eardrums are affected by the frequency of the call and the way they are connected to muscles via interneurons (1999, p106).

The cricket phonotaxis case is definitely a 'one-cue → one-response' story, involving a one-dimensional use of interneurons- they always activate leg muscles. Even so, low road representation still seems at play just in virtue of something (purposefully) mediating between stimulus and response, and, perhaps, standing in, in the cricket's design for the male cricket's presence. As for high road representation, the gap between cricket phonotaxis and deciding whether to cheat or not on a tax return seems big enough for such representation to slip through.

---

[6]We see such affordance driven behaviour in psychiatric conditions such as utilization behaviour or

But, perhaps, the 'preparing the tax return' defence of representation is a bit beside the point. Clark and Wheeler (1999 p104) claim that most cognition, most biological intelligence, is not the offline (planning or imagination) type, anyway, but the kind that produces ongoing flexible response to 'a constant stream of incoming stimuli'. Moreover, they accept that the range and effects of non-representational factors that are involved in this type of intelligence, the seriousness of the 'causal spread', is such that some of the flexibility in behaviour doesn't need to be explained by instruction or code.

> sometimes, the kind of work done by additional factors and forces [additional to representations] turns out to be precisely the kind of work previously assigned to a more isolated, typically inner system (1999, p130).

> ...the other forces and factors ... reveal themselves as the unexpected root of *a good deal of* [my emphasis] the kind of flexibility and subtlety normally associated with representation based control (1999, p131).

I think the 'representation for flexibility' story still stands intact. Phrases like 'a good deal of' are too imprecise to do much damage. Clark and Wheelers' own examples don't fill the representationalist with fear. Maybe the sexing of alligator eggs depends on temperature in the nest as much as putatively representational genetic instruction in the eggs, but this isn't quite the flexibility we had in mind. Nor is the tracking of con-specifics in the insect world by using one's own body design to 'process' communication much more flexible than a plant following the sun. A plant, too, uses its body design to track the signal. The same lack of real flexibility applies for developmental processes such as infant walking- another favourite example of complex behaviour just 'falling out'. These are not the sorts of problems, that animals face, requiring rapid information and stored knowledge integration.

That is, despite what Clark and Wheeler say about the majority of animal online cognition, we might argue that nature just isn't that generous as to let non-

---

alien hand

representational solutions succeed. Sterelny makes the point that quasi representationalist but mindless cue detection works best in benign environments where the cue is constant and comes from a disinterested party, eg, the sun. However, when signals or cues come from prey or rivals or predators or even just from changeable environments then fitness enhancing behaviour may well not just fall out from cued responses, since the cues may be misleading. In order to detect misleading cues some kind of analysis is required of the cue and its context. Cue detection needs to turn into cue comparison, at the least.

> Adaptive behaviour targeted on the inanimate world (and biologically indifferent parts of the animate world) can often be controlled by simple cues of environmental structure. But much behaviour takes part in the hostile world of predation and competition. ...Predation results in epistemic pollution. Prey, too, pollute the epistemic environment of their predators. Hiding, camouflage and mimicry all complicate an animal's epistemic problems (Sterelny, Forthcoming, p 28).

Perhaps this means that whenever environments get tricky, which is pretty often, stronger forms of representation are preferable to weaker. Were, for example, cricket phonotaxis to be open to false calls from a predator, then presumably higher road representation would be a useful tool for a cricket faced with these epistemic problems- in particular, informational states that were available for integration with other informational and representational states.

Actually, as representationalists themselves, Clark and Wheeler make a case for picking out some casual factors as (strongly) representational amongst other contributing factors to behaviour. To set representations apart from other brain, body and world states, they set the bar reasonably high. According to Clark and Wheeler, representations are:

1. arbitrary, in that 'what matters is … their role as information bearers' and not their form.
2. consumed for their content by consumer systems, where the information they are putatively carrying matches the type of consumer system they are consumed by, and,

3. capable of combination to some extent that is appropriate to the target domain (1991, p124).

In contrast, Markman and Dietrich would presumably only require criterion 2- where a consumer system may only be something non-specific like the motor cortex, or even the body itself (as in the case of the cricket). Purposeful informational mediation does not seem to commit us to either arbitrariness or combination/compositionality. The requirement of arbitrariness seems linked to the following two ideas:

a) that it is how a representation is different from other representations that really makes it informational. For example, Jackendoff 's (1992) definition of mental information is as follows,

By a form of mental information, I mean an organised combinatorial space of distinctions available to the brain. ... it is not the symbols we use that are significant but the distinctions possible in the system of symbols... these are claimed to be homologous to the organisation of the relevant subsystems of brain states.

And b) a representation has to mean something to something, it has to be decoded somehow – so its effects are more important than its nature, even though its nature determines in part the effect. It's not that a representation has to *communicate* something to a consumer system (Cummins' communication fallacy) but as Clark and Wheeler, put it, a representation has to secure effects without being understood (1999, p125). From this, Clark and Wheeler conclude that representational form must be arbitrary.

Wheeler and Clark summarise matters for representation as follows:

The three features [of a representation]… fit together like this, arbitrariness (the fact that any suitable information bearer could play the same adaptive role) is what forces us to depict certain states or processes in contentful terms; content based consumption emerges as a kind of 'inner decoding' in which some specific information-bearer is enabled to guide behaviour; and systematicity [i.e.,

combination] is the economical icing on the informational cake, allowing structurally related inner states to guide different but related behaviours. (1999, p128)

On these matters,

1. I don't see why real representations can't possibly have a non-arbitrary form. Differences between representations of non-arbitrary form will, presumably, be just as useful as differences between representations of arbitrary form. But more importantly, it may not be the differences between representations that count so much as a representation's intrinsic properties- as in a strong resemblance theory of content. In this sort of case, of course, consumption may be required to make use of a representation, but it needn't determine in what way a representation is contentful. 'Content based consumption' needn't be 'content-making consumption'. In fact, matters may be the other way around: content may shape its use. In this case, no decoding is required to allow us to treat the representation as contentful. In general, I argue that responding to the physical form of a representation can be responding to its informational content, and not creating it. As I noted in the introduction, just because use is required of a representation for it to be a representation does not mean that its use constitutes it content.

2. Why should representations necessarily be combinable? Interpretation may require combination of some sort- if only between a representation and a representation of a context. But, could this be handled by implicit representation? Certainly the angle of a spindle arm doesn't seem to combine with other representations. Couldn't the brain contain such one-task modules with minimal low road representational machinery? If representations *have* to be multiply useable, couldn't they be multiply consumed in the absence of combination with other representations?

I make these points in anticipation of a pluralistic approach to representation that I construct in later chapters.

Tacit and implicit representation

Cognitive science has had to make some sense of the different ways knowledge can be said to belong to a cognitive system. Stored knowledge might be intuitively thought of as explicitly represented knowledge- as in an encyclopaedia or instruction manual. But, as philosophers like Ryle have pointed out, such knowledge *qua* representation isn't going to account for much intelligence by itself if it is not put to use in an intelligent way. This, of course, applies to 'written down' knowledge about how to use other written down knowledge. How are we, then, to account for the intelligent use of this kind of explicit representation? The obvious thought is that this kind of 'how to use' knowledge must be stored in a different form- such that when it comes into contact with the 'written down' variety, the right things automatically happen to the written down representations. Such know-how would be in the system's 'wiring' or organisation, so to speak. A little confusingly, both the terms 'tacit' and 'implicit' have been used to refer to this kind of knowledge–sometimes *qua* representation, as in 'tacit representation'. I now consider some versions of these notions, in the literature, for the purpose of shedding some light on the question of what counts as representation.

Millikan (in Dahlbom (ed), 1993) distinguishes between tacit suppositions and intentional icons- we might describe the latter as a form of high road representation. As for the former, Millikan points out that an organism's design can embody assumptions about the world without explicitly representing them, eg, that its surroundings are similar in colour to its skin, or that, in the visual world, corners and edges signal objects. Her camouflage example is consistent with Dennett's idea that '[e]volution embodies information in every part of every organism' (1998, 'Things about things'). Dennett, for example, refers to information about tree branches being in the structure of a primate's hand. This is consistent with the idea of 'causal spread'- bodily facts can be informational as well as brain states, but in different way.

As Dennett puts it, this sort of information 'doesn't have to be 'represented' in 'data structures' in the nervous system' (ibid, p 5). Millikan's 'corners *means* object' example, where the wiring of the visual system is such that representations of edges and corners lead to representations of objects, is obviously a case of representation in

the nervous system, but it is also tacit; it falls under Dennett's 'Intentional Stance' notion of tacitly represented information, that is, information 'hardwired' into a system.

Do these notions of representation shed much light on low road representations? It does seem true that in wiring cases such as the 'corners means objects' example, the wiring facts intervene or mediate between sensory input and behaviour. Perhaps it could even be said that they *stand in* for the relationship in the world between corners and objects. We can note, though, that they do not covary with that relationship in the informational sense of 'covary'; still, perhaps this is because the relationship between objects and corners doesn't itself vary. Such hardwiring is not informational in the sense that it is directly related to changing sensory input.

Perhaps, for this reason, Markman and Dietrich would take a sceptical attitude to wiring facts, such as connectionist weights, as representations. It *is* true that connection weights alter with 'sensory' experience, but it is always sometime after input information has made its way through a system that connection weights are changed. If the low road/high road distinction applies to tacit representation, then states like weight states might make some sense as low road representation in virtue of being background 'informational' states, or stand ins for long term relationships in the environment.

As putative high road representations, it is unclear if wiring type facts are capable of multiple uses. A weight state, taken as a whole, might underpin many input-to-middle layer transitions. However, it is unclear if each of these underpinnings counts as a different use or as different instantiations of the one use of a representation. Certainly, when weight states are used within the one network they don't seem to be consumed in the different ways that a symbol in mentalese might be, in some theories. In fact, the weights seem to be the consumers themselves, namely, of activation state representations. On the other hand, when Clark and Thornton (1997) imagine problems being fed into various networks in many-branched searches for computational traction, I suppose each branch of that search consists of a use of that network. I consider the representational status of network states in more detail throughout the thesis, and weight states, in particular, in the last chapter, when I

consider, again, whether there might be some multiple-use story to be told of weight states as maps.

To move on to *implicit* representation, or information implicit*ly* represented: for Dennett (1986, p 216), implicit information is information 'implied logically' by *explicitly* represented information[7]. Explicit representation is the basis of inference, whereas, implicit representation cannot be combined with other information because it hasn't been made physical (and discrete?) yet. Implicit representation isn't much use to us here, although it does help to make the notion of explicit representation a little clearer. Explicit representation *is* available for inference, and perhaps combination.

Millikan ties her notion of explicit representation, what she calls 'representation proper', to inference (1993, p103). An intentional icon is acquired by an organism in response to the environment- something capable of varying in correspondence with a (possible) variance in an aspect of the environment (ibid, p 99). Moreover, *qua* representations, intentional icons are capable of delivering new information via some kind of inference. Millikan has a broad notion of combination/inference in mind. For example, what makes a mental map a representation is that it can be overlaid with other maps to produce new information.

> ...two vehicles of information [can be] combined, *using a middle term* [a shared term], so as to produce a third vehicle containing new information. Accordingly, the maps are 'representations' (ibid, p104).

So Millikan, like Dietrich and Markman, wants real representation to be closely linked to sensory input (she wants them to vary as the world varies) and also to combine with other representations. Unlike Clark and Wheelers' position, there is no requirement of arbitrary form – in fact, the opposite is suggested in that intentional icons vary as the environment varies. Although, we should note that this does not require that representations are structured to resemble the environment in any obvious

---

[7] Though, it needn't actually be 'reachable' by a system containing that explicit information for practical reasons such as time constraints.

sense; but there is, at least, the requirement that they be capable of changing in some way as the environment changes, and also capable of sharing components. This last point about shared components seems to tie 'explicit' representation to representations with identifiable constituent structure- to representations containing parts with specific content, and which are tokened in different representations.

This notion of determinable shared components in explicit representation is keenly fought over in disputes over the format of mental representation. Some connectionist research seems to suggest that constituent structure can be represented in a non-obvious manner, in compressed or superposed representations, eg, in the middle layers of auto associative networks. Clark (1992), influenced by Kirsch (1991), wonders if, under the right circumstances, such representations might even count as *explicit* representations. Clark is inclined to treat 'explicitness' as relative to the retrievability and usability of a representation- rather than it being a matter of a representation being obvious in its 'deliniability' and structure. For example, a representation in compressed or superimposed form might still be able to explicitly represent providing it's content is easily useable by a system.

Clark also treats 'explicitness' as coming in degrees- increasing with the range of uses to which a representation can be put. Accordingly, any superimposed tacit know-how in a network's weight states does not automatically become explicit just because it may be easily retrievable. Even if, for example, it is possible to ask a network questions about the information contained in its weights, eg, as in a network that answers questions like 'Do dogs have fur?' In this case, while you might say that 'the retrieval tool is built into the knowledge representation itself' (1992 p385) and so the information is easily retrievable, the knowledge is only good for retrieval, not for any other use.[8]

On the other hand, any evidence of such compressed or superposed representations being not only retrieved but also manipulated is taken, by Clark, to count towards their full status as explicit representations. Clark refers to Chalmer's (1990) claim to

have shown how the superimposed middle layer representations of a Recursive Auto Associative Memory network[9] could be manipulated by another network. The original network was able to auto associate representations of linguistic tree structures, namely, representations of sentences of English in active form. Its compressed middle layer activation state representations of those tree structures were then inputted into a standard feed forward network.

This second network, after training, took the compressed representations as input and was able to transform them into another (compressed) grammatical form, namely, their passive forms. Chalmers treated this as the utilisation of *implicit* representation (1990 p60). Clark suggests such compressed representations in the context of a transforming network are perhaps closer to being *explicit* since they are readily usable, even if not multiply usable (1992, p389). However, as such, he concedes that they do not possess the same computational dexterity as LISP atoms, for example, which are roughly as usable as the programmes that manipulate them are 'writeable'.

Consequently, I think we can conclude that Clark fails to fully motivate a redefining of 'explicit' representation as being relative to computational context. On his own admission, he is unable to tell a convincing story about how connectionist candidates for tacit representation might be open to promiscuous use. He does make a case for considering the augmenting of any system's processing power by external resources like pen and paper (ibid, p392). However, this does not really suggest to us how our most promising example of tacit knowledge, namely, weight state representation, could be made more flexible. Increases in memory capacity don't obviously turn know-how into declarative knowledge; rather it just makes that know-how more potent and opens up the possibility of further know-how to develop about manipulating those external resources.

All this suggests that explicit representation remains, at this stage, best tied to the notion of discrete single content representation. That is, it seems that for a system to explicitly represent something, a delineable part of that system must have that task

---

[8] Perhaps Clark is alerting us to link between representation and multiple use being a matter of explicitness rather than representation-hood, itself.

and no other task- at least at any one time. Rather than blur the distinction between tacit and explicit representation, it might be better to keep the two ideas apart. The simplest stance is that cognition depends on both tacit and explicit representation. Explicit representation is arguably required in a signalling/ information-bearing role and as the bearer of new premises for inference and, possibly, new conclusions. Tacit representation is required to 'make these things go', so to speak, and in a manner that reflects the way the world is. This, in turn, suggests that, prima facie, tacit representation can stand in for the world's (functional) structure. What we really need to know about these two forms of representation is how they work in more detail- both in their syntactic and semantic aspects.

I finish this chapter, here, then, having motivated a wide or liberal use of 'representation'. This includes reference to what I described as low through to high road intentional states. I will use the notion of low road representation in criticism of Cummins' account of representation. In discussing the explicit/tacit/implicit distinctions as they apply to mental representation, I have left open the representational status of 'wiring' facts- noting that they may share some features of low and high road representation. I also noted, using connectionism, the distinction between representation that is *tacit* in the sense of being a wiring fact and *tacit*, or *implicit*, in the sense of being 'disguised' or 'non-explicit', as in being superposed. Superimposed activation state representations of the type Clark considers were argued to be 'non-explicit', at least, if not clearly tacit or implicit in Dennett and Millikan's use of those terms. In the next chapter, I consider the notion of syntactic structure in more detail, and in particular, the disguised, non-explicit syntactic structure of superposed activation states. I argue that the virtues of structured representation that Cummins makes much of cannot be credited to superposed representations.

---

[9] Auto-associative networks output their input, unchanged, having taken it through a compressed middle layer.

Chapter 2. The psycho-syntactic tool kit

Intrinsic and relational syntax

In this chapter, I discuss the notion of a syntactic property, or syntax, in general. The main questions posed here are:

1. Could syntactic properties all be relational properties?

2. What is the relationship between syntactic and semantic identity in computational explanation?

If classical theories of cognition rely on systems being sensitive to form over and above other physical properties, it may be that we need to distinguish syntactic form from shape- such that it is possible for the same form to be instantiated by similar but divergent shapes. This may be just a matter of treating form as a higher-ordered property than shape. A more interesting possibility is that syntactic form isn't so much a matter of shape, but, rather, a relational property of some kind. We saw signs of this idea in Clark's notion of explicitness as relative to ease of use. I will argue that if such 'disguised' explicit representations enjoy syntactic properties, these depend on the processing environment they find themselves in for their existence. Against this possibility, I want to make a case for the centrality of the notion of syntactic properties as intrinsic properties.

It is sometimes claimed of cognitive systems that they can only be sensitive to syntax and not semantics. For example:

> If mental processes are formal, then they have access only to the formal properties of such representations of the environment as the senses provide (Fodor, 1981, p231).

Turing's big idea, after all, was that sensitivity to form could be as good as sensitivity to content, for the purposes of computation, at least. If Turing's idea is useful at all, if brains have anything important in common with Turing machines, it is important to be clear about what constitutes form or syntactic type. (We might also want to know if a *processing* device, in any case, could be sensitive to the relational properties of a representation?)

Fodor (1981, p 227) wrestles with the notion of a syntactic property,

> being syntactic is a way of not being semantic. …I see no responsible way of saying what, in general, formality amounts to. The notion of formality will thus have to remain intuitive and metaphoric, at least for present purposes: formal operations apply in terms of the, as it were, shapes of the objects in their domains.

And, in a later work,

> the syntax of a symbol is one of its higher order physical properties, we can think of the syntactic structure of a symbol as an abstract feature of its shape (1988 p 18).

So even if syntactic form, or type, is an abstract feature of an abstract physical property, namely, shape, it is still the sort of property that a processing device might be causally influenced by. As Fodor puts it,

> the syntax of a symbol might determine the causes and effects of its tokenings in much the way that the geometry of a key determines which locks it will open (1988 p19).

So, how could syntactic properties be relational properties? Consider logical form; it might be said, for instance, that a conjunction of the type 'a&b' has a logical form, namely, the form of being a conjunction. Is this something like a geometrical property that will open a lock? 'Logical form' is usually ascribed to arguments. If we think of an argument as a complex representation, it does seem as if the logical form of an

argument is one of its intrinsic properties. For example, an argument of the form 'a&b ∴b' such as, 'Trees grow and plants grow; therefore, plants grow'- seems to enjoy its logical form much the same way as a square blackboard enjoys the property of being a square.

In logic, though, formal properties seem to come in two varieties. On the one hand there is a kind of abstract shape, such as the shape of being a conjunction or an instance of modus ponens, and on the other, there are the implicational properties of being a certain symbol type. For example, what makes 'a&b' a conjunction, within a particular formal scheme, is, in one sense, its shape, but, in another, its implicational relations to 'a' and 'b'. Similarly, what makes '&' mean *conjunction operator* is its relations to other symbols. So, being a particular interpreted symbol is both a matter of intrinsic (shape) properties and extrinsic (relational) properties- both seem formal or syntactic properties, even if the relational properties might also be meaning making, as in the case of logical connectives.

We can note a similar type of relational syntax in language. Verbs, for instance, enjoy relational syntactic properties. 'Give', for example, enjoys the properties of being a transitive verb, of requiring a subject, and being able to take an indirect object. That is, in a sentence containing the word 'give', someone has to receive something, someone has to do the giving, and someone can give something to someone. Similarly, nouns also have relational syntax; for example, 'He' and 'himself' in 'He gave himself a black eye' enjoy the relational property of co-referring, whoever they refer to, in part because of their relative positions in the structure of the sentence.

Representational schemes, then, depend on both kinds of syntactic properties- both the physically identifying properties of representations and the relational properties that comprise the use of representations. Perhaps, though, it might be argued that intrinsic properties are the bedrock of representational schemes. In an automated scheme, relational properties will have to 'supervene' on intrinsic properties in some way, since relational properties have to be implemented by some set of intrinsic properties or other. Intrinsic properties, then, seem to have first claim to causal

efficacy. (As Fodor points out in explaining why there still has to be a language of thought, some relational properties may not ever be instantiated, if they are merely logical relations for example). An intention of this thesis is to explore the usefulness of focussing on syntactic properties. The starting bias is that intrinsic syntax is likely to explain more than relational syntax. I argue, then, that Clark, for example, is mistaken to think of superposed activation states as explicit representations, not just because they are not multiply useable, but because their representational structure lacks causal efficacy- their structure is matter of relational syntax only.

This is to support Fodor's position on the syntax of mental representation. Fodor states that

> the syntax of mental representations is like the syntax of natural language sentences in the following respect: both include complex symbols (branching trees) which are constructed out of what we will call classical constituents (1998b, p 92).

> For a pair of expression types E1 and E2, the first is a classical constituent of the second only if the first is tokened whenever the second is tokened (Fodor and McLaughlin, 1990).

So, clearly, on Fodor's account, syntactic properties are structural properties of the representations themselves- something distinct from their use or relations to other representations. We can note, here, a parallel with semantic theories that hold semantic properties to be distinct from facts about the use of a representation, most obviously resemblance theories, but possibly causal theories. Most obviously in the resemblance case, the syntax itself goes a long way to explaining the content.[10] The payoff of an intrinsicist view of syntax, not unlike an intrinsicist view of content, is, hopefully, that such identifying structure goes some way towards explaining how representations, in turn, explain cognition- in particular the language-like features of

---

[10] Likewise the connection between the notion of relational syntax and use theories of content seems tight; a representation's relational syntax can be closely tied to its content. For example, 'give' takes an object and indirect object because it is possible to give someone something. It is the fact that relational syntax of 'give' and its meaning are entwined in a way that perhaps makes syntactic typing in this case not separable from semantic typing (although, linguists do distinguish between semantic and syntactic roles). If content were determined by use, on the one hand, and syntactic type was a matter of intrinsic syntax, on the other, then syntactic and semantic type would be estranged. One would be determined before the other, content would depend on the 'wiring' that controlled the syntactic types

thought such as systematicity and productivity. Rather than rehearsing those arguments, I will now consider alternative views of the relationship between structure, syntactic type and cognition.

By way of preparation, it is worth saying something about the notion of discreteness, since it seems to be at the heart of 'constituent structure'. If a structure has constituents (perhaps, as opposed to mere regions), it must have delineable, i.e., discrete parts. (Intrinsic properties are really best enjoyed by discrete objects, since they have those properties to themselves.) Discreteness is best contrasted with continuousness. By continuous representations, I have in mind representations which merge into other representations so as to lose their form. Representations can be continuous with one another by being:

1. superimposed (as opposed to concatenated) as some connectionists take the representations in neural networks to be, or,

2. (physically) continuous with those representations physically closest to themselves- for example, in the way the states of something like an expanding balloon merge into one another. (This second sense is from Deitrich and Markman.)

Thus, when Millikan (in Dahlbom (ed) 1993, p99) describes an *intentional icon* as a response to the environment that is capable of varying in correspondence with a variance in the environment, then she may well describing a continuous representation. That is, if the variation in the form of the representation matches a continuous change in the world. On the other hand, the idea that change in a representation can represent change in its referent does not necessarily require that such representations be continuous in form with each other. The integers, for example, may be used to represent change as a non- continuous change.

Markman and Dietrich characterise a discrete representational scheme as 'gappy'.

A set is discrete if and only if it has gaps between its members. If the set has no gaps whatsoever between its members, it is continuous (and visa versa). For

example the set of rational numbers is discrete. … The set of real numbers is
continuous because this set has no gaps- it is missing no numbers (ibid, p101).

Hence, in the case of the continuously changing Watt Governor's spindle arm angles,
there are no discrete representations because there is no gap between any two angles
separated by an instance in time. The same applies to the bimetal strip in a thermostat
but does not apply to the mercury vial switch attached to the bimetal strip; this has the
discrete states of being *on* or *off*.[11] Some connectionist network states are represented
as potentially continuously valued. For example Elman (1992, p138) states that
'[d]istributed representations provide a high dimensional, continuously valued space'.

The superposed version of continuity is synchronic continuity. 'Discrete', in this
context, means something like *finite physical magnitude* in the sense of not blending
with another representation. Dietrich and Markman use the particulate quality of
genes to make the notion clear. Genes don't blend, but if inheritance ran on blood
instead of genes, then our parents' genetic endowments to us would be as continuous
with one another as blended blood.

'Retrievability' is related to discreteness, because, as Dietrich and Markman put it,
when 'snow' and 'bicycle' are combined in 'snow bicycle', for example, 'snow' and
'bicycle' are still discrete and so retrievable (although, Clark would say this is relative
to the processing device). In contrast, when fluids combine, it's difficult to get the
constituents back again. So, Dietrich and Markman argue, '[t]he fact that constituent
concepts are easy to extract from combination suggests that they do not combine in
the manner that fluids do' (p 110). In other words, an important aspect of cognition
may be about keeping representations extractable- and it looks like that is going to
require discrete representations both in the sense of not being lost in a mix of some
kind and in having a particular magnitude that is gettable. This is the point- every arm
angle of the Watt governor has a particular size, but is not easy to get hold of. We can
note that retreivability in cognition is the type of thing Fodor thinks constituent

---

[11] Discreteness looks a bit like digitalness, and continuousness like analogicity, at least according to
that way of understanding the digital /analogue distinction.

structure will help representation to explain.[12] (Certainly, retrievability seems a feature of multiple usability.)

Millikan's relational syntax

Why would a representational system not use discrete constituent structure? The structure of complex lisp symbols, for example, intuitively seem more easily readable, extractable and multiply useable than the hidden constituent structure of Chalmer's superposed representations from chapter 1. One reason might be that constituent structure requires the syntactic typing of complex and simple representations by their form and this might be a difficult thing for the human brain, for instance, to achieve. Millikan is one theorist who is interested in structured representations but rejects the idea of what she calls 'mental typing' by form.

We have already come across her notion of intentional icon as representation 'proper'. Millikan (in Dahlbom (ed), 1993) depicts these as objects capable of combination, possibly in the way two maps might be overlaid to produce new information in virtue of corresponding components being placed on top of each other. It would seem, here,

---

[12] Just on this point of separability, Ross Gaylor (2003, conference paper) has recently claimed to have demonstrated superposition and retrievablity within a connectionist network; that distributed representations which have been superimposed are 'gettable', providing they are put into superposition in the right way. Namely, by using the right kind of vector multiplication and by choosing the right vectors. Gaylor argues this allows connectionism to meet Jackendoff's (2002) challenges to the connectionist community concerning the virtues of structured representations, such as the massiveness of the binding problem, the problem of dealing with multiple instances, the problem of variables, and the compatibility of representations in working memory and long-term memory. He states that the 'essence of these problems is the question of how to neurally instantiate the rapid construction and transformation of the compositional structures that are typically taken to be the domain of symbolic processing'. If Gaylor is right, we might begin to consider the possibility that Dietrich and Markman may be wrong about the importance of discrete representations in the sense of particulateness. However, it would leave open the question of whether discreteness in the diachronic or blending sense is important since the choice of vectors as representations that Gaylor talks about is partly a matter of the vectors being distanced from each other so that they can be put into superposition safely.

Gaylor's claims are similar to those of other connectionists like Horgan and Tiensen and Smolensky (below) who likewise seem attracted to a hybrid cognitive architectures that run on symbolic or name-like representations that are nonetheless distributed or structured, in some way, in isolation but also capable of combination in superposition. The key question is, what computations can be done on such structure in superposition?

that these shared components would need to be matched somehow. If this takes place according to their form, then we seem to be talking about constituent components, in the Fodorian sense of the term. Yet Millikan decides mental typing by form is not a good idea:

> ... certain physical similarities, and not others, group tokens into types only because someone, say the general public, reads the tokens that way. Physical similarities, if salient, are so only because made so by symbol users. Similarly, if there are symbol types in mental language, these must correspond to ways the inner system reads mental tokens. ... I will argue that it is most unlikely that typing for mental tokens rest on physical similarity (1993, p111).

It seems, then, that representations are to be matched and typed not by their form but by the use that is made of them- that is, how the system reads them. This would seem to leave Millikan's reference to maps and their shared structure sitting out on a limb since shared components are now to be identified in the use of them as shared components. Normally we think of maps and their components as something to be made use of in virtue of their structure, not as structured in virtue of their use.[13]

Millikan's view is what we might call a 'use' notion of syntactic type- it is the fact that different representations are 'read' in the same way, or play the same role, that makes them the same representation. She suggests that mental typing (the determination of which equivalence class a representation belongs to) could be a relational property between tokens, as, for example, in co- reference between nouns and pronouns. In 'He asked himself a question', although 'he' and 'himself' are very different shapes, their relative positions in the sentence determine their co-reference. This is a little confusing, since although a system may come to read the two terms as being co-referring in virtue of their relative positions in a tree structure, we might ask if 'relative position in a tree structure' is itself an intrinsic or relational matter? How are two cases of the same relative positions being occupied to be identified? Surely it is a matter intrinsic to the representation of the sentence.

---

[13] As we see in the next chapter, Cummins (2003) makes these sorts of points against Millikan.

Millikan, in any case, seems to want to rest her view of mental typing on relational syntax, and give up on intrinsic syntax. Millikan wonders if

> the brain groups tokens into types by wiring them together so that whenever one token of the type lights up they all do. Or perhaps different tokens are merely different lightings up of the same individual symbol, its place in a variety of iconic structures being held by external connections that can be activated either simultaneously or alternatively (connectionism)(ibid, p117).

In these imaginary cases, the connections between representations count, not the physical structure of the representations. The sameness of two tokens is something tacit in the representational system, not something that is 'judged' according to their form- it is more like something that it is shown in the competence of the use of the representations:

> …recognition that elements within two intentional icon tokens represent the same occurs only when the icons are used jointly, pivoting on the overlap or middle term, to produce a new icon or an adapted icon (p 116).

Millikan's approach seems, then, to involve collapsing semantic and syntactic typing together. (Providing there are not two types of roles in a system such that one type establishes syntactic type and one establishes semantic type.) *Use* seems to establish both semantic and syntactic type. But, again, this seems a little at odds with her notion of intentional icons, above. Remember, firstly, these icons enjoy 'variation in response to variation in the environment'. Such variation would seem in part, at least, to be a matter of intrinsic syntax- if it just variation in use, you start to wonder what all this use is predicated on. And, secondly, her notion of common components, as noted above, is hard to understand in the absence of the typing of those components by form- even if it is the *use* of common components as co-referring that, metaphysically speaking, makes them co-referring. In the above quote, she doesn't explain how the two middle terms are matched, only that their matching tells us the icons they are part of are the same icons.

Whether or not Millikan's views on representation are consistent, her case against typing by form seems to rest either on the biological implausibility of the brain being able to make consistent enough tokens, or the general claim that syntactic type is an observer/user dependent notion. The first possibility resonates with Dennett's cognitive wheels 'slur' against formal symbols. This is the worry that formal representations are too neat or too single purpose to have been designed by nature-rather like wheels. My response to this is, 'who knows?' Natural selection built the eye, the immune system and DNA- which seem to have either single purpose parts of repeatable, finely structured components identified by their form.

The second objection is that 'mental type' is a functional notion. This is potentially a bit ambiguous. Millikan could be referring to mental states either as syntactic or semantic entities, though in this case she *is* discussing syntactic properties. Generally, though, it is not always clear in what sense people are referring to representations. They may also be referring to representational states as psychological states. Fodor, for example, seems to think that in some ways mental states are functionally individuated, *qua* beliefs or desires etc, but that that fact does not rule out the possibility that they enjoy constituent structure.

> what's at issue, however, is the internal structure of these functionally individuated states. (1987, p 138)

Fodor, then, can consistently maintain that syntactic or constituent type is not a functional notion. He has in mind, here, mental states with complex semantic properties or objects, and given that constituent structure requires discrete representations identifiable by their form, he can't also hold these simple constituents to be functionally individuated - at least not as syntactic objects. (We can also note that *qua semantic* objects, it's a different story, though still not a functional one. In 'Psychosemantics' he thinks 'concepts are individuated by reference to the properties they express, thoughts by the states of affairs they correspond to, and so forth, (1987, p91).)

By contrast, Millikan seems to want simple representations to be syntactically typed by use, and at the same time she wants to hang on to constituent structure. I'm not sure this is possible. If inference depends on common components of a representing structure to be matched, then surely some property of the common components must be the basis of the matching; that is, the use of the representations cannot be the matching itself, otherwise the inference that results from the matching is not based in the premises so to speak, but merely the act of matching. There must be facts about the premises that determine the match and, thus, that the inference is possible.

I consider now another more concrete attempt to marry constituent structure with relational syntax, and then make some general remarks about the relationship between constituent structure, cognition and syntactic typing.

The connectionist version of constituent structure

Fodor argues that it is best to explain overlap in complex representations/contents by appeal to common constituents. He has in mind, here, complex representations composed of simples that don't physically change when they form complexes. We might say that they remain discrete. Hence, if we can type simple representations by their form, we can type complex representations by theirs. However, as we saw in the first chapter, a non-standard connectionist interpretation of Turing has also been floating around idea space in the last decade or so. The idea is that it is possible for a representation to encode (syntactic) structure but not in virtue of part whole relations. For example, in Elman's (1990/1992) language parsing networks, an activation state representation of 'boy *qua* subject' does not contain a 'subject' and 'boy' part. In one sense, it is a complex representation- because it has complex semantic structure that seems to be read by the network, given that subsequent states of the network seem to share appropriate *relative* syntax with it.

Elman (1992) makes this point and more besides,

> … the spatial organisation within the token space [eg, consisting of tokens of boy] is not random but reflects differences in context. …the tokens of <u>boy</u> which occur

in subject position tend to cluster together as distinct from the tokens of <u>boy</u> which occur in object position. …thus the network has learned not only about types and tokens, categories and category members; it has learned a grammatical role distinction which cuts across lexical items.

This introduces us to State Space Semantics as *token space* semantics, where divisions in a network's activation space are taken to reflect or actually *be* conceptual divisions. For Elman, the structure of his network's token space reveals a 'knowledge' of types and tokens, and categories and category members. Churchland (forthcoming), as we shall see, takes the same strong view of the conceptual achievements of such networks.

These are tricky cases for it seems relational syntax and intrinsic syntax get mixed up. The intrinsic syntax of a 'boy *qua* subject' vector apparently allows it to play a certain role in the network but, at the same time, the role determines, or at least reveals to us, the supposed constituent structure of that representation. In other words, form and use seem to co-determine each other. But, also, form and structure come apart. The supposed internal semantic structure of these activation states is not reflected in the intrinsic syntax of the representation. Thus, we have a debate about the nature of causally effective syntax. The question is, must useful syntax be 'part-whole' syntax- the kind of syntax that is amenable to a programme- or can it be the type of syntax that is revealed in the running of the programme, or just the workings of the representational system? The explanatory order of cognition and representation/syntax seems at issue.

I turn to another example of relational syntax as the syntax of structure in more detail. Horgan and Tiensen (1996) make the case for the existence of encoded syntactic structure in Berg's (1992) connectionist parser. This network, like Elman's and Chalmer's above, purportedly produces propositional representations containing syntactic structure but not parts.

Given the nature of the information-processing tasks performed by the model... and given that the trained up system has the capacity to perform properly on inputs not among the training corpus, the representations in this model clearly encode *syntactic-constituency relations* [my emphasis] in a way that makes the

representations amenable to a certain kind of structure-sensitive processing: parsing. Thus, the representations exhibit a rudimentary form of effective syntax. Yet they do not encode syntactic constituency via a part/whole relation (1996, 74-75).

That is, we have, here, no parts and wholes but we do have the encoding of syntactic-constituency relations. Moreover, this encoding is decoded or processed, which fulfils the use requirement of representation, even if it does not amount to multiple use. Berg's parser is based on a Recursive Auto-Associative Memory architecture. In effect, the encoding and decoding in question amounts to the packing and unpacking of complex representation. Because of the hour glass shape of the network, i.e. there are fewer nodes in the middle layer than in the input and output layers, its middle layer representations are considered fully distributed across those nodes, since they are representing at least the same content as the input nodes have represented. (The extra representational work is taken up evenly amongst the middle layer nodes). Words, phrases and sentences can all be represented or stored by the same hidden layer and at the same time. This is the purpose of the network being recursive.

The network is recursive in that hidden layer states are fed back into the input layer as new input is added to the network. Representations for phrases are, thus, successively built up out of representations for words (and sentences built up of phrases). As the syntactic structures of the middle layer activation state representations get more complicated, no extra hidden layer nodes are required to encode the extra structure. The reverse happens when a complete middle layer representation is gradually decomposed in successive output vectors- as the encoded syntactic structure gets simpler, the middle layer representations remain the same length in nodes.

The evidence for structure-sensitive processing lies, in fact, in how a sentential hidden layer representation can be redistributed into four output pools. These pools stand for the categories of a Chomskian (1981) 'X bar' structure: 'specifier', 'head', 'first complement' and 'second complement'. After a first decomposition, the specifier and complement vectors can be copied back into the hidden layer and further decomposed. For example, a verb phrase can be broken up into a verb and preposition phrase, which can in turn be fed back into the hidden layer to be broken up into a preposition and noun phrase. In the end, 'each of the atomic constituents (i.e., the

words) turns up at one point in the head-word (or specifier) pool of the output layer'
(1996, p59). The relational syntax of the whole sentence then can be unpacked in this
way.

That a middle layer activation state can be unpacked like this suggests, at first glance
at least, that it encodes recursive structure- it enables the network to create a tree
structure in time, rather than in space. And, since all such representations are the same
length, a part whole story seems incorrect.

Horgan and Tiensen employ a dynamical systems theory perspective to make sense of
the system's behaviour and the way it seems to encode syntactic structure. The points
in the system's activation landscape come to be organised so that when the system
passes through those points, it does so in a clever way- that is, in a way that reflects
the structural relations between those representations the points stand for. 'The
realisation relation and the landscape topology end up 'made for each other''. In other
words, the dynamical landscape of the network is such that its representations follow
each other as if constituent structure were driving their order. And hence, according to
Horgan and Tiensen it does:

> [Training]…systematically moulds the activation landscape and systematically
> positions words and phrases on the landscape, so that a place is made on the
> landscape for words that it has not seen. Representations of clauses and sentences
> become similarly systematically related to the representations of their constituent
> phrases and clauses. That is, syntactic constituency relations become systematically
> embodied mathematically in the relative positions of points with representational
> content on the moulded activation landscape; that is the representations....exhibit
> nonclassical, *dispositionally realised syntactic structure* [my emphasis] (1996,
> p62).

In other words, all the relational syntax between words, phrases and sentences that the
system needs to handle and produce, even for using words it hasn't seen, gets packed
into the system's weights and the system's choice of middle layer representations-
which are, of course, a product of those weight changes. This is how the activation
landscape gets built. Given those weights and those choices of representations, the

syntactic structure of phrases and sentences, as represented in the middle layer, becomes 'dispositional structure'- if that makes sense[14]

Fodor, we can imagine, would be inclined to deny that 'dispositional structure' does make sense- along the lines that structure is one thing that isn't dispositional. Structure is the source of dispositions, not the effect (as we shall see, much like Cummins might think that represented content is the raw material of cognition and not the effect). For Fodor,

> mental representations are, by definition, co-tokened with their *classical* constituents. But it is not the case that they are, by definition or otherwise, necessarily co-tokened with their derived constituents (1998 b, p119).

This draws attention to the strangeness of the notion of 'derived' or 'relational' constituents. For one thing, how could such constituents get processed if they are only potentially derived?

> Now, I suppose it goes without saying that merely *counterfactual* causes don't have any actual causal effects.' (1998, p118)

Still, Horgan and Tiensen think effective relational syntax is doing work in Berg's parser,

> Since the relative position relations [between vectors] play an essential role, in combination with the topography of the activation landscape, in ...[the] performance of its task, this constituency-reflecting relational structure among the representational points constitutes *effective* syntax (p 162).

Horgan and Tiensen, thus, claim the input, output and middle layer activation vectors of Berg's parser constitute a LOT of sorts. As complex representations, with their

---

[14] For determining the strength of the 'proof by existence' factor, here, the 'has not seen' clause is important. Connectionists see the training of a network as setting it up for novel inputs. A classicist might point to Pinker's work on the past tense (1999) that purports to have cast doubt on how efficient it is to try to anticipate inputs in this way, rather than use a rule. And in general, the doubt for this story is how many sentences the network could deal with- how many and what kind of sentences the network could anticipate in its structure or in the topology of its state space.

components superposed, they stand in definite positions in state space to each other in virtue of their physical properties. Horgan and Tiensen argue that the relational structure of the state space (or, perhaps, the activation states) constitutes effective syntax – apparently because it is 'constituency reflecting', and perhaps because it is a syntactic matter, in that it is about shape, albeit relative shape.

So, on this story, the syntactic structure of a complex representation needn't be realised by its intrinsic properties alone- the relations between complex representations can determine the syntactic structure of the relata. This way of putting things is initially attractive because it looks like syntactic structure is still being physically encoded in the representations themselves- it's just that the decoding of that structure requires comparison with other representations; it cannot be done in isolation from other representations. However, we shouldn't be blinded to the fact that syntactic structure is being determined by use, here- the relative positions of representations in activation space simply reflect how the representations are used. Aydede[15] (1995, p 15) puts it that such syntactic properties are '(metaphysically) determined' by a computational (causal/functional) role.

The notion of structure as use, which we saw Millikan appeal to, is a difficult notion. It seems to bring a whole lot of mereological problems into psycho syntax. Because *use* now determines type, a representation's syntactic properties are determined (and perhaps shared, if that makes any sense) between itself and the processing medium around it- in this case, the weights of the network in question. At least, the weight states, or the topological landscape they create, *determine* the syntactic structure of a complex representation. Aydede tries to put the focus on the physical properties of the representation itself. He argues that a network places an input vector in a region of its middle layer activation state space according to that vector's abstract shape, i.e.

---

[15] Aydede (1995) finds much the same take home message as Horgan and Tiensen from RAAM networks and Smolensky's tensor product representations. He depicts the idea that common structure reflects common components as just a *formal* description or requirement.[15] It says nothing about how this requirement is to be met. All that matters is that complexes *are formed* out of simples; and operations on those complexes are sensitive to this structure. Syntactic properties are '(metaphysically) determined' by a computational (causal/functional) role (ibid, p 15). We are being asked to imagine, then, multiply realisable syntactic properties- where concatenation is just one way of realising syntactic structure. Where Fodor is wrong, on Aydede's account, is in assuming that forming complexes out of simples *requires* concatenation; Goedel numbers are a counter example, as are superposed representations in middle layers of RAAM networks.

'specific numerical values at specific positions. There is a clear sense in which this is the 'shape' of this kind of representation made computationally relevant...' (p14).

> Its 'shape is indeed radically different at some level of analysis from the shape of concatenatively realised symbols of conventional Von Neumann style computers. But from the perspective of a properly understood LOT, they should all count as symbols in LOT, and the processes are properly called symbolic processes, because what counts is the reliable transformation of representations themselves... (1995, p16).

Horgan and Tiensen agree with Aydede: 'syntax is simply the *systematic* and *productive* encoding of relationships'. As long as the system can decode the relationships between representations so as to decode representational relationships within complex representations, then that it is all that matters. However, this seems to me an equivocation, of sorts. If syntax is just the 'systematic and productive encoding of relationships', then, in this case, that encoding includes the topological landscape, as well as the representations themselves- in the absence of the landscape, or the system's dispositions, the shapes of the activation states don't encode syntactic structure. We might say that the representations are too dependent on the processing medium. Consequently, to anticipate Cummins in chapter 3, I think Aydede and Horgan and Tiensen really are using cognition to explain representation/syntax and not representation/syntax to explain cognition.[16]

---

[16] In the literature, the debate followed a slightly different course. It came down to whether network processing of the kind of compressed representations we are talking about counts as structure sensitive processing. Fodor and McLaughlin (1990, p 200) argued that since the constituents are not actually there in the compressed representation (since there are no physical delineations), if the answer is yes, then the structure sensitive processing in question must be a miracle.

Talk of miracles may have been a little hasty- at least on one ungenerous reading. It may be possible to set up a network or cognitive system to make all the right moves in advance, so to speak- if all those moves either can be known and prepared for in advance or if they are all much of a muchness, or all of the same basic kinds. Following a trajectory in dynamic activation space is, after all, just a matter of churning out the usual associations. This setting up process, as the connectionist research shows isn't a miracle. This is the appeal of dispositionalism or associationism or behaviourism, and I think this appeal *vis a vis* human behaviour depends on how limited our behaviour is. If some of our behaviour repertoires are as limited as those of certain connectionist networks (unlike our mastery of the past tense, according to Pinker), then maybe intrinsic syntax is not a the important idea it seemed to be in Turing.

Hence, Aydede points to Chalmer's 1990a network as an example of such a 'miracle'. This network took previously compressed representations of active sentences and learned to transform them into their

I should note that Horgan and Tiensen also think connectionist representations can and must have identity criteria physically built into them and into the system that uses them. They don't give up on intrinsic syntax completely- they just don't want to use it in the same way that Fodor does. Horgan and Tiensen see the importance of representation identity as follows:

> For each individual i and property P that the system has a way of representing, a way of representing that 'i has P' is automatically determined... And [that representation] must automatically have its context appropriate causal role (p 82)

According to Horgan and Tiensen, systematicity follows just if the components of the state representing 'aRb' can equally be used to represent 'bRa', and no other components have the same role as 'a', 'R' and 'b'. Something like Fodor's productivity purportedly ensues from some minimal combinability of the components of the representational system. Horgan and Tiensen don't want to buy into too much language-like representational machinery- there needn't be recursive combinations - just minimal combination of names and predicates without quantifiers or connectors (p 91). However, they clearly think their version of syntactic structure buys them the benefits that a LOT delivers; in this case- by positing discrete simples with physical identity criterion and a way of encoding complex syntactic structure. (Of course what they mean by using a simple to make a complex isn't what Fodor had in mind.)

Interestingly, Horgan and Tiensen also agree with Fodor's well-known pessimistic conclusion (eg, 1983, 2000) that as long as computation is seen as essentially algorithmic, problems of relevance and abduction will remain intractable. Horgan and Tiensen look to intractable cognition, instead; they claim that their 'non-algorithmic' cognition is, however, still a species of classical cognition- after all, giving up on

---

active counterparts, even though it had not learnt to compress the representations in the first place. Chalmer's network is interesting because unlike auto associative networks, it does more with an input than regurgitate it at the output layer- this seems a little more like real structure sensitive processing.

But again the same problem arises, since it is not really accessing structure it has to depend on 'guessing' that structure on the basis of the similarity relationships between representations. Again, this is all fine, as long as the tasks our brains perform as so accommodatingly structured. *Qua* structure sensitive processing, in one sense, such cognition would still be guesswork.

syntactically structured representation would mean throwing away your best tool for explaining semantic coherence (p 54).

I think this is trying to stretch the Turing's insights about syntax too far. Their idea is to keep syntactic types, but to give up on part-whole relations between those types, and rules that apply to the processing of those types in combination[17]. Horgan and Tiensen are hoping a 'high dimensional activation landscape' will do the computational work non-algorithmically, instead. But if syntactic type is determined in consumption or by use, we seem forced to abandon the formalist's dream. We seem to lose the distinction between syntax as the rules of inference and syntax as the intrinsic properties of the representation.

That is, as I argued above in relation to Millikan, if it is *use* that determines the formal identity of a representation, and not its intrinsic properties, then *use* (or, the one *use*) cannot also constitute implementation of the rules of inference as well. And I think the formalist dream relies on the two type of syntax being present.[18] An automated formal system has to be able to recognise a form in virtue of its intrinsic features and then change or combine that representation as the system's rules of inference require. It cannot do both at once because it cannot determine the identity of a representation whilst it is using it. It is true that Horgan and Tiensen have given up on (hard) rules, in any case, but that doesn't leave their relational syntax in the clear. Remember that one of the points of positing syntactic types is to explain semantic coherence. If Horgan and Tiensen are relying on relational syntax, then they are also relying on cognition to explain typing in combination and cannot be using typing in combination to explain semantic coherence.

---

[17] Though, this still leaves open the possibility of soft laws, or 'defeasible cognitive tendencies subserved by the non-linear dynamics of the brain's neural networks', as O'Brien (1998) puts it.

[18] Consider the ~ symbol in propositional logic. That ~ means 'negation' is clear from the use of ~ : it is legal in propositional logic to make the transition from ~ ~ A to A. Negation is the only operator which can do this coherently. Its relational syntax fits its semantics (as Cummins' (1989) tower bridge has it.) However, this cannot be what determines the syntactic type of ~ as well. This seems to be a matter of being a wavy line- even if there is some slack in that notion.

If it were true that the content of ~ was somehow in the structure of ~ then the resemblance theorists would be correct in this instance about content inhering in form and syntactic type and semantic type would amount to the same thing. In this case that is not true: ~ is not isomorphic to the concept of negation. (Not that any resemblance theorist would claim it was, anyway.) In this case, content does seem to inhere in its relational syntax. That said, use of ~ still seems to require syntactic typing not by use but by intrinsic properties.

I think we give up a lot if we give up the brain's capacity to type by structure. If anything is right about the formalist or Turing's idea, then the brain has representations typed by physical form. Whether the rules of cognition are represented explicitly or implicitly, or perhaps even if there are no such laws, real constitutive structure is still going to give the brain a head start in representing structure and computing over that structure- half the representational and computational load will be carried in the form of the representations. This is where Cummins has a point to make against theorists like Millikan, Clark, Horgan and Tiensen, and Aydede. Curiously, it seems to put him in the same camp as the (anti-resemblance) classicists.

Two anti-representationalist responses to Horgan and Tiensen's position

Horgan and Tiensen receive criticism from both sides: from the believers in real constituent structure, and intrinsic syntax, and from those who don't see representation as the most important tool in our armoury for explaining cognition, such as the DST theorists. I mention these responses to Horgan and Tiensen because they shed light on my claim that their position is not consistent. Consider that,

> The argument that syntactic structure we discern in cognition must be mapped back into the *causes* of cognition is challenged by DST (Garson, 1998, p311).

Garson is referring to the 'effects mirror causes' intuition that he rightly identifies as an important plank in classicism. Garson thinks Horgan and Tiensen are still under its spell to some degree. Bechtal (1998, p307) might say Horgan and Tiensen are caught between providing a mechanical explanation and a dynamical systems type explanation of cognition. Mechanical explanations tend to assume localisation in a system, i.e., that there are components of system that preform various tasks; whereas, DST type explanations are sometimes taken to reject this kind of task decomposition and localisation. In Horgan and Tiensens' complex representations, there is no locating sub-representations, but there is locating the whole complex representation and the simple representation before combination.

Bechtel argues that some DST approaches are compatible with mechanistic explanations, eg, as in the Watt Governor case, but others are closer to covering law explanations rather than mechanistic explanations (1998, p311). The question is, what type of explanation are Horgan and Tiensen attempting? They seem to want some localisation for representation in their explanation, but in their appeal to dynamical notions like 'attractor's they seem to be engaging in a covering law type explanation. That is, an attractor is less like a component and more like a description. This, in effect, is my point that their syntax is part intrinsic and part relational – their reference to attractors is another way of referring to that relational syntax.

Garson points out that the 'Cartesian notion of Principle P' (which is expressed in his quote, above) may be inappropriate to cognition if the brain is a chaotic system. The notion of self-organised structure discredits the intuition that patterns at one level of organisation map onto patterns at a lower level. If we think of cognitive types as volumes in phase space in a chaotic brain, perhaps one driven by different attractors, as Horgan and Tiensen argue, then these volumes may well be very complex and gerrymandered. Thus brain states corresponding to tokens of cognitive types may well be very dissimilar to one another- represented by distant point in phase space. To use Garson's example, the concept 'happy' may be realised by two very different brain states which nonetheless are part of the same collection of brain states that are about happy people and lead to the same type of inferences eg, that the happy person is smiling. In other words, relational properties may be multiply realised within the one system.

This seems to stretch Horgan and Tiensens' claims to a language to breaking point. The difference with the LOT picture, as Garson imagines things, is that the physical properties of the representation don't drive the semantics of the system in virtue of being a particular physical type, as Horgan and Tiensen and Aydede try to argue. There may be no physical similarity that explains the role of 'happy' representations. Similarity in their causal roles may emerge sui generis, and, 'the presumption that internal syntactic structures must exist to 'keep the brain in line' loses its appeal' (p 310-311). And, here, this thesis concurs- dynamical theorists don't care for intrinsic syntax.

I think Hardcastle (1997) agrees when she argues that DST explanations don't fit well with Turing type notions of computation. She rejects Horgan and Tiensens' attempts to adapt Marr's three levels of description of a cognitive system to their soft law account of cognition. Horgan and Tiensen distinguish the *(non algorithmically) computational* from the *representational/mathematical* from the *implementational*. Hardcastle argues that, in effect 'these three types of description each presents us with different ways of accounting for the same set of trajectories which, in turn, describe (probably) a set of neural firing rates' (ibid, p 322).

For example, the two higher levels- *mathematical transitions* and *cognitive state transitions*- are not separable from each other. A high level description of cognitive state transitions, e.g., as a Lorenz attractor, is just another description of the mathematical state transitions, albeit perhaps more general (p 377). And, the implementational level, according to Hardcastle, also becomes impossible to keep distinct from descriptions of function. Dynamical accounts don't really require an implementational explanation,

> [i]f Horgan and Tiensen are correct in emphasising the centrality of dynamical systems in psychology, then the traditional distinctions between descriptions of function and descriptions of implementation disappear, for they both amount to the same thing (p 377).

Perhaps Bechtal can help us explain this. If the system is to be explained as an instance of a covering law of dynamics (and Horgan and Tiensen are hoping that 'we can uniquely characterise any dynamical system by its dimensionality and the topology of its attractors' (Hardcastle, p 376)) then, there is no need to specify any causal powers of any implementing components. The law, after all, is 'doing' the implementing, so to speak.

Perhaps, another way to describe matters is to say that, in Horgan and Tiensen's vision, it is relational syntax all the way down. Consequently no mechanistic explanation is involved. If trajectory in state space explains what is going on in the brain, it doesn't make sense to say that the system is *computing* a particular complex

trajectory. That is, no algorithms are required that, in turn, are implemented by computations over representations with intrinsic syntax. The trajectories of the brain, the mathematical transitions in the brain and its implementational properties amount to the same thing. This is consistent with the claim that the complex representations Horgan and Tiensen are talking about don't have the right kinds of parts to do the right kind of implementing or to take part in any mechanistic explanation, anyway.[19]

Interestingly, Bechtal might still agree with Horgan and Tiensen that the middle layer activation states in Berg's parser were representations of structure. I say this because Bechtel thinks the representations in Elman's (1990) network, such as 'boy *qua* subject', help provide a mechanical account of a dynamic system. According to Bechtal, we may not find a complete localisation of functions, but we do find information bearers. That is, we do not find a localised component responsible for representing the 'subject' part of 'boy *qua* subject', but,

> Nonetheless, Elman is able to show that the relevant information is captured in the representations on the hidden units. …what must be done in such accounts is to explain how information is carried through the system and made available to other parts of the system that use it (1998, p 313).

Bechtal is not necessarily claiming Elman's representations have constituent structure but that they carry information about constituent structure –and, presumably, that is enough to make them representations of constituent structure. However, we might ask Bechtal how they are carrying such information if they don't possess the appropriate syntactic structure- where is the information about 'boy', 'qua' and 'subject'? There do seem to be gross representational components in such networks, as their dynamic state space analyses suggest. However, the parts only seem to go down so far- after a while decomposition ends, just where the encoding looks like happening, and hence, I

---

[19] We can note here a general criticism of connectionist representations *qua* representations (see Ramsey, below) pertaining to Marr's levels of description. It may seem, in general, as if the physical facts about connectionist networks, namely, the facts about connections between nodes are also the algorithmic facts- the connection facts are in effect all the networks 'algorithms' superimposed together in an indescribable manner. Perhaps, then, the two levels collapse. It might then be further argued that there is little point in speaking about connectionist representations, since the whole point of representations is that they be made subject to algorithms.

would argue, no encoding and hence no information-carrying representation seems to be in play. In general, perhaps state space divisions are best used in a covering law account of the network, rather than a representational *qua* mechanical account of how the network works.

If Bechtal, Elman, Horgan and Tiensen and Aydede are correct, there must be some other way of being informational about constituent structure apart from using constituent structure as the bearer of that information. Perhaps an abstract resemblance story might do the trick. Ironically, it is perhaps easier to tell a simple resemblance story about classical representation since physical boundaries are used to represent syntactic and semantic boundaries, in effect, minus all the other syntactic and semantic paraphernalia of tree structures etc. In contrast, a State Space Semantics version of resemblance representation will probably use a second order structure where the relationship between representation and represented is not so straightforward. I examine this option in subsequent chapters.

How should we identify representations?

I finish this chapter by addressing the problem of understanding the relationship between syntactic and semantic properties in light of the importance of representing constituent structure. As I noted above, we can individuate a representation semantically, for example, as Fodor does, in virtue of what property or state of affairs it expresses, and we can individuate a representation syntactically, for example, in virtue of its formal properties. Representations enjoy both syntactic and semantic properties. But, we can ask in what way representations share these two features-separately, co-dependently, or in some other way.

It is tempting to leave semantic properties out of the picture altogether and attempt to individuate representations by their syntactic properties alone. Stich's (1983) syntactic theory of mind (STM), which he contrasted with the representational theory of mind (RTM), is a case in point.

The STM views mental states as relations to tokens of purely syntactic objects. Generalisations detailing the interactions of mental state tokens describe them in terms of their syntactic types. On the matter of content or semantic properties, the STM is officially agnostic. … the STM is in effect claiming that psychological theories have *no need* to postulate content or other semantic properties like truth conditions (1983, p186)[20].

Stich defends STM against such doubters as Pylyshyn, who doubts if it is possible, in explaining behaviour, to leave formal representations uninterpreted. The putative problem is that certain patterns in behaviour might not be explicable, and certain generalisations not possible, in the absence of semantic specification. For example, we might need to explain not only the mechanics of behaviour, but why those mechanics arise in a given situation- that is, what the connection is between the world and those mechanics. Why, to use Pylyshyn's example, should the smell of smoke and the sound of a fire alarm both lead to the syntax of escape route detection and selection. (Even to call such syntactic processes by that name 'escape route detection and selection' is to interpret that bit of syntactic mechanics.) Surely, argues Pylyshyn, we need to allude to the fact that something about those mechanics is *about* fire, danger and the avoidance of danger[21].

Without going into the details here, Stich thinks STM doesn't miss any such generalisations about the cogniser's interest in the physical environment,

The thrust of my argument throughout this chapter is the STM theories can do all the explanatory and predictive work of content-based theories, and they can do it better (1983, p182).

---

[20] Strong RTM claims that the generalisations of cognitive science advert to content, weak RTM does not. Weak RTM does insist that syntactic objects enjoy semantic properties (p186-187.)

[21] Pylyshyn argues that 'Simply leaving them as uninterpreted formal symbols begs the question of why these particular expressions should arise under what surely seem (in the absence of interpretation) like a very strange collection of diverse circumstances… of course why these occur under such diverse circumstances is precisely that they represent a common feature of the circumstances- a feature moreover that is not found solely by inspecting properties of the physical environment. (E.g., what physical features do telephone calls warning of a fire share with the smell of smoke?)' (1980, p160, quoted in Stich 1983, p172)

I raise this issue just to make clear that I am not arguing for a syntactic theory of mind - I'm not sure that purely formal objects could count as representations, in any case. I am arguing, though, for an emphasis on intrinsic syntactic properties and syntactic typing. This follows from the claim that real constituent structure is a driving force in cognition.

What the problem of syntactic typing comes down to, I think, is deciding on whether to take formal logic as any kind of guide to mental representation or not. This involves accepting a view of the role of syntax *vis a vis* semantics. *Form* is what *formal* logic is all about, at least at first glance. In a formal logic, the syntactic properties of terms are established independently of the establishment of any semantics via an interpretation schema. Of course, as Millikan points out, this relies on the recognition abilities of the logician. But we can note that computers are able recognise syntactic types, and so we can question any 'question begging' accusations. As we know, the hope of the formalists, or classicists has been that if formal identities can be established, and we get the relational syntax right, i.e., the rules of inference, the hopefully we can let the 'semantics take care of itself' – there will be no need to worry about contradictions or about truths that generate falsehoods. Admittedly, so far, there still seems to be a need to worry about abduction, relevance and the frame problem, but those are other matters.

This way of thinking about representation suggests syntactic properties are separably accessible, by the system, from semantic ones. The formal view of representations conceives of a system capable of typing representations by their form both in isolation and in concatenative syntactic structure. Whether and in what way, in virtue of this typing, the system is, in anyway, gaining of access to the representations' semantic properties at the same time is a contentious point, which I consider in the next chapter. The resemblance view of representation also conceives of a system that is capable of reading structure, but a system that is definitely gaining access to explicit content at the same time. It needn't matter, as far as the resemblance theorist is concerned, so much that the syntactic type of the representation should be determined in this way because no extra access to its content is to be gained via recognition of its syntactic type. A system can go straight to content. (If the representation is of syntax-

then the structure of the representation would be the 'syntactic content' the system is after.) Views such as Millikan's are different again, in this regard, because they treat the recognition of syntactic type as something that is either implicit in the 'wiring' or emergent in the workings of the system, not something that is read as structure.

The formal and resemblance views both imagine a system that is sensitive to representational structure *qua* physical structure, of some sort. Of course, all systems will be sensitive to the physical structure of its components, but not necessarily *qua* representational structure. For the formalist, representing structure is always syntactic structure, even if it resembles a gross semantic structure as well. Whereas, I think it's fair to say that for the resemblance theorist, the representing structure of a resemblance representation can be more a matter of semantic detail and not really a syntactic matter at all. Maps and models seem in some way 'unrequiring' of syntax given the explicitness of the semantics they display.

This is not to say that, on the formalist view, semantic properties become redundant, or that evolution ever built syntactic properties before it built semantic properties. But the formalist view does seem to require that something about the machinery for representation might be prearranged, somehow- the sorts of things that could carry the simplest units of meaning might be 'organised in advance', so that brains might be able to recognise the representations it comes to use, by their form alone. I have argued, against Horgan and Tiensen, that the same applies to representations in combination.

If this is true, then the intrinsic syntactic properties of representations would need to be referred to in order to explain cognition, semi-independently of those representation's semantic properties. I don't think this is possible using relational syntax, alone. So, in the next chapter I'm going to half agree with Cummins that the intrinsic properties of representation should explain cognition- I say 'half agree' because I'm not sure that it is also true that intrinsic properties must also completely explain the semantic properties of representation.

Chapter 3. The big picture: does cognition explain representation or does representation explain cognition?

In the previous chapter, I made a case for intrinsic syntactic structure over relational syntactic structure in the representation of language. In his book 'Representations, Targets and Attitudes', Cummins argues that the *semantic* properties of representations, similarly, should be thought of intrinsic properties and not as relational, or 'use' properties. This leads Cummins to *resemblance* as the ground of mental content. In this chapter, I outline some of Cummins' argument and sketch a landscape of possible replies. I begin with some explanation of *use* and *non-use* content.

The psycho-semantic tool kit

We might think of the *use* or *role* of a representation as a matter of when it is tokened and how it is consumed. This might encompass both long and short arm notions of role; that is, roles relating to the world and roles relating to other representational machinery. Cummins' account of *use* is tied to the notion of a 'target':

> To use a representation is to apply it to a target. Uses, then, are simply applications. To specify how a representation is used on a particular occasion is to specify a particular target. To specify a general use of r…is to specify what targets r is (or can be) applied to. …The fundamental idea is simple: in a case of correct use, content = target (1996, p29).

By 'target', Cummins means whatever it is that a representation is being used by a system to deal with- 'the notion of representational target is essentially a functional notion.' The 'target' metaphor seems to suggest a mental representation must be capable of being 'aimed' at something in the world to count as representation. This requirement is possibly compatible with Markman and Dietrich's notion of a

mediating state, as well as Bechtel's approach to representation, though (purposeful) mediation and co-ordination are not quite as 'active' as *being targeted at something*. More precisely, they don't presuppose a targeting machine, which Cummins' calls an 'intender'. We can also note that being targeted at something (external) does seem incongruous with the role of putative representational devices analogous to logical connectives, or *internal* indicators, or pointers. In the end, as we shall see, Cummins wishes to rule out all non-resembling states as representations. This puts some putative roles for representation out of the picture since the role fillers, themselves, will not be representations, on his account- given that they are non-resembling. This makes it awkward for Cummins to talk about some kinds of use- since, strictly speaking they aren't uses of representations.

Cummins, for instance, has to describe the putative *uses of representations* in conceptual role semantics:

> the idea is to think of a meaning of a representation in a system Z as fixed by the cognitive transitions it enables in Z (1996, p30).

So, in this framework, non-resembling states may enable cognitive transitions, and hence be representations. We should note that this is not to describe conceptual role semantics as just a short arm, or narrow, theory of meaning:

> what Z applies *r* to will turn on what Z believes (and its other attitudes), together with the stimuli that impinge on it and its computational architecture. These factors are just *r*'s conceptual role in Z. So, conceptual role determines use and use determines meaning (ibid, p30).

Here, 'stimuli' are in the picture, so it seems Cummins has in mind something like Dennett's position on content,

> What makes a data structure about Julie Christie is that it is part of a system, the presence of which explains my capacity to pick her out of a crowd, answer questions about her in quiz shows, etc., etc., (1998, p 283)

This account of content relies on a wider notion of use than a 'targeting'- it's not exactly like pointing an arrow or, especially, 'pointing' a map at Julie Christie, which is the sort of targeting of targets by representations Cummins seems to be thinking of.

The various psycho-semantic theories on offer in the field involve different approaches to the relationship between representation and use. We can make out the following broad categories:

- Functional role theories such as conceptual role semantics. Representation *R's* content is a matter of the causal role *R* plays in the wider system, or perhaps the cognitive transitions it enables in the system, as Cummins puts it. What is clear, here, is that the focus is not on the intrinsic properties of the representing vehicle but on the vehicle's relations to other vehicles, perhaps things in the world and how the vehicle contributes to behaviour.

- Proper role theories, which are a subset of functional role theories. I take these to hold that meaning of *R* lies in *R's* role as selected for in evolution or the development of the organism. Dennett states that,

    [my notion of function] is a 'teleological' notion of function (not the notion of a mathematical function or a mere 'causal role', as suggested by David Lewis and others). It is the concept of function that is ubiquitous in engineering… but also in biology (1998, p359).

    That is, certain brain states, representations, can come to play a role that is part of a larger, evolved purpose. Here, the emphasis is on the history of the representing vehicle's relations with past referents, past vehicles of representation and past cognitive architectures- such that *R's* existence and role within the host system is picked out and accounted for.

    A version of proper role theories, namely, Teleo-informational semantics, holds that representations both indicate and have the function of doing so. (Cummins views proper role theories as teleological versions of causal theories (ibid, p55).)

- Informational or denotational semantics. If we subtract the 'teleo' part of 'teleo-informational' semantics we are left with *informational* semantics. According to the subtypes: *causal* and *covariance* theories, representation *R* means the object or property that *R* is caused by or *R* covaries with. Of course there is the complication that, in order to determine *R's* content, we must also specify which of *R's* covariances or causings are the content determining ones without referring to any teleology – or what R is *meant to* covary with.

Again, the emphasis, here, is not on the nature of *R* but on its relational properties it shares to objects in the world. Cummins argues that because such theories must distinguish between correct and incorrect tokenings of a representation, these theories are, in fact, use theories of content- cases of holding that content is constituted by correct use, rather than holding that content is constituted by some intrinsic property of the representation, or the mere fact of covarying with some object. For Cummins, there is always more that the informationalist has to refer to than the facts of covariance.

In contrast, an 'informationist' or 'covariation' theorist might contend that that the correctness of a tokening is a matter of the properties of the connection existing between the token and the object- and not the 'expectations' of the system, or any downstream, post-tokening events. Without dealing with this now, I think it important to point out that a post-tokening role can be separated from a tokening role in principle, at least[22].

---

[22] It is always a temptation for an informational theorist or a role theorist to distinguish the tokening of a representation from its subsequent use/role in a system. For example, Fodor (1994) raises the possibility of dealing with *sense* and *reference*, as properties of representations, by assigning reference to covariation and sense to the post-tokening use. Similarly it may be tempting to distinguish between a *proto* content determined by a useful covariance from a *full* content or an *understood* content that is determined by the post-tokening role of the proto-contentful representation.

On the other hand, as we noted above, there is no reason that a functional role theorist couldn't include both a representation's causal antecedents and consequences in the delineation of a role/use. Dennett, for instance, in his 'Do it yourself understanding' paper (reprinted in his 1998) takes Dretske and Fodor to hanker foolishly after an information / meaning distinction where the later arises as the result of processes in the central systems where understanding 'happens'.

Cummins would agree with Dennett that a (correct) tokening of a representation is just part of a larger use of that representation. He would also argue that it is only possible to know that a tokening of a

Finally, our range of semantic theories includes Cummins' preferred option, namely:

- Resemblance theories of content. These hold that *R's* meaning is what *R* resembles, or the resemblance *R* has with something else. Resemblance can be understood in the straightforward sense of being 'like' something, but it is used to also refer to abstract relationships of isomorphism. In resemblance theories, the emphasis *is* on the intrinsic properties of the vehicle - thus, representation structure comes into play. Connectionism could be taken to be at home here because it deals most naturally with structured representations. I will consider whether Churchland's State Space Semantics belongs here as a resemblance account- but O'Brien and Opie (forthcoming) are clearly resemblance theorists when they say that content lies in the structure of the resting network, more particularly the weight structure of the network, which resembles the task domain[23].

In general, Cummins places great store on the claim that the semantic properties of a resembling representation can be specified without recourse to any notion of 'correct use' or 'correct tokening' of that representation. A representation's resemblance, or resemblances, to the world, and hence its content, or contents, are contained within itself, so to speak. This, argues Cummins, gives him room to move in explaining representational error- something he holds central to a theory of representation. According to Cummins, if we can refer to use separately from content, we can refer to the incorrect use of content without begging any questions as to the nature of that content- and thus explain error.

---

representation is correct when we know what property in the world its tokening is meant to coincide with; Cummins would say, when we know how it is meant to be used- or what it is meant to be used to detect, Moving from the epistemological to the ontological, what makes a tokening correct is, therefore, something about its full role in system- it is determined by what information is 'expected' by the system in the event of a tokening. It follows that informational semantics are determined by post-tokening use, as much as by covariation.

[23] It is possible to argue that a functional role theory of content might be compatible with the notion of resemblance (eg, von Eckardt 1993). The functional relations within one system of brain states might resemble another system in the world. (A dynamical State Space Semantics might fit this description, except that principally, in a state space story, it is physical similarities between representations that resemble similarities between things in the world.)

Cognitive science's chicken and egg problem

Cummins argues that getting the relationship between representation and cognition right makes all the difference when thinking about mind, in general, and the content and form of mental representations, in particular. He suggests that the proper explanatory order in cognitive science is that representation explains cognition, and not the other way around. It is certainly true that if we privilege representation over cognition as our explanatory foundation, then representation becomes the jewel in the explanatory crown. In broad outline, representations are (almost by definition) about stuff: objects and states of affairs etc. Hence, we can explain why thoughts and perceptions are about stuff, as well - providing we can explain how representations get to be about stuff themselves, in the first place.

Moreover, since representations are amenable to taking part in computation, we can explain an organism's intelligent interaction with the world as the result of computation over these things that connect the agent with the world separately from any computation that they enter into. Accordingly, what intrinsic properties representations require in order to be about stuff should be determined before anything else, so that the whole story is anchored by representation.

Of course, this is the opposite of a role theorist like Dennett's view of representation, in which content does not explain the use of the representation, but, rather, the use of a 'Julie Christie' data structure, for example, explains why it is about Julie Christie. As a role theorist about content, Dennett must think the role of a representation is more fundamental than the representation itself. Certainly, on his account, semantic typing is an 'after the fact' matter, a matter of interpretation. As we know, this is not to say Dennett is an irrealist about representation and meaning- more that, on his account, and in contrast to Cummins', 'meanings don't pull levers' or drive cognition; they emerge out of cognition instead.

For Cummins, this is a bad idea: given the centrality of representation in our explanations, it's not smart to be a *use* theorist about representational content. In fact the whole approach is close to question begging. Once you have assumed the existence of a set of roles, you have already helped yourself to cognition, since roles only emerge once problems have been solved; that is, when you already have cognition- which is the thing representation was meant to help you explain in the first place. *Role* cannot explain representation and cognition at the same time. Hence, role theories of content cannot get off the ground because they get the explanatory order back to front. So, if Cummins is right, our lesson in the previous chapter about Turing's notion of computation and syntax also applies to cognition and content- in both cases you have to type your key quality- syntactic or semantic type - before you do anything else:

> Current cognitive science typically seeks to explain cognition by appeal to representation. This explanatory strategy is undermined if representation cannot be defined in a way that makes it independent of cognition. For example, a definition in terms of rationality or inference (Dennett, 1987; Pollock 1989) evidently threatens to undermine the standard explanatory strategy (1996, p3).

> Representation is an explanatory primitive… we do the foundations of [cognitive] science no service… if we define representation in mental or cognitive terms (1996, p4).

So, to continue with Cummins on matters *intrinsic*, if role theories of content are out, then so are accounts of mental representation that posit an arbitrary form for those mental representations- providing all representations of arbitrary form 'achieve' content in virtue of use[24]. (I will question Cummins' acceptance of this below.) Representations must, then, have their content in virtue of their own intrinsic properties- before role or use comes into the picture. And that suggests to Cummins, and others like O'Brien and Opie[25], the idea that content must reside in the structure

---

[24] Obviously, someone like Dennett or Clark and Wheeler would not want to claim that the form of a representation in a intelligent system could be literary arbitrary, but, just as far as the issue of content goes, form is not the issue.

[25] See O'Brien (1999) and O'Brien and Opie (forthcoming).

of a representation. After all, structure is one thing that the world and representations inevitably have in common - something that can be shared to a greater or less degree.

I should mention here, by way of polemical context, that multiple sharings of structure between representations and the world has been taken as a problem for resemblance theories. We may find an unseemly proliferation of content if a representation shares its structure with several objects, and, conversely, every brain state threatens to become a representation of something or other, if shared structure isn't hard to find. In reply, it might be argued that multiple instantiations of particular complex structures are not that easily found in nature[26], but Cummins bites the bullet here and concedes the possibility of multiple contents to representations (1996, p102).[27] This is the price of not isolating a correct usage of the representation to determine it true content- of course, use is out of the picture for Cummins, so there is no way of pinning down a 'canonical' content for any representation. This may not be such a high price if, in general, we shouldn't be put off by the odd puzzling semantic property.[28]

Another traditional objection to resemblance as a ground of content is that resemblance is a symmetrical relationship whereas representation is not- a person doesn't constitute a representation of a photo of themselves, for example. Resemblance advocates of a tripartite account of representation seemingly have an easy out here, namely, that, while the use of a representation by a larger system does not contribute to its resembling content, it does contribute to its status as a representation- which is not the case for the object being represented (see O'Brien and Opie forthcoming).

---

[26] O'Brien in conversation

[27] Millikan (2000) attacks him on this point arguing that an isomorphism between two sets of relata/objects will generally mean many since one mapping between the objects and relata between the two sets will be convertible into another mapping.

[28] Although Cummins doesn't enter into this spirit of generosity, himself, when he seems to make something out of the problem of multiple covariances for informational semantics - of the 'cow' vs. 'ungulate' type. He thinks these threaten the connection between content and *computationally relevant* form for symbols (p 68); how is the right form coupled with the right covariance? The resemblance theorist can argue, here, that multiple contents are less of a problem for resembling representations since they have form to fall back on to guide computation. The question for them, though, is 'how does the system know which form or structure to make use of?'

But leaving traditional objections aside, and to continue with the upside: for Cummins, since structure can be computed with (in neural networks, for example, and other analogue computing devices), content can enter into computation. All that remains then is to show how brain states can share structure with the world and, then, structure can take its place as the rightful bearer of content in computation and the mind. (We will see below that, for some connectionists, structural properties are not straightforwardly intrinsic properties. For other connectionists, perhaps content can actually be found in the physical structure of the network.)

I note a parallel here between distinct but connected philosophical debates in the philosophy of mind and cognitive science. It's interesting to note, as O'Brien and Opie do, that this privileging of representation over cognition corresponds to the position of the old identity theorist *vis a vis* the functionalist in the debate over the metaphysics of mental states.[29] The functionalist claims that the intrinsic properties of mind stuff, i.e. the brain, is only important to the extent that it supports the right kind of relational properties between its various components. Similarly, the role theorist of content and the arbitrary form proponent might argue the same for mental representations. It's how mental states are related to each other, their relational syntax, how they are wired up together, that is crucial- not their intrinsic properties. The identity theorist meanwhile claimed the nature of the stuff, i.e. the brain, is part of the nature of the mental states that make up mind.

In the current debate about mental representations, as opposed to mental states, the resemblance/structuralist theorist is in the position of the identity theorist; the nature of the stuff, in so far as we are talking about structure at least, is crucial to the nature of the representation and hence to the intelligence it supports. Thus, the debate has moved from the nature of phenomenal mental states to the nature of *intelligence supporting* states- whatever the associated phenomenology. The key properties of the brain are taken by the 'identity theorist' of mental content *qua* resemblance to be structural properties.

Resembling structure

Before outlining some replies to Cummins, I should introduce some notions of how structure talk might actually be translated into computational talk, so that we can consider various types of representation with resemblance in mind. Firstly, there can be mappings between a resembling structure and its referent which are more or less complete. That is, there may be a one to one mapping between all or some of the objects and relations across the two structures. If the mapping is complete (i.e., with no bits left over in either structure) then the two structures are said to be isomorphic to each other (Cummins, p96, O'Brien and Opie, p9).

In that case, according to Cummins,

1. An object in R can represent an object in C.
2. A relation in R can represent a relation in C
3. A state of affairs in R- a relation holding of an *n*-tuple of objects- can represent a state of affairs in C.

> Note, however, that R does not represent C because the objects, relations and states of affairs in R represent a relation in C, but the other way around: things in R represent things in C because R is isomorphic to C. … there is no such thing as an unstructured representation, except in the derived sense just introduced (p 96).

This idea that structured representations can bottom out in unstructured representations has already cropped up in our consideration of relational syntax and will feature throughout the rest of the thesis, for example, when we consider whether State Space semantics counts as an implementation of Cummins' thesis. For now, we can note that Cummins is at pains to point out that there are no independently semantic constituents of any representation. Any constituents of a representation only represent in virtue of being part of a larger representing structure. (Cummins (p 97) doesn't think this implies that all representation is holistic but I will argue that State Space Semantics is a holistic representational scheme).

---

[29] In conversation with Gerard O'Brien.

Given this notion of isomorphism in representation, which Cummins refers to as a 'picture' theory of representation, mental images would seem to be the most obvious candidates for mental representations. Understood in an obvious way, images (including mental ones) can be clearly isomorphic to what they represent since they use spatial relations to represent spatial relations. This kind of resemblance, where intrinsic properties are shared between structures or objects, is called first order resemblance. Some cognitive scientists believe that the brain makes use of some straight forward, first order, isomorphisms between itself and the world. Pinker, for example, is tempted to believe that some mental operations such as rotation really might be the rotation of neural activation in the visual cortex (1997, p287).

Most analogical or resemblance or isomorphism talk is not this intuitive because it doesn't make use of first order resemblance . These approaches are more likely to be about the brain using representation-to-representation relationships to represent relationships in the world. That is, relations, usually of some kind of similarity between things in the world (perhaps different versions of the one thing), are represented using relations of similarity in the representing medium, eg, between connectionist activation states or vectors. This is termed second order isomorphism (see O'Brien and Opie forthcoming, p 8).

State Space Semantics runs on second order isomorphism. Those familiar with NETtalk will recall the cluster analysis of the network, which revealed a set of family relations between middle layer representations for phonemes that mirrored linguistic theories about the relationships between phonemes. For example, NETtalk's representations for /p/s were close in their physical structure to /b/s and appropriately distant from its representations of vowels such as /o/ and /i/.[30] Earlier, we saw that Elman thinks his 1990 and 1992 networks similarly capture relative sentential structure in activation paths through activation space.

This all requires that representations have some intrinsic syntax, but not for the same reason the Classicists would argue. Representations need structure because they need

---

[30] (Actually, NETtalk's cluster analysis produced a dendogram, which is not exactly the same thing as a state space diagram. For example, it does not show the dimensionality of a network's state space, just

to differ in the right way from other representations- representations need to have structures that stand out in structure space. The finer their structures the more fine grained the relationships between representations that can be captured. The system of representations will need to be of as fine a grain of structure as the structure it hopes to 'catch' in the world (Speitzer, 1997).

These relationships of second order resemblance are most clearly captured using state spaces, such as an activation space that represents the activation states of the middle layer of a connectionist network. However, the idea is older than that no doubt – it is visible in the Tractatus, for example, where Wittgenstein drew attention to the relationship of resemblance between relata of various kinds- for example, the grooves in musical record contain the same relationships between each other as in the equivalent written music and as between the actual notes (1961, 4.014). In this case, each note or groove in no way resembles a sound, but the whole set of grooves or notes could be taken to abstractly resemble the melody they are related to.

Some broad options for resisting Cummins

I develop these responses to Cummins in subsequent chapters, but introduce them here to give the lay of the land of the debate and the thesis.

Option 1. Information is not use.

This is roughly the idea that intrinsic syntax can mark informational content, but, unlike resembling structure, not be the basis, or ground, of content. This is to have it that content is established before use, since a representation presumably has its intrinsic formal and covariational properties established before it is used by a system. An informational semantics, then, might be consistent with a non-use semantics for non-resembling representations. As we noted above, Cummins argues, instead, that informational semantics is really a *use* theory of semantics, one reliant on more than

---

the distance relations between representations' average positions in state space (Churchland, 1998, p14).)

covariation to determine the content of any representation (and hence ruled out in virtue of getting the explanatory order in cognitive science wrong).

Option 2. Content is emergent and explanatory.

This option represents a denial of Cummins' explanatory framework. Cognition is not to be understood as the use of pre-existing content to solve problems but the use of representational form (and tacit representation) to solve problems such that content emerges, and is, in a different way, part of any solution.[31] A related idea, here, is that states of the brain with indefinite semantics are possible. Dennett's (1998, 'Things about things') 'thing about redheads' provides us with an example. This is an imaginary bit of cognitive machinery that comes into play whenever the topic is redheads,

> and that adjusts various parameters of the cognitive machinery, making flattering hypotheses about redheads less likely to be entertained, or confirmed, making relatively aggressive behavior *vis a vis* redheads closer to implementation that otherwise it would be, and so forth. … the contribution of [this thing] could be perfectly determinate and undeniably contentful and yet no linguification of it could be more than a mnemonic label for its role (p 1, on-line version).

Dennett thinks trying to express a content for this thing in words, i.e., via an act of 'linguification', is just to shift attention away from what matters, namely, what it does and how it does it. The point for us is that what the thing does is contentful, not so much what it is.

According to this view of cognition and representation, within a cognitive system certain objects play certain roles in virtue of their physical qualities, presumably their

---

[31] A related idea is that representation is as much about organisation as information capture. Discrete, re-identifiable brain states may be a source of organisation in the brain. Here, the key insight is that order emerges out of order, at different levels and scales. Consider Simon's view of complex systems-namely, that they depend on lower level regularities and hence modular entities on which more complex orders can supervene[31]. Newel (1991) (another arch classicist) also claims stability in complex systems requires levels and hierarchy. Perhaps, representations are the bottom layer of this pyramid, holding the various layers of order up. To borrow from Dennett (Things about Things, 1998) uniforms, i.e., identifying syntactic properties, can be useful in getting individuals to work together.

form, without that form constituting an explicit, determinant content of sorts- that is, without those forms enjoying semantic properties intrinsically. As such, content wouldn't explain cognition in the same way intrinsic syntax might- and in the way Cummins thinks it must- but it might be possible to argue for another kind of representational explanation of cognition, here, namely, a functional one.

Option 3. The pluralistic option.

Perhaps there are different types of representations and content in the human brain. If there were, this would still defeat Cummins, even if it were shown that resembling representations were more basic in an evolutionary, or developmental, or, even, semantic sense. By a 'semantic' sense, I mean a case where 'use' representations emerge or sit on top of non-use representations and rely on those for their own content. (Though I don't see why covariation is not as fundamentally useful to a brain as resemblance and not easier to build). The story might go that once content comes into the system, perhaps via a form of resemblance, it becomes possible to build on that content, to abstract from that content in some way, to redescribe, as others have put it[32], and get an arbitrary form/ functional role semantics in operation[33]. Maybe a bootstrapping process could defeat Cummins' order of explanation objection, allowing content to play different roles in cognition. In general we can ask, 'Why should the content making properties in a brain all be of the one type?' Is there anything about content that requires a uniformity of basis?

---

[32] Clark and Karmiloff-Smith (1993).
[33] Dietrich and Markman (2003).

Chapter 4. In more detail: two (and a half) responses to Cummins

Response 1: Bringing implicit content to cognition

Our first defence, against Cummins' argument that only resembling representation can explain cognition, is the claim that symbolic accounts of mental representation don't rely on the use of symbols to explain symbolic content. Consider a Fodorian thesis that says meaning = nomic covariance with a referent (perhaps, with some asymmetric dependence thrown in to further nail down which covariances are the meaning determining ones). On the face of things, content determination, here, of representations of arbitrary form, seems separate from how the symbol is used or interpreted- it is just a matter of covariance. Ramsey (1997, p 42) points out that Dretske stressed that the reason such a covarying representation could come to be recruited into a system and play a useful role was precisely because it already had a content of sorts in virtue of the covariance in question. Accordingly, Ramsey, unlike Cummins, depicts indicational content as content that does causal work in a cognitive system.

Perhaps, then, Fodor's causal theory of content could be seen to leave him on the side of the 'identity theorists' of representation, or the intrinsicists- at least, in the sense that he eschews any kind of functionalism about representational content[34]. Both the 'actions' of resembling and covarying do not depend on down stream processing- they seem a matter primarily between a representation and its referent.

Apart from the problem of explaining error in 'covariational' content, the main task for this view of content possession is to explain how the content of such representations gets to do causal work. On the face of things, Cummins is in a better position to make use of the 'representation $\leftrightarrow$ referent' relationship as a causal player since, on his resemblance account, its basis, namely, shared structure, is better suited to involving content in computation. The structure of a resembling representation

---

[34] O'Brien in conversation.

more obviously, or explicitly, 'contains' content compared to the non-resembling intrinsic properties of a symbol, which provide no obvious hint of the covariance the symbol enjoys with its referent. O'Brien and Opie put it that resembling structure meets what they call the 'causal constraint', namely, that (a theory of) content should be consistent with its content's causal role in shaping appropriate behaviour (Forthcoming, p2).

But, doesn't Dretske have a point? Covarying with something important in the world is a useful property for a representation to bring to a brain. We can use the fact that a representation has that property to explain the post-tokening role it has. In other words, we can explain part of cognition in virtue of the properties, admittedly relational, of the representation. The appearance of that representation, i.e., that form of brain state, will signal the presence of its referent object or property. This seems to be a matter of content determining what happens next in cognition, not the other way around.

Cummins, however, thinks this is not to understand causal theories, or informational semantics as it is applied to representation in general. According to Cummins,

> causal theorists don't think representing is carrying information- even though causal theories derive a lot of their plausibility from the idea that the basis case of representing is carrying information (p 65).

This critique of informational semantics alerts us to the fact that causal theorists, particularly if they are telling a language of thought type story, accept that tokens of representations are not always used in actual indication, but rather to refer to the property or object they might have originally indicated, or otherwise happen to covary with. Cummins points out that, on that kind of story, most tokenings of a representation might not carry information about the state of the world:

> … as Fodor (1990b) has emphasized, most uses of representations are not indicator uses (what Fodor calls labelling uses) anyway. The occurrence of a /cat/ in /If cats were less selfish, they would make better pets/ does not even have the function of carrying information about cats. Thus it is that causal theorists quickly concede that

even correct representations of cats do not typically carry any information about cats at all (1996, p65).

However, as a response to Cummins, Response1 is an attempt to argue that covariation nonetheless fixes *a meaning* to a representation, a meaning that copies of that representation enjoy even though they may be tokened otherwise than in an act of covariance. It is not to say that indication fixes information *about* a referent to a representation, just that it 'captures' an important property in the world and assigns it to that form of brain state. Cummins' response to this is that in…

saying that a symbol's role in detection simply fixes its meaning… one abandons altogether the idea that representation has anything to do with carrying information; one just adopts a trick that will associate the right meanings with the right symbols without telling us what representation is. The plain fact is that if you like the kind of combinatorics that symbolic schemes give you, informationland isn't where you should be looking for meaning (p 65).

Cummins, then, is not impressed by the semantic properties of classical representations. At the most, reference isn't a very useful kind of content, since it doesn't tell you much about what is being referred to, except perhaps its 'name'. At the least, reference isn't informational in any sense. In other words, as far as informational content goes, unless a representation is actually indicating- it doesn't have any semantic properties. A covarying representation, in the act of covarying, cannot keep any covarying content for later use, as it were, or share it with other versions of itself.

But, why not? The fact that a representation is the same type as other actual indicating representations seems to be a useful property for a brain to 'know' about. That way, the brain can use the same processing tools it uses for actual indications to consider hypothetical indications or for inference etc. This is meant to be one of the advantages of Barsalou's (1999, see below) perceptual symbols, namely, that perception slides efficiently into imagination.

I'm asking why representation has to be informational in the immediate sense of being 'hot', as it were? I'm suggesting a scenario where indication slides into reference as it *cools down*, so to speak. Cummins thinks that reference is not representation because it is not informational and so will not help us understand cognition 'in the way a theory of cognition is supposed to' (1996, p93-94). But, I argue that there may be a case to be made that reference *qua* cold indication is informational, in a *second hand* way. Reference may not be informational about the here and now, but it may be implicitly informational about what sorts of things there are in the world, or at least the sorts of things the whole cognitive system has experience dealing with, and in particular, about which thing in the world the system is 'contemplating' now. A cold indicator, or covarying representation, may represent the property that hot versions of it indicates in virtue of being the same shape as its hot cousins. This content would be implicitly encoded, rather than explicitly encoded as a resemblance.

How this happens needs to be explained, of course. Of course, the brain cannot take a symbol and, by its form, 'judge' what thing or property it, or an antecedent version of it, covaries with. But brains never 'read' content like that in any case, not even of resembling representations; they don't literally *infer* from representation to world, or represent the world again on the basis of a first representation. They just *use* representations to solve problems. The question, then, is, 'Can a brain use non-resembling representations to solve problems, and, in virtue of the content of the representation?'

The answer here may be that a) the brain uses indicators in virtue of their co-varying properties- it uses the covariation without actually touching it, i.e., not in the way it may come into physical contact with resembling structure- it just 'assumes' the covariation is with a particular property/object in its workings; and b) the brain uses down stream versions of these indicators in virtue of an implicit coding of that covariation in their form. Somehow, due to learning or evolution, the brain uses the form of those representations as if it were reading a symbolic language. The cold symbols refer to the hot versions of themselves, and so the brain reads the cold symbols not as indicating a property in the world right now, but as referring to the state of affairs which hold when hot versions of the representation are tokened.

It is true, as Cummins points out, that arbitrary mental symbols cannot carry much information about the properties they detect or stand in for by themselves. (Indeed, it may be that this is part of the reason why this type of mental representation, i.e., *information lite* representation, is able to give you the kind of combinatorics a formal system offers[35]. In any case, no-one doubts that information can't be conveyed digitally.) Cummins claims that mental symbols are most likely to function as mere triggers for procedures, mere cues for stored knowledge and as constituents of complex representations (1996, p70). But, according to the argument, here, this might be to overlook a kind of property capture that such indicating/labelling representations might contribute to cognition. In any case, it still doesn't follow that those other roles in cognition that Cummins refers too are not useful or explanatory.

Being a trigger for a stored procedure[36], being a cue for stored knowledge, and being a constituent of complex representations seem pretty important roles. Cummins would presumably argue that they are parasitic on the content that other (resembling) representations bring into a system. However, on the one hand, I'm not sure that fact, if true, makes those roles representationally impotent, so to speak, and on the other, that those roles wouldn't make more sense if the representations in question didn't carry some indicational or labelling content themselves, as I've argued is possible. Surely it would be a useful thing that the brain's triggers, cues and combinatorial constituents maintain some connection to the world, namely, via their form. We could think of their form as a kind of memory of what they refer to. This may actually help to keep the world and the brain's activities in synch, just as Cummins hopes resembling structure promises to do. However, whereas, the problem with resembling structures may be that they carry too much information with them, *qua* too much form, to play certain roles in cognition, this may not be the case for non-resembling representations with their content attached more implicitly. (I develop these themes in subsequent chapters.)

---

[35] As Fodor (1998) argues, concept combination doesn't seem to involve the combination of much structural baggage

[36] …which itself might be composed of other triggers for other stored procedures/triggers, see Newel 1990

O'Brien and Opie would no doubt object that this notion of an implicit encoding of content in the form of a representation doesn't meet their causality constraint- that (the basis of) content be consistent with a representation's causal role in shaping appropriate behaviour. The shape on an indicating state or name may help record a covarying relationship between a representation and a referent but it won't reflect the 'world' part of the covarying relationship, i.e., the referent, in any way. Consequently, the covarying relationship- what O'Brien and Opie call the grounding relation of the mental content in question- won't explain how the representation contributes to the production of intelligent behaviour (forthcoming, p 4).

My alternative view of the situation is that the grounding relation as implicitly encoded in the representation might affect the use of the representation. The representation's intrinsic properties will be causally linked to the covariation - the basis of its content- in that the covariance relation has come to be recorded by that particular shaped indicator over time.

I think Cummins would say this historical connection between the form of the representation and its content is not direct enough to get it off the charge of relying on use for content. Consider this quote from Cummins:

> Causal theories are use theories … for they identify the content of r with its target when it is applied by a successful *detector* [my emphasis]. (1996, p29)

One thing Cummins wants us to remember, here, is that there is no indication or reference without a *detector* or a user of an indicating state. Indication is dependent on there being detectors – and so indicating states, in order to be informational, are dependent on their use by a particular kind of detector in particular conditions. Indication, in the context of a detector, is really a kind of application or 'judgement' about the world. In other words, it's the relational properties of an indicating state to the 'judgement' machinery that matter. To use Cummins' example, indicators are like ticks in boxes on questionnaires. A tick has no meaning in itself, and so is not a representation.

Since, on the cold indication story I'm telling here, an indicating state is unlikely to carry it's intender or relational properties around with it everywhere it goes, we might be in trouble if we think indicating states can get their meaning in the act of indicating and then trot off with it in their pockets.

My first reply is that in one sense, any state only has content, in virtue of being part of a cognitive system- even point to point reproductions of the world require a user system to count as representations. Such representations and ticks are at opposite ends of the representational spectrum, but the same spectrum. So why should resembling be any more representational than covarying when they are both equally contentless outside an appropriate computational context?

My second, more serious, reply is that Cummins seems to assume that indicating states are like ticks – that is, they are representations which have no defining features. However, surely we can imagine indicating or covarying states with identifying intrinsic qualities. Moreover, we can imagine that other representation-using machinery might recognise these individuating properties. In that case, it would make sense to say they carried their reference-type content in their form and didn't require their 'home' intender machinery or 'home' relational properties to give them content.[37]

The misalignment problem

Even if the assignment of content to an indicator does originally depend on relational properties like what intender a representation is originally tokened in (which I'm not sure is necessarily the case), there may be no reason why that should stop an indicating state taking on that content in a more permanent way. Cummins points out doubtfully that there would need to be some process by which the right roles come to be assigned to copies of the right indicating states. (In this case a representation's role would reflect its content, but I don't see why it should be understood to determine that content.)

---

[37] Perhaps, this is more easily imagined if we think of the sort of abstract feature detectors Markman and Deitrich have in mind, rather than lower level indicators that have a more limited role to play.

Cummins seems to think good matchings between role and content are unlikely, in any case. As we saw above, one reason is that arbitrary form representations don't give away any clues as to what they covary with. Cummins puts it that

> the structure that is crucial to cognitive explanation [the structure that disciplines representations and allows them to take part in inference, for example] is extrinsic to symbols but internal to things like maps (1996, p93).

That is, the intrinsic properties of a map give a system much more of a head start in achieving intelligent behaviour than the intrinsic properties of a symbol. (In the previous chapter, I claimed the story was different for symbols in concatenation.) This is why Cummins thinks reference, as well as indication, isn't representation – it doesn't represent much about the thing being represented. In contrast, if the structure of a representation itself contains information, there is less chance of misuse of the representation. For example, /cow/ /ungulate/ or /apparent cow/ may all get tokened in the presence of cows- appropriate use of these arbitrarily different symbols (so that their content tracks their form, as Cummins puts it) is going to depend on some pretty good software. Moreover, if the use of a representation doesn't match the content it seems to have as an indicator, then content wouldn't seem to be of any explanatory use (1996, p68).

Perhaps the information semanticist can find an unlikely ally here in Dennett. Dennett has more faith in natural design processes than Cummins, he isn't worried about what he calls the problem of pre–established harmonies between a representation's meaning and its causal role, since the two are the same thing for him. (Not that he would rule out nature making use of resemblance). I quote him now to shore up the defences against the misalignment problem:

> it is no accident that events with the meanings they have get to play the causal roles they play (and Dretske in fact gives us a good account of this), but the other side of the coin is that the odds are astronomical against the occurrence of an event or structure that lacked the relevant meaning somehow arising to cause a bogus B-type

[downstream processing] event- at least more than once or twice. It is not just that As cause Bs, but that if As cease to mean what they do, they will (shortly) cease to cause Bs, thanks to the power of an adaptive (or learning) system to select against those of its structures that prove inappropriate. That is a tight a relationship as one can hope for, and it is tight enough to explain the admirable if imperfect human capacity for being responsive to content (1998, P77 'do it yourself understanding').

Although Dennett is described as a use theorist of content, he seems, here, to be almost on Fodor's side: he could be read as saying that it is no accident that *events with the meanings they have*, that is, indicating events, play the *non-meaning making* but *meaning respecting* roles they do. In any case, perhaps the moral, here, is that we shouldn't presume to know the limits of nature's engineering capacity and so hastily rule out some representational formats as unusable. We shouldn't assume that the connection between a representation's covarying properties and its syntactic properties (only implicitly contentful) may yet get to be co-ordinated by evolution, if concepts are innate, or by learning.

My other response to Cummins' misalignment problem between role and content is to say that such fine grained content discrimination, as in 'cow' versus 'ungulate', is too much to ask from a theory of content from the start. Or that failing to account for it, isn't reason to give up on an approach alone. Fodor (1994)makes this move in deciding against abandoning informational semantics in the face of such rare anomalous cases. Moreover, I'm not sure a causal theory of content need attempt to explain this fine-grained sort of content determination alone in any case- it needn't be the only explanation of the semantic properties found in any cognitive system. And, in any case, we can ask how a resemblance story is going to do any better with fine-grained content? Perhaps functional role does come into the game when we get to his sort of content (as per the pluralistic option below).

Out of interest, we can note that both Cummins and Fodor have uses for 'use'- apart from fixing representational content. In The Elm and Expert, Fodor distinguishes between representational *content* as determined by covariance and representational s*ense*, or narrow content, as determined by use. Fodor does so in order to deal with

'Fregean' cases such as the 'morning/evening star' problem where different instances of the same covariation seem to have the same referent but different senses. Cummins opts for a 'use' component of meaning as well – this is what he calls 'meaningfor'[38], as opposed to representational content, or simple meaning (1996, p87). The meaningfor of a representation to a system is what knowledge that representation activates or is associated with.

As we know, Cummins claims we cannot take such a functional approach to representational content, or simple meaning, because this will leave no difference between a representation's content and the 'content' of its use, eg, its targeting - we need both to explain error, as a mismatch between the two. In fact, a main motivation for positing this content dualism is to find room for the possibility of misrepresentation or error. We can note that Fodor wouldn't want to use post-tokening *use* to explain error since he's already using it for Fregean cases and anyway there remains the possibility of using tokening to explain error by privileging some tokenings over others- i.e. as in the asymmetric dependence account of error.[39]

Finally, there is one other point to make about the semantic powers of the syntactic structure of arbitrary representations. We have been arguing about the nature of simple name-like representations, so far. However, in the case of complex representations, there may be another sense in which their form contains their content. The structure of something like a quantified statement in predicate logic reflects the structure of its propositional content- quantifier, predicate and term are discrete within the one representation and perhaps resemble the world's division into objects and properties. The point of the arguments in the previous chapter was, similarly, that explicit representing structure in complex representation can be a useful tool for a cognitive system.

---

[38] This is distinct from application content, that is, the result of representational content being ascribed to a target.

[39] I'm going to keep out of the error debate except to say, I don't think it should lie at the heart of your theory of everything. Since this is a low road friendly thesis, error, is also not so high on its list of 'to explains'.

Millikan (2000) is puzzled why Cummins doesn't accept this kind of structure as representation (as per Wittgenstein's picture theory of language). Cummins treats language itself as non-representational – for one thing language doesn't carry the map-like detailed structure-based content the Cummins wants pushing cognition around. However, this 'discreteness within complexity' that we find in language is why information semantics goes so well with language like combination. Since indication, compared to picturing, doesn't require much structure, its products are more easily combined *in structure* since there is less structure to combine. Ironically, this ease in forming complex and new structures may give cognitive systems the sort of computational advantages that Cummins reserves for more obviously resembling representations

Response 2: Content and functional explanation

In this section, I attempt not so much to argue for role theories of content but to support their claim to explanatory relevance, or respectability. (This is in part an insurance policy against failing to show that informational semantics is not a use theory of representation.) Suppose it is the case that content *emerges* out of cognition rather than being its raw material as Cummins thinks it must be if it is to count as content at all. How could emergent content be explanatory? Well, perhaps we should look to syntactic properties in order to understand the nuts and bolts of cognition and to semantic properties to understand the functional organisation of cognition.

Ramsey (1997) identifies two approaches to positing representations in any account of cognition. One is more or less Cummins' own approach to using representation in cognitive explanation: look for causally active content such that states of the system which represent give the system 'extra powers' in virtue of the intentional properties the representations bring into play. This is certainly compatible with content being computationally powerful, resembling structure.

Ramsey's second approach is close to our Dennettian defence of representational talk, such as talk of a 'thing about redheads'. This approach is to agree that the intentional properties of representational states are causally inert but note that treating a cognitive

system as a representation system allows an understanding of how the system is organised (1997, p 37). (I explain this in more detail below.) This is to deny Cummins' explanatory order for content and cognition whilst trying to keep representation in the game. At one point, Cummins (p 97), himself, states that 'use can give us evidence of content, but it cannot be constitutive of it' So perhaps Cummins would agree that this approach is useful, but disagree that it justifies representation talk as descriptive[40].

The identification of representations can make sense of the functional structure of a cognitive system as follows: a system is broken up into subsystems, certain (representational) structures are viewed as inputs and outputs to these subsystems in a way which makes sense of these subsystems. For example, we might want to posit some part of the human brain that calculates niceness. To make this claim, we might point to inputs to this putative niceness calculator that can be understood as representing aspects of a person's behaviour that are germane to their niceness, eg, signs of their ability to listen, or sensitivity to the frequency and geometry of their smile etc. We might also point to neural pathways to our module from relevant areas of the sensory cortex.

Thus, we can answer the question, 'What makes you think that is a niceness calculator*?*' with the answer, 'Well, it seems to receive such and such kind of information'. We might then try to ascertain the destination of its outputs and whether consumers of these outputs could, in turn, be thought of usefully as consumers of information about niceness.

Decomposition doesn't have to stop there. Within our niceness calculator we might even try to propose a smile identifier and treat its inputs and outputs as representational states germane to its function, i.e., as carrying information about the geometry of a mouth. So too, for the processing machinery within these calculators: the software of the system could be understood as either explicitly or implicitly coded- but in separate task related chunks. According to this way of modelling cognition,

---

[40] Cummins is also pretty clear that he thinks treating content as use makes content talk trivial (p 41).

> if you want to know where a specific element of a system's 'know-how' is encoded (in order to change the system's behaviour, for example), it is often possible to locate it in these distinct data structures (Ramsey, ibid, p40).

In response to Ramsey, Von Eckardt (2003) claims that it doesn't follow that the internal workings of a something we might identify as computational, such as a niceness calculator, must be representational. Certainly not just because its inputs are representational and the operation of the system is coherent, or makes sense to us. What Von Eckardt terms 'minimal representation' may be in operation here, namely, representation with formal properties only. So, on her account, a niceness calculator may not actually run on representations even if it consumes and outputs representations (p 435). But consistent with the role of representation in aiding functional delineation, we still seem to have a niceness calculator on our hands.

Having described the motivation for this kind of representation talk, I turn now to Dennett's approach to the role of content in cognition, in a little more detail. Dennett, like Cummins, thinks that covariance alone is not enough to get you content. In a covariance case, such as our direction of movement indicators, Dennett would think of content as supervening over both the tokening and (designed) post tokening use of such indicators. Take Dennett's (1987) 'two bitser' thought experiment. Here a state in a vending machine flags correct coinage by covarying with correct coinage. At least Fodor might see it that way, if he thought such a machine were capable of intentionality. Dennett would add that covarying with an intended object is part of the role of the flagging state- as is serving as a trigger for the release of a soft drink from the machine. (Being of arbitrary form, Cummins would say the coin indicating state has no content at all.)

The thought experiment continues that the machine is moved to a new country (Panama-'the poor man's twin earth') where another type of coin activates our indicating state- a coin identical in size and shape to the coin that the designers of the machine had in mind. Our indicator state now seems to be either in error or to have a new content due to a new covariance/role- even if the state continues to have the same causal effects: namely, the release of a soft drink. One of Dennett's morals, here, is

about gradual change in meaning (as opposed to determinate change) as the role of a representation changes over time. Whether the state (or machine, as Cummins would have it) is in error will depend on the role of the machine in its new country, which in turn will be a matter of how long the machine has been there and the intentions of the owner.

Dennett sees the new situation as role change- not just as covariance change. The key point, here, is that, on Dennett's account, content doesn't drive the workings of the system since it depends on matters external to the system, such as the attitude of the owners of the machine, in this case. If they are happy to swap soft drinks for the new coin, then the new coin becomes the content of the flagging state.

If we recall Cummins' worries about the misalignment problem, I suppose Dennett's though experiment is a case of the misalignment problem. The causal role of the indicator state seems to have come apart from its indicating content. But what Dennett claimed above would happen in nature, happens also in human affairs: if the owners of the machine were not happy with the coinage they were receiving for their cans of drink, the days of that indicating state, and that alignment, would be numbered. Moreover, until the machine comes to have some role or other, the content of the indicating state remains indeterminate or non-existent, so that one might argue that there would be no content or roles to even be in misalignment with each other.

Thus, on Dennett's account, content is required to explain the workings of the two bitser; in particular, why the indicating state exists within the machine or what it is for- *vis a vis* the interests of the owners. The indicating state exists because of the role it plays- which is another way of saying that it exists because of its content, in a sense. On the other hand, in this account, nothing about the indicator's intrinsic properties really need to be specified to explain its content, in direct contrast to Cummins' position. Dennett would presumably argue that even if the indicating state were structured to resemble the appropriate coinage (indeed, some slot in the machine may have just this feature) that resemblance still would not be the source of its content, nor would reference to it add to the kind of explanation our intentional explanation of the machine offers.

Consider Dennett's approach to Millikan's intentional icons, from above. Dennett (1993) claims that although Millikan looks like a 'mappist', or an intrinsicist of sorts, she is still a use theorist. As we saw, Millikan reserves the term 'representation' proper to what she calls intentional icons (1993, p103). By intentional icon she means a structure that varies as its referent varies and is capable of 'combining' with other representations to deliver inference. Millikan's intentional icons, then, seem, at first glance, to have intrinsic properties that determine their content. Icons adapt an organism to a particular mapped feature and that feature is its content. Presumably, maps map in virtue of their structure- so their structure looks like being the basis of their content. Yet, Dennett disagrees about this take on Millikan,

> Since the content... depends as much on the user part of the system as its 'producer', no item gets content independently of its role within a larger system. For instance, [Millikan] notes that 'A mental name exists only when its thinker has a competence to reiterate it in a variety of different correct and grounded whole representations...'(p 119) (1993,Dennett and His Critics, p223)[41]

Of course, Cummins might say the same about mental names, if he thought they were contentful, but Dennett says it about mental maps as well:

> The individual structured items may be picked out 'atomistically' but whether they are intentional icons, and if so, what content they carry, is still a holistic matter (1993, p223).

It looks like Millikan agrees. Her (2000) criticism of Cummins is that isomorphisms tend to multiply- one isomorphism tends to mean variations on that isomorphism are possible. To use a simple example, any picture is isomorphic to the mirror image of what it is a picture of. Consequently, the actual representational content of any isomorphic structure will depend on its use by an intender or consumer system; since isomorphisms cannot be kept down to one, the 'intention' of the consumer *vis a vis* the representation will have to come into play. Millikan also points out that we cannot just assume an intender or consumer system will make the 'right' use of the right

---

[41] Clark and Karmiloff-Smith use Evan's generality constraint in similar fashion.

isomorphism. Take the case of a simple map- we can't assume that a user will implement a mapping from representation to world that has things on the left in the map also on the left in the world. The opposite isomorphism, where things on the left in the map are on the right in the world, is also available for use. As a response, Cummins (2000) is prepared to accept multiple contents for the one representation - and stress that a user is 'free' to make what it can out of what isomorphisims/contents the representation contains.

In contrast, for Dennett, any structure (resembling or otherwise) that a representation enjoys cannot constitute that representation's content since content is a 'holistic matter'- it depends on matters external to the representation. Of course, having to take into account these externalities to pin down content rules out content explaining cognition in the kind of way Cummins and O'Brien and Opie want.

The trick, here, in making sense of this standoff may be in distinguishing between interpretation/use/holism as a necessary requirement of the presence of a representing system, at all, and interpretation as a necessary component of content. If you are a generous soul and inclined to believe that the unpublished writer is still a writer, then you will have no trouble imagining uninterpreted content as analogous to unpublished writing. You might also think that interpretation will create a further content, as Cummins own theory allows and requires, namely, the application content he talks of. You might take this to be analogous to the creation of a *published writer* and *published writer's cont*en*t*.

What has Dennett got against such a two-tiered account of content? Just that it's only at the second tier that content is determined – and comes into existence. This is to contradict Response 1 above, as well as Cummins. In his favour, he still has what he takes to be non-semantic properties of representations in his armoury that Fodor and Cummins appeal to- namely, covariation and resemblance. But for Dennett, they don't carry content into battle, they just make a contentful role possible. Dennett's representations can do as much causal work as anyone's, but their contents can't really take part in the sort of mechanical explanation that Bechtal was referring to, even if the representations, themselves, can.

Arguing as I am for the explanatory power of syntactic notions, I am inclined to accept that Dennett can claim to be explaining cognition. In any case, I'm hoping I can take a neutral position between Fodor and Dennett and still meet Cummins' arguments against non-resembling representations not explaining content. In other words, Fodor and Dennett seem to have all the options covered. Either content can be 'pre-use' for non-resembling as well as resembling representations, or it is still assignable to particular states of a system in a way that makes sense of the properties of those states and the role they play. Either way, non-resembling representation is possible.

I finish with two problems for Dennett that Cummins raises against Millikan. Firstly, surely we can imagine representational structure coming to be used better over time. Cummins (2000) asks us to imagine a situation where distances on a map come to be read off the map when previously only relative positions were. Surely that content, or information, about distances was there all along (p 115-125)? Isn't this the kind of information extraction that connectionist learning, for example, is all about? Secondly, couldn't a map be misused by a system in such a way that content was not made use of at all. Doesn't this make sense, as well?

I suppose Dennett's reply is going to be that, in these cases, the content was not there all along waiting to be better used or misused, but that the potential for it to be used, and thus created, was there in the representation, all along - precisely because of the representation's resembling structure. There was always the potential for content to emerge out of the representation's better use, or for it to disappear after the representation's misuse. (What we should call that potential, if not content, I'm not sure). This all goes to show, that a representation's syntactic and co-varying properties are not arbitrary with respect to content in the system- without the right properties, a representation will not contain the potential for particular contents to emerge.

When we ask Dennett how is he using content to explain cognition, he has recourse to claims about the functional properties of the system being real properties, along with his familiar moves about the predictive usefulness of the intentional stance. These can be backed up with claims about being a semi-realist anyway about semantic

properties, along the lines that 'content is in the system like centres of gravity are in objects', etc. Intrinsic properties may do all the work in a system but we can pick out the right syntactic and covarying properties and the right roles with the aid of content ascriptions. (Perhaps Dennett could appeal to Fodor's (1974) defence of the special sciences.)

To be honest, I have some sympathy with the 'really' realists about content in cognition and their hope of explaining how it pulls levers – and it does seem as if a resembling relationship in particular can be used to show how content (*qua* structure) is used to produce intelligent behaviour – furthermore, it does seem ungenerous to not call that thing content in virtue of it being pre interpretation (O'Brien in conversation). On the other hand, as I will argue in the following chapters, the notion of 'content as resembling structure' seems limited at the moment in explaining a range of putative contents and, in this way, leaves interpretation and role still in the picture as determiners of content.

Response 2½: The co-determination of content and use

I include here a further hybrid option, or 'take', on the relationships between form, content, role and cognition. It combines some of the moves made above in responses 1 and 2 to argue that representation and cognition may shape each other over time in a way that makes 'information content' and 'role content' interrelated. I am imagining a process that matches the form of a covarying representation with its covarying and post covarying roles.

The story goes that the right covarying shapes (of representations), *vis a vis* certain roles in cognition, could come to covary with the right objects via a process of mutual realignment. Firstly, the covarying content of representations may come to partly shape cognition because the post-covarying roles of those representations will evolve to fit that content. Were not this or that property in the world being potentially signalled by a covarier, certain roles in cognition that make use of this fact would not

come to exist. This would partly meet Cummins' requirement that content of a representation should explain cognition. A representation's indicational content wouldn't explain cognition as it was happening, but it would partly explain why the system was making a particular use of an covarier, namely, as a use that made the *most successful use* of a co-varying content. Secondly, and at the same time, it might also be possible to say that the use of the covarier might partly explain the covarier's intrinsic and perhaps even its indicational properties, namely, as having best suited that particular useful role.

To clarify: it is possible to understand teleo-informational semantics in this way such that content gets spread over proper role, as well as indication. Via evolution and learning, the covariation facts could come to shape the role facts over time, as certain uses of the covarying representation contribute more to genetic success than others. And, conversely, the success of the role would prolong the covariation itself, and, also, come to shape the shape of the covarier, as certain versions of it did better than others *at being made use of*. On this story, we would be stuck with two sources of (non-resembling) representational content- namely, role and covariation- and a bi-directional causal connection between representation and cognition.

I raise these considerations to cast some doubt on the clean divide between *use* content and *non-use* content on which Cummins bases his account of cognition.

Chapter 5. Abstract representation

In the following chapters, I point to some considerations that seems to conflict with Cummins' position that only resembling representations can explain cognition[42]. This is not to deny that resembling representations might be useful or likely, but it is to suggest that the same holds for non-resembling representations. I admit that Cummins, himself, thinks indicators, for example, are a key component of cognition (1996, p74). And, that he merely doubts that indicating states could be representational. As I argued above, I think several cases can be made that non-resembling covarying states can be representational[43]. In general, I think the low road approach to representation supports indicator type states as representations. I also attempt to show, here, that there are some representing jobs in cognition, including indication, which resembling representations seem unlikely or unable to do; in particular, I argue this applies to a putative connectionist implementation of resemblance representation, namely, State Space Semantics. I begin with the role of representing abstract properties.

Capturing abstract properties

As I pointed out in chapter 1, our ability to generalise and recognise different versions of the same type of thing is one reason to believe in representation. There are at least two ways to depict this connection between representation and generalisation. Either, as Fodor (1998) believes, so that there is no intervening psychology between the abstract property in the world and the representation that 'captures' that property, or, alternatively, so that the brain contains something like a hierarchy of representations, as Markman and Dietrich think, such that lower level feature detectors feed into higher level feature detectors. On this story, abstract properties are detected via detection of their less abstract properties, or more complex properties by their more simple constituents.

---

[42] I don't plan to focus on expressive inadequacies of pictures or maps associated with quantifiers and logical connectives. Newel's (1990) discussion of annotated models suggests a way of combining resembling and non-resembling representations, in order to combine the benefits of both formats.
[43] This stands even if it is true, as Cummins claims, that indicating an X doesn't always require representing an X.

Cases of connectionist pattern extraction, such as in Cottrell's face recognition network (see Churchland 1995, p38-55) and Elman's parsing networks, present us with perhaps a third depiction of generalisation. Churchland or Elman would argue that the divisions of a successful network's activation space, which we find using tools like cluster analysis, are evidence of the network generalising over inputs to form concepts. Churchland depicts the divisions of a state space as evidence of a network's 'portrayal' of its world- that is, its background conceptual division of the world into parts (Forthcoming, section viii). So, are these sorts of concepts mediated by other representations?

There are several options here. Depending on how you see the weight structure out of which such an activation space is built, you might see this abstraction as either brute causal or mediated by representation. If a network's weight structure were taken to *not* harbour representations (eg, as Ramsey (1997) argues), then the state space divisions would be brute causal - not in the sense of being unlearnt but in the sense of not being mediated by other representation. If weight structures, on the other hand, are representational, then it would seem that state spaces are generated by superposed representations - which is presumably not what Markman and Dietrich had in mind, anyway. That is, there is no hierarchy of compositional representation from the particular to the abstract in play between a network's weight structure and its state space divisions.

On the other hand, we might focus on the activation states themselves and consider their relation to their components- the individual nodes. Churchland points out, in relation to these matters, that the middle layer nodes in Cottrell's face recognition networks, of which there are 80, have no easily described preferred stimuli in the input array the network receives. They are, in one sense, highly complex and span the network's 4096 input node array. Churchland goes on to argue that

> it should come as no surprise that networks can and regularly do have 'complex' concepts without their having classical simples as their compositional constituents. For on our view, all concepts are complex. All of them have an intricate internal

structure, a structure with a dimensionality equal to the number of neurons in the space that embeds them (section vii).

So, here, is another way of understanding conceptual structure as representational structure- on the one hand, there is the distance relations between concepts/representations that are said to mirror relations between properties in the world, which Churchland describes as making up an 'abstract picture' (section vii), and, on the other hand, there is the internal structure within each point in activation space, which cannot be taken as implementing any simple account of conceptual structure. In this chapter, and the remainder of this thesis, I will consider whether a State Space Semantic understanding of generalisation, based on the abstract picture account of state space, offers an implementation of Cummins' resemblance approach to representation.

Returning to the general connection between resemblance and abstraction, on Fodor's account of abstract feature detection, there doesn't seem much role for resemblance to play- it is hard to imagine the brain just 'coming up' 'brute causally', as he puts it, with resembling representations of abstract content. On Markman and Dietrichs' 'hierarchy of detection' account, there does not obviously seem a role for resemblance to play. Here, the semantics of higher order representations would seem to be a matter of those representation's relations to lower order feature representations, the indicator semantics of those, and/or the more abstract representations' (mediated) indicator relationship to their referent properties. If these semantics are not indicator semantics, they are role semantics. There may be a case to be made for role semantics as resemblance semantics, but the resemblance in question would not be structural as Cummins wants it to be, but functional.

So, at first glance, resemblance might seem a non-starter for higher-order abstract concepts as a matter of course; in any case, how, we might wonder, could something as seemingly formless as democracy, whatever that is, be represented via a resemblance?[44] One possible answer to this is that the thing in the world that is

---

[44]Not that 'democracy' is the type of concept, you might argue, a theory of representation should necessarily spend much time on, ahead of getting some simpler representation under control (O'Brien, in conversation).

democracy does have a structure of some kind and that structure could be somehow realised in the brain. Another possible answer is that if conceptual schemes are holistic in the Quinean sense of gaining their content from their relations with other concepts, then a mind's whole web of concepts *qua* representations might resemble a web of properties in the world, somehow. This might apply to the set of component concepts of the concept 'democracy', or to the web of abstract concepts that includes the concept 'democracy'.

Conceptual webs and semantic holism are a natural philosophical basis for State Space Semantics. Churchland, for instance, marries a holistic view of concepts with state space divisions arguing that we only get representation when we have families of representations, when we have 'second order isomorphism between an entire family of concepts and the entire range of objects or features that they represent, as opposed to any 'first order' isomorphism between concepts and objects taken singly.' (Shepard ((1980) as quoted by Churchland, forthcoming).[45]

However, whatever the family requirements of representation, I argue that the moral for representation from generalisation or recognition is overall one about non-resemblance. My reasons are, roughly, that abstract representations are useful to a mind, not least, because of their 'decontextualised' nature, and as such are likely to be less structured, and less 'en-structured'. That is, 'a-contextual' representations, which contribute the same content across different representational contexts, might be easier for a system to do important tasks with, such as combining them to build more complex representations, or using them to track particulars across contexts- in particular, using them to represent properties out of the context in which the system first learnt to represent them. Admittedly this is less about the process of abstracting and more about downstream use of those abstractions, but the abstraction process presumably must produce usable representations. Of course, 'less structured' doesn't necessarily mean 'less resembling', but I hope to point to some reasons to think that it does.

---

[45] I'm not sure Cummins would agree here. First order resemblance is fine if it can be realised.

My second reason is that second order resemblance, at least in its State Space Semantics guise, which is the most prominent candidate for resemblance in the brain outside of the possibly first order resemblance of mental imagery, doesn't cut the mustard as a component in mechanical explanation. In particular, I claim that second order resemblance is akin to relational syntax – it doesn't deliver constituent structure *qua* effective representation.

To give a flavour of the general case here, when Barsalou (1999) argues that cognition and representation run on the same kind of representations- on what he calls 'perceptual symbols'- he admits the cognitive versions of these may exist in a slightly schematic form. I attempt to show, here, that this acceptance of 'slightly schematic form' is the thin edge of the wedge to be used against resemblance and complexly structured representations.

Like resemblance theorists, Barsalou sees value in conceptual representations from the world in a way that captures information. What Barsalou terms 'perceptual symbols' are

> neural representations in sensory motor areas of the brain; they represent schematic components of perceptual experiences… they are records of the neural states which underlie perception. These records are then available for recombination in imagination, that is, for thoughts about how the world might be (1999, p582).

Unlike mentalese, these symbols come in a sensory mode- eg, as part of vision or hearing. The promise is that this will explain the semantic grounding of these symbols (perhaps via resemblance, though Barsalou doesn't explicitly consider this option) and help explain how they are used, i.e., in virtue of the information they carry within them. Cummins, O'Brien and Opie would agree here, no doubt- providing the information capture was in a causally relevant form.

The neuroscientist, Rolls (2001) agrees with Barsalou about the virtue of cognition-guiding sensory content in representation,

> ...symbolic representations must have a great deal of information about what is referred to in the world, including the quality and intensity of sensory events, emotional states, etc. The need for this is that the reasoning in the symbolic system must be about stimuli, events, and states, and remembered stimuli, events and states, and for the reasoning to be correct, all the information that can affect the reasoning must be represented in the symbolic system, including for example just how light or strong the touch was, etc (2001, p12).

Roll's seems to agree with Cummins that representations do need to bring a lot of information to cognition, but disagree that resemblance is required. But, he also seems to have in mind representations of non-arbitrary form- given that he describes them as 'symbolic'.

Whether or not this is a viable picture of representation, there is a natural tendency to assume that the derivation of cognitive representations from sensory representations takes representation towards a non-resembling form and a non-resemblance based semantics. Whilst we can imagine sensory representations being somehow physically shaped by their referents, the more the abstraction process extends inwards, as it were, the more it seems that shape becomes a matter of how the representation is used by the cognitive system, as Clark and Wheeler argued above. Perhaps this is to forget that schematic form can still be intrinsically useful, but it *is* hard to shake the intuition that the more abstract a concept is, the less it is concerned with surface properties of things, so to speak. Hence, it is tempting, if not necessarily valid, to think it less likely a representation with such a content might 'carry' that content in its own physical structure.

Clark and Karmiloff-Smith (1993) give a useful preliminary account of such a 'deriving' or 'abstracting' process, which they call 'representational redescription'. Although they don't explicitly argue for symbolic representation, the tenor of their explanation of the process seems to suggest at least a movement towards non-resembling form. It is true that the authors are sympathetic to a connectionist programme in mental representation in which a resemblance theorist has most hope of finding succour. However, their argument is primarily for a 'freeing up' of the

representational results of perception- a decontextualising- which I think can only involve the loss of properties tied to the history of the representation.

Clark and Karmiloff-Smith's main theme is that cognition involves a disentangling of knowledge gained through perception. They take a data driven, connectionist view of perception as a kind of network-organising response to regularities in the world. And, as I said, this has the flavour of network states coming to be 'matched' with the world via a process of learning and so it, at least, puts resemblance in the picture. However, the authors also see cognition as depending on the transition from example-driven computation to more multi-purpose computation over context-neutral representations. We might say that these representations need to cast off some of their relational baggage that ties them to their original learning context. Now, this would seem to be a particular problem for holistic accounts of concept or representation identity, such as second order isomorphism accounts, where such relational baggage is defining of the content of the representations in question.

To give an example of the point of this decontextualising process, Clark and Karmiloff-Smith point out that connectionist system NETtalk's representations are not available for less example-driven computations. NETtalk is unable to count vowels, for instance, however much of the world of vowels it captures in its connection weights or the structure of its activation space. Its knowledge about vowels is therefore not much use to any other system that might want to make use of the concept *vowel*, and is therefore not available for higher cognition purposes.

Hence, Clark and Karmiloff-Smith argue for a process of representational redescription or a recoding of sorts to produce such multi-purpose representations. Empiricist minded psychologists and philosophers have long pondered over possible mechanisms by which perception and perceptual processes might yield vehicles appropriate for cognition. Abstraction, here, is a common metaphor. In connectionism, the technical process of skeletonization is a more concrete candidate for the process of redescription, but it serves well as an abstract metaphor anyway. Any brain implementation of cluster analysis, or of any other analytical tools for analysing network activity, might be a candidate, as well. However, as we noted above, (holistic) NETtalk's representation of vowels, as revealed by cluster analysis,

is inextricably bound to its representation of consonants. It is the gap, or distance, between the two types of representation that gives the representations their content. Therefore, if 'vowel' is going to be redescribed, so will 'consonant' have to be, and at the same time. That is, representations in holistic representational schemes travel in groups. This, I think, may turn out to be limitation for resemblance representation, namely that it best serves whole domain representation rather than between domain representation.

Leaving aside the tensions between holism and decontextualisation for the moment, Clark and Karmiloff-Smith do not have anything too specific to say about the nature of the vehicles that representational redescription might deliver, but the basic notion is 'reduced form of some kind' (p 492). Again, it is true that 'reduced form' is compatible with a more *schematic* resemblance- as Barsalou would put it - however, it is not clear that that is what the authors have in mind. Inevitably, representation seems to involve an element of abbreviation, and it needn't be 'resemblance preserving', so to speak. It is in the nature of concepts or representations that their content is abstract to some degree. Indeed, Cummins points out that, in one sense, only particle to particle duplications are not abstract- with any other sort of representing, something has to be left behind (1996, p109). Sometimes this process is described using the notions of analogue and digital formats. As Lyon (1997) summarises Dretske-information flows from the analogue to the digital as it becomes more precise- as digital information is extracted, a sieve or grid has been applied to the flow of information (Dretske, 1981, p142).

So, at the very least, more abstract content is going to mean more abbreviated form. Could such 'decontextualised' or 'redescribed' representations remain as structured entities that resemble their referent? Perhaps. The advantage of being intrinsically informational would seem to usefully hold of more abstract representations as much as less. On the other hand, I will argue that the task of combining with other representations is possibly hampered by structure. (Although, we should note that it has been claimed that combining structured, information rich representations has the advantage of producing an immediate integration of information in the new structure.)

Perhaps, there is at least evidence that abstract representations should be discrete, rather than continuous. Markman and Deitrich claim that abstract content is best captured by discrete representations, and for similar reasons to those of Clark and Karmiloff-Smith that have to do with leaving the influence of sensory stimuli behind,

> continuous representations are not abstract ... They are typically highly specific, because they correspond to some specific continuously varying input. In contrast discrete representations are often abstract. In fact, creating abstractions produces discrete representations (p 112).

Markman and Deitrich also claim that the abstract contents of such representations have to be determined by functional role and only discrete representations can have functional roles with respect to each other (p 111). It would follow that in order for there to be abstract representation there must also be discrete representation with functional role semantics- presumably of arbitrary form. To go down this path, I would have to side with Dennett against Fodor over the holistic semantics of at least some representations. Presumably, though, Fodor would not be unhappy with the conclusion of arbitrary form in discrete representation.

But why should abstract content be 'use' or 'role' content? Markman and Dietrich (2003, p111) claim there is 'significant evidence that peoples' concepts are highly interconnected and that concepts derive their meanings in part from these connections.' For example, the observation that categories are organised into hierarchies suggests discrete representations are in play because people keep these levels distinct. In general, according to Markman and Dietrich, 'the basic facts of conceptual organisation argue strongly for discrete representations'. However, whether the facts of conceptual organisation point merely to discrete representations or, further, suggest that the content is a function of that organisation between discrete representations is another question. If functional organisation is constituitive of content, then organisation and (structural) resemblance are not compatible sources of content. On the other hand, perhaps the facts of conceptual organisation are germane just to the matter of discreteness.

We might ask though, if, in any case, resembling representations are capable of this kind of organisation- can we have in our heads hierarchies of maps, for example? Certainly, it is possible that higher-level representations, in a hierarchical scheme of resembling representations, may be more schematic in form than lower level representations. This would seem to leave a resemblance semantics intact. However, this pyramid of isomorphism would only be possible if it were possible to unite all lower level examples of any concept in a common schematic form they all shared. And, it seems unlikely that all concepts work that way, since all instances of a type may not share many properties in common. For example, in the following simple conceptual hierarchy:

Modes of transport
Car, bike, horse, swinging rope

it is hard to imagine any abstract structural properties the items along the base of the pyramid all share. We can imagine structural/informational baggage within each component concept just confusing the issue, since it wouldn't relate to the hierarchical relationship being represented.

However, as we noted above, the resemblance theorist has another hand to play here. It may be that the representation of a whole domain such as *transport* or *political theory* or *democracy* is the proper locus of resemblance in representation. Maybe, it is whole domains that enjoy interesting abstract structure and, moreover, of the sort that might be captured by an abstract structure in a network, or its activation space. NETtalk is a good example of an abstract domain, namely, the phonemes of English, being captured somehow by the network. No NETtalk activation state resembles the phoneme is represents, but its relationships to other activation states resemble relationships that phoneme has to other phonemes. So perhaps this kind of resemblance can be used to represent something like the abstracts concept of 'democracy' and 'transportation'. (Of course, Fodor (1998) argues concepts like 'political theory' aren't structured in this rich way, in any case.)

Thus, the resemblance theorist reminds us that sub-domain representations needn't resemble anything. Moreover, the holistic, second order resemblance we find in state

space semantics doesn't fail as an implementation of Cummins' account of content just because the individual component representations involved, while being physically structured, are not themselves resembling. As we've seen, Cummins is likely to reply here that nothing is ever represents outside of a larger representing structure of some kind. A representation, then, needn't resemble if it is part of such a resembling structure. This means, though, that the larger the domain of resemblance, the greater the number and levels of sub-domain representation that needn't be resembling. (I ask below, what does it say of a theory of representation that it doesn't apply to the great majority of the representations it puts forward)

I think several problems open up, here, for the resemblance theorist. These problems are related to the fact that, in the end, when we are using second order isomorphism to represent sentences, faces, or whatever, constituent structure will not be captured explicitly in a single representation but only by the representation's relational syntax with other representations. For example, as we saw in chapter 2, a 'noun *qua* subject' representation in an Elman type network will only capture that structure in virtue of being in the subject part, of the noun part, of the activation space - that is, in virtue of being similar and different to the rest of the activation states the network might use.

A possible fundamental obstacle for this relational type of representation is that the apparent measurements of similarity (albeit sophisticated ones that are capable of reflecting complex structure) that we find in activation space, for example, may not qualify as representational. A sceptic might ask, can it be that just because one set of network states are physically similar to each other and dissimilar to another set that the states themselves represent things in the world similarly related to one another? Ramsey (1997) is such a sceptic. His catch cry is 'responding is different to representing'. Responding similarly to similar objects doesn't amount to representing them as being similar. It just means that the processing/causal processes in question follow similar trajectories. If response is representation, then that is a weak notion of representation, and we seem entitled to ask what doesn't count as representation? It seems many systems react similarly to similar situations, eg running water, or Ramsey's example, the digestive system without representing the situations that caused them.

The dispute lies, I think, in how robust we want our 'use' component of representation to be. As we saw in chapter 1, finding isomorphisms between a problem solving device and its problem won't interest most philosophers of representation unless something 'reads' this isomorphism in some way or is pushed around by it - representations have to be used as representations by a larger system to count as representations and that goes for representations of similarity and difference, i.e., second order representations. Somehow isomorphism as intrinsic syntax eventually has to do some work to count as representation. Remembering, then, that the isomorphism in question is one constructed out of activation states and does not exist within particular activation states, I am tempted to believe that although facts about individual activation states obviously do causal work, facts between activation states representations do not. How could they, when each activation state is consumed in isolation? The representing structure in question, itself, namely, the relational properties of the whole set of activation states, never appears in the network, just its parts. (I make this claim (somewhat monotonously) in various guises in several places below.[46])

Generally, the *larger representing structure* defence of non-resembling representation leaves us asking, is resemblance at work in the activities of the component non-resembling representations? In a case of whole domain representation such as a map, the answer seems to be 'yes, the whole resemblance makes the component representations useful.' However, that leaves cases unlike a map where the component representations are not accessed all at once by a consumer system. The limiting case, here, is when a component representation is used in representation outside representation of the original domain.

A part of this problem for second order resemblance is that, even in the representation of its home domain, it threatens to peter out when there is work to be done because, in the terms of chapter 2, it doesn't bottom out in the only effective syntax we can imagine, namely, intrinsic syntax. In other words, while it is consistent with Cummins for a dot on a map of Australia to represent Perth in virtue of being part of a larger

---

[46] Actually Ramsey goes further and questions activation states as representations by themselves, given the way activation states seem locked into the one type of consumption.

structure resembling Australia, it is not consistent for an activation state representing something like a sentence to represent the sentence's constituent structure in the same manner- since unlike the dot representing Perth, the activation state is purportedly representing structure. (Also, using the activation state to represent that sentence in another network will be like using the same dot to represent Perth in a written sentence or in a table of figures.)

Activation states represent structure *implicitly* - and certainly not *explicitly*, as Clark tries to argue, and not in Dennett's sense that the structure in question is logically implied. In an Elman type case, the 'as if' structure might seem to do some implying- namely, which representation should come next. That is, if there is syntax in 'boy *qua* subject' - it lies in the fact that the network thinks the next word might well be a singular verb or a pronoun in a relative clause. But my preferred take on this from chapter 2 is that such 'as if' structure doesn't imply anything, but is revealed and *exists* as an effect and not as a cause. It cannot then be 'informational' in the way Cummins' wants structured representations to be.

By way of illustrative contrast, this is certainly not the case in one of Cummins' favourite examples- 'the autobot' (p 95). Here, a car runs through a maze while a card, with a complex slot in it, sits under a car, connected to a steerable rear axle. The card is the putative resembling representation. Cogwheels on the rear axle pull the card through the car such that the car's wheels are steered by rear axle's contact with the card's slot. The trick, of course, is that the structure of the card's slot is explicitly isomorphic to the desired route of the car. But, this is exactly *not* the relationship between an activation state and the constituent structure the activation state is taken to represent. The network is not guided by any such resembling structure when it consumes such an activation state. I think this helps make clear why it is hard to imagine a network being sensitive to implicit structure *qua* structure and hence why it is also hard to see how such 'as if' representation can be part of a mechanical explanation of the network's use of structure.

So, while it may be true that a state space diagram, or the state space itself, has constituent structure, it also seems true that that structure doesn't work in the manner that Cummins autobot card or a traditional map does. In the end, state spaces

themselves are not really like maps, contrary to Churchland's position. Churchland must be mistaken, in Cummins' eyes at least, when he treats 'the lasting system of prototype points, and … the similarity and difference relations that structure the activation-space that embeds them' as representational. For Churchland, activation states are 'specific activations of that background machinery [that] typically constitute specific portrayals of the world's local character here and now' (p 39). In effect, an activation state says where in the state space, where in the divisions of similarity and dissimilarity, or in the conceptual map, the input lies.

The big difference between Churchland's lasting system of prototype points and Cummins' pictures of maps is that Cummins' map metaphor suggests synchronic resemblance, whereas state spaces do not. That is, a map usually exists in its entirety at some point in time. Now, it is true that a state space *diagram* will also exist in full at any instant; and that a weight state might, as well. Even a lasting *system* of prototype points might do so also. But what cannot do so is the whole set of activation states which make up the state space. State spaces are never, therefore, in place to drive cognition, even if the weight states that give rise to them are. The GPS display (for hire cars) metaphor that Churchland employs, where *current position* representations are taken as analogous to activation states, is therefore inappropriate.

Abstract representation in smart search

Contrary to my position, Clark and Thornton (1997) argue that Elman's networks provide us with a *good* indication of how abstract representation might be instantiated – and, in particular, how it might be generated during difficult problem solving. According to Clark and Thornton, some problems are difficult because their being solved, or learned, depends on the detection of abstract properties hidden, in some way, in the sensory or input representations that face the problem solver. Such learning, or search, requires some 'insight', then, into the target domain- something more than just run of the mill, connectionist, pattern extraction. Clark and Thortons' key idea is that some kind of *recoding* of the input representation is required. The search space in question is, hopefully, narrowed down and the problem becomes

amenable to solution via simple pattern extraction. In this way, a kind of poverty of the input is overcome via a two-stage presentation and re-presentation of the input.

In Clark and Thortons' terminology, *type-one* problems are solvable through exploitation of observable statistical effects in the input data (e.g., probabilities). Pinker (1999) notes that connectionist pattern extraction of this sort neatly explains how children might learn the *family resemblances* of irregular verbs in English. ''Type 1' problems are, in this sense, 'statistical'- connectionist networks might deal with novel or degraded input just by knowing the most likely answer, given the patterns it encountered in its learning regime.

Type -2 problems, on the other hand, are 'relational' (p 58) – they require more than the extraction of probabilities. One of the authors' examples is a parity learning case. The task in question is to recognise a parity across inputs, namely, oddness in the number of ones in the input of three binary values, eg, 1, 0, 0, or 1, 1, 1, . That these two inputs should be treated alike is relational problem give that oddness, here, is not statistical, i.e., there are 4 ways of being odd and there are also 4 ways of not being odd, such as, 1, 0, 1,. This means that the problem cannot be solved by calculating probabilities of parity between, or oddness in, any two inputs.

In Clark and Thorntons' testing of the application of backpropagation learning to this problem, a 3-layer feedforward network was able to learn the problem (after 50 000 iterations) but crucially was unable to generalise to unseen examples. They concluded that,

> Parity cases… do not really warrant the customary optimism concerning the chances of back propagation in a multilayer net hitting on the right recodings to solve type-2 cases. Instead as we move toward larger scale, more realistic cases, we find a robust pattern of failure (p 60).

According to Clark and Thornton, a typical 3 layer feedforward network lacks the computational power the we engage when solve type 2 problems such as parsing

language; it lacks the representational power to make the computational dimensions of the problem manageable. Clark and Thornton suggest that humans use better representations, than such processing power can deliver, to do less computation. Better representations partition the problem better, and allow for better analysis of the input (see Dennett, below, on the power of better labels). This sounds much like Karmiloff-Smith's notion of redescription, but, as the authors note, her idea is about 'increasingly general, flexible and abstract recodings' that are made 'available for use *outside* the original problem domain' (p 63)- not *inside* the problem domain.

What clues, then, do we have about the learning or search that delivers these recodings or partitions of a problem? Clark and Thornton suggest Elman's parsing networks as an example of successful higher order pattern extraction through recoding. We can remember that these networks seemed to extract the grammatical structure of complex sentences in carrying out their task of predicting suitable next words given that a) they had some success in its task, and b) network analysis seemed to suggest a partitioning of the problem in accordance with that grammatical structure.

Elman explains the success of one network as depending on it making the right distinctions early in its training:

> … the network appears to learn the basic distinctions- noun/verb/relative pronouns, singular plural etc. – which form the necessary basis for learning more difficult sets of facts that arise with complex sentences (1992).

This is certainly compatible with Clark and Thortons' two-stage representation. Elman achieved this two stage learning by tailoring the network's training regime to allow it to solve the simpler problems first – such as representing nouns, verbs, and subject-verb number agreements. He achieved this by first restricting the training samples to simple sentences. He then introduced more complex sentences containing relative clauses, for example, to a network that was now able to 'make basic sense' of these sentences so as to find more difficult relationships in the data, such as agreement across embedded clauses. In other words, in order to 'follow' complex sentences with clauses, the network needed to be able to first recognise lexical

categories and subject verb agreements. Only then could 'the higher level grammatical properties pop out' as type-1 regularities.

We are back, then, to considering whether a state space semantics is compatible with Cummins' position on representation and cognition. Given the account of gradual learning we find here, it seems unfair to suggest that the early distinctions mastered by the network are not put to use by the network in the detection of more abstract relationships. Something seems to have been 'learnt' about grammatical structure, and learnt in a particular order, and apparently represented in the process. When we question how it is represented, we seem forced to return to analysis of the network's activation states where we find that the network directs inputs through different regions of middle layer activation space according to the grammatical structure of the sentence that is being constructed. In other words, the organisation of the network's own dynamic state space seems to resemble the relative structural properties of the sentences constructed[47].

How can we, now, deny that such organisation of the network is not representation in action? Certainly, no complaints about holism and decontextualisation apply here, since there is no claim to be moving representations outside of their problem domain. However, on the other hand, the same claims I made above about the causal efficacy of such holistic representations do seem to apply here. In summary, individual representations, namely, activation states, suffer from representational impotency. Their resemblance-based content cannot drive cognition Cummins' style, since it is not contained within any representation but between representations. Clark and Thornton are concerned with the representation, or at least the 'coding', of concepts like 'noun'. But, while it might be said that a region in the network's state space encodes 'noun', no single activation state will have only the content *noun* nor any

---

[47] Actually, there are two points here. Firstly, similar sentences carve a similar trajectory through activation space; and, secondly, because these representations are dynamic, in that they are built up over time (the networks are recursive) – the actual activation pathway through activation space of a single sentence representation might resemble in its gross details the structure of the actual sentence. In general, analysis of Elman's networks state space may show an abstract resemblance between its set (of sets of) activation states and the sentences that they track. Similarly structured sentences will carve similar paths through dynamic activation space.

specific subpart of any activation state. If an activation state represents something as a *noun* it does so in a way that isn't physically realised in its own structure[48].

Alternatively, the entire family of activation state representations considered as the one object seems impotent, as well. As the entire family never exists in total at any one time, it cannot be used as the one representing object, which we agreed was a qualifying height that even low road representations needed to pass over. We saw in chapter 2 that Elman thinks that his networks learn both types and token, categories and members. However, we have just noted that the tokens lack causally efficacious structure, and that the same seems to apply to the types. If Cummins is right about representation, then state spaces don't seem to be the right representing vehicles for a system to rely on.

What a critic of my position might say here, of course, is that Elman shows us that a mastery of explicit representation of compositionality is not required in order to solve type 2 problems. In the end, despite the fact that no one representation explicitly represents a syntactic structure, the system is sensitive to such structure, and indeed complex examples of it. What other explanation do we have of this fact other than that the system represents the grammatical structure of English sentences, somewhere in its architecture and occurrent states? Of course, a big question hanging over this explanation of smart search is whether or not such networks can scale up to deal with the kind of novelty and heavy traffic we find in real language. If they could, I suppose 'seeing would be believing'. Although, another response might be to say that what ever is going on here either isn't representation, or if it is, it lies somewhere else than in activation space.

My main concern is with state space structure considered as a Cummins' style resembling representation. Given the success of Elman's networks, my position may seem to be unable to make sense of the power of tools like cluster and principle

---

[48] Elman writes that 'some states may have significance not simply in terms of their similarity to other states but with regard to the ways in which they constrain movement into subsequent state space… (1992, p159).' If these states do this in terms of their content, then this is not resemblance content at work, but more like functional role content. We are back with Horgan and Tiensen's notion of 'as if' guiding structure, which isn't, really, guiding structure at all.

component analysis. Surely, it must show something – if not the actual representations that the network uses, then at least how the network divides up its problem? As we shall see in chapter 8, perhaps, for instance, the representations revealed by state space semantics lie in the weight structure themselves, as O'Brien and Opie claim. There is certainly potentially a better case to be made for synchronic representation, and representation with actual physical constituent structure, using weight states. We should note that the state space semanticist may have another card to play if it can be shown that state space diagrams are comparable across networks. Churchland claims they are,

> We can perfectly well define and use a purely internalist notion of sameness and similarity of *configurations-in-activation-space* for prototype families across distinct networks. Further, we can define what it is for two prototype points in distinct networks to occupy metrically identical or metrically-proximate positions within their respective prototype families (p 12).

Churchland is making a case, here, for sameness of content across representing vehicles (albeit of the same domain). If he is correct, then perhaps this will tell us something about what different successful networks for any particular domain have in common. Perhaps state spaces themselves can be shown to be isomorphic to each other and this fact used to pick out divergent networks containing the same kind of knowledge and equivalent representations. Although two networks might solve the same problem using different weight states, and even different numbers of input and middle layer nodes, the divisions in their weight space might be shown to be abstractly equivalent.

Suppose this is true. Would that convince us that concepts are sets of regions across different state space? A domain might well shape different networks in isomorphic ways; but, it doesn't follow that similarly organised state spaces are similar representations or concepts of that domain. We could ask the same question we asked about state spaces themselves- is this even more abstract property doing any representing work?

Sutton (1995), however, takes a different view of the causal efficacy of state spaces, and presumably of 'meta' state spaces. His view is based on a treatment of the concept of causation, which I outline below.

State Space Semantics and reductionism

Sutton considers the ontological implications of connectionist explanation; he discusses Clark's (1988, 89) move to postulate features that causally explain but don't cause 'in person', so to speak. This is Clark's take on beliefs and desires etc; one that makes use of Jackson and Pettits' (1990) distinction between process explanations and programme explanations. On this story, belief /desire explanations are programme explanations rather than process explanations and the same goes for 'cluster analysis' or 'state space division' explanations. Programme explanations don't have to be about implemenatational properties, or intrinsic syntax. They posit entities or properties that program causation across a range of similar cases but are multiply realisable eg, fragility, or increase in temperature or believing that the ice is thin, or having a certain kind of structured state space revealed by cluster analysis, or, presumably, the meta cluster analysis Churchland talks about. Process explanations, on the other hand, posit entities that cause first hand, so to speak, eg, connection weights and what we have been referring to as intrinsic syntax.

According to Sutton, Clark is happy with cluster analysis explanations as explanations of mastery of a domain because 'they capture constructs which, though themselves causally inefficacious, highlight important facts about an important range of cognitive constitutions.' Thus, we can match two differently weighted implementations of NETtalk and understand this match even though what they share in common, their abstract organisation, plays no causal role in their behaviour. Moreover we can rule out any other networks as capable of mastering the domain of NETtalk, the English grapheme to phoneme function, on account of a mismatch in their state space divisions with the canonical one. Clark considers this is an epistemological matter, more than an ontological one.

This is close to Ramsey's second justification for representation talk, which we made use of in our Dennettian reply to Cummins, namely, that representation talk reveals the functional organisation of a cognitive system. Although, a difference, here, may be that the functional organisation Ramsey takes representation ascriptions to reveal is closer to an 'algorithm revealing' organisation than the structural organisation *fragility* or state space diagrams reveal. We seem close to Bechtal's distinction, mentioned earlier, between mechanical explanations and covering law type explanations. Perhaps state space explanations are more like covering law explanations- if so, this would explain why they don't meet Cummins' account of representation in cognition.

In any case, Sutton wants to get more than Clark out of state space divisions. He argues, against Clark, that any kind of explanation makes ontological commitments and so he looks to get more ontological work out of cluster analysis and state space analysis. Higher-level explanations do have epistemological value, but they also require ontological support and make ontological commitments. They must be reducible to lower level explanations in the sense of being implemented in some identifiable (if ridiculously complex) way so as to be causally efficacious.

Sutton thinks state space divisions do causal work,

> My hunch is that cluster analysis can give us genuine causal process explanations of the results of the processing. (1995, p362)

His main argument seems to be that programme explanations only get up and running because they supervene on whatever realises them. For example, an increase in fragility causes a container to break because that increase is supervenient on a particular change in the containers physical properties. Hence, being multiple realisable doesn't rule out being causally efficacious; indeed most so called process explanation properties will turn out to be multiply realisable *and* causally efficacious, even connection weights.

Sutton, then, makes a case for ontological reduction securing causal efficacy:

> Causal program explanations in general must be identical with or supervenient on lower level process explanations to be legitimate. The reducibility explains why they are valuable: because the partitions in state space which they cite, for instance, are really there in state space, constituted by, realised by, and supervenient on the (numerical level) connectivity weights in the particular case (p 362).

There are echoes, here, of O'Brien and Opies' position that the physical properties of the network do the actual representing and not mathematical descriptions of the network; and in particular, its weight state. Except that, while we are focusing attention on the weights that give rise to activation states, rather than the activation states and their abstract relationships that compose the state space, we are explaining the causal efficacy of the activation states in terms of the causal efficacy of the weight states. The question to ask is, can one thing's causal powers, at the same time, be another different thing's causal powers, as well? Can a weight state's causal powers also be the network's activation states' causal powers? After all, on the story so far, it is the activation states, or the divisions they create, that are meant to be the representations.

If the causal efficacy of state space divisions rests on the properties of the weight structure, how is Sutton to answer someone like Ramsey who claims that there is, therefore, no need to invoke the programme explanation, the state space explanation, given that it doesn't function to explicate the workings of the network? State spaces and activation states start to look epiphenomenal. Sutton may be correct that state spaces supervene on the mechanical facts such as weight states, however, that doesn't make the components of the programme explanation mechanical components. Clark and Sutton might say that the higher order explanation does help us make sense of the mechanical facts and give us a useful way of approaching them, but this is a different point.

In summary, Sutton's attempt to get causal work out of state spaces is also an attempt to marry second order resemblance (perhaps of a diachronic kind, at least) to Cummins' notion of content as an actor in cognition. His notion of programme explanations supervening on process explanations is probably not enough to do this

given that content, on Cummins' account, is part of a process explanation that makes the mechanics of cognition clear.

Chapter 6.   Combining Representations

In this chapter, I examine the role for representations of combining with other representations and assess its compatibility with resembling representation.

The trials of combination

In general, with representation, less can be more at times. Markman and Deitrich put it that,

> the benefit of losing information from continuous representations is the production of a set of discriminating, potentially referring, discrete representations that are combinable (2003, p112).

Beyond discreteness as a virtue, Fodor (1998) has argued that, for all the evidence of concept combination in cognition, there does not seem to be evidence of the combination of complex structure. This is not to lump resemblance accounts of representation with a prototype/stereotype view of concepts, however, if concept combination is representation combination, it doesn't seem as if complex representational structure is necessarily being combined. To use Fodor's well known example as an indication of the problem, whatever the prototype of 'pet' and 'fish' are, or however they might be represented as resemblances, the concept of a 'pet fish' doesn't seem to be an amalgam of two complex structures standing for, or resembling, pets and fish.

Being capable of combination seems to be an important property of representations. We can remember that, in chapter 1, Clark and Wheeler made systematic combination a defining characteristic of representation (Millikan and Von Eckardt take similar positions relating to high road representation). Indeed, the whole point of Clark and Karmiloff-Smiths' representational redescription is to redescribe existing representations in order to construct more combination-friendly representations. In sum, combination seems an important role for representation- it serves in inference, and simply in representing the world's objects and properties standing in new relationships to each other than as initially experienced.

On the virtues of unfettered combination, Clark and Karmiloff-Smith make an ontological claim about thought: real thought conforms to Evan's generality constraint. In other words, thoughts have to be capable of being used interchangeably. Perhaps the same applies to representation (or, 'high road' representation, at least), as Clark and Wheeler claim. Cummins might reply, here, that, if we are thinking of abstract concept combination, this is the wrong way to think about abstract concepts. In the first place, concepts are not single representations but more analogous to books or libraries (1996, p134). But, to reply, surely the brain is capable or latching on to abstract properties in a more simple way than that? And furthermore, somehow, we do seem to combine representations of abstract properties in our thinking. If concepts are made up of lots of representations, how do such complex entities combine? Surely, Cummins thinks there must be *some* combination in the brain, if not of concepts, then at least of representations, or if not representations, then at least of discrete bearers of intentional properties?

The trouble for Cummins-type representation is that flexible combination seems to suggest less structural baggage rather than more, that is, simpler, smaller, and easier to combine representations. Not everyone draws this conclusion, however. For example, Lycan (1993), even though he thinks that it is obvious that the mind/brain contains a collection of semantic primitives and a recursion procedure for combining them into complex thoughts. Lycan is confident that given that we can think about almost anything, then there has to be some recursive combination going on; it's the only way to get unboundedness from finite resources.[49] Yet despite this emphasis on recursion and combination, Lycan explicitly eschews any account of representation that includes language-like representations. He states that the representationalist is not committed to physically salient representations or the language-like syntax of predicate calculus. The distributed representations of connectionism are valid candidates for representation and indeed the state space divisions of NETtalk, revealed by cluster analysis, are candidates for concepts, as far as Lycan is concerned.

---

[49]The argument is partly by way of analogy with Chomsky's argument for compositional rules applying to morphemes given the productive nature of speech, made against Skinner in the 50s. Although it stands on its own, as an inference to the best explanation.

As the reader can imagine, my question for Lycan is going to be something like, 'Are state space divisions, or the activation states that compose them, the sorts of things that are going to be easily combined- particularly, given the holism of the representational format?' Surely, activation spaces, and the activation states they are composed of, are just the sort of things that are tied to a particular learning and representational context- as components of a solution to a particular problem, each part will only make semantic sense in relation to the other parts. Although each part may be structurally simple, as opposed to the whole scheme, each part cannot be combined without constraint because of the inherent holism of its 'home' representational scheme. Like components of a map, each part, in effect, cannot carry its content intrinsically or essentially.

One way around this problem of cross-scheme combination, and perhaps this is what Lycan had in mind, is to suggest that our concepts are all part of the one holistic representational scheme- the one giant state space. This way, there are no cross scheme identification problems to deal with. While no one presumably thinks the brain is one giant network, perhaps it is possible that somewhere in the brain the concepts available for recombination are represented in the one network's state space divisions. Thought might turn out to be something akin to the sequential activation of this super network's middle layer. That is, combination would be more like concatenation in time rather than space. (In chapter 7, I consider Dennett's argument against spatial concatenation.) Our giant conceptual maps of the world, on this account, would guide us through planning and inference by guiding the order of representation activation.

This one map/web solution still leaves problems with '*between* brain' conceptual identity. That is, we reach the sort of cross-scheme identity problems Fodor and Lepore have raised against holistic representational schemes in general and State Space Semantics in particular (eg, Fodor and Lepore, 1992, 1999). Could, for example, any two people ever be said to have the same thoughts, given that that the state spaces of any two of our supposed 'whole conceptual scheme' networks are bound to be different. Similarly, how could the learning of new concepts not alter the

nature of all the old concepts in virtue of a forced rearrangement of existing state space divisions.

But putting those disputes aside in order to concentrate on the role of resemblance in this account of thought, as a resemblance compatible position, this one-network solution may still face difficulties living up to Cummins' notion of cognition-guiding resemblance. The idea is that our conceptual network's state space would count as a resembling representation of the world. Consequently, the non-resembling activation states of this network would be part of a larger representing structure. However, as I note elsewhere, surely these non-resembling activation states themselves would be the appropriate causal players to focus on- the entities from which consequences follow- not so much the whole set of state space divisions which is never instantiated as the one representing object? After all, it is the activation states that are meant to do all the combining here- they are the explicit representations- and combination is supposed to be where representations are causally efficacious. In other words, the epiphenomenal resembling character of state spaces still seems a problem, and this is exactly what resemblance theories of content were supposed to avoid.

Perhaps there are other ways of thinking about the combination of resemblance representation, in general, and state space representation, in particular. Consider, again, Millikan's example of the overlaying of two maps of the one domain to produce new information about that domain not contained in either of the original maps. In connectionist terms, perhaps whole state spaces may be combinable in this way. However, as I suggested in previous chapters, this process would seem to rely on the non-relational typing of what she terms 'shared middle terms'. That is, in order for two maps to be overlaid, some shared components need to be identified. The problem is that unless the two maps are the same maps, in which case no inference is possible, these middle terms will not share the same relations with their fellow representations[50]. Unless there is some way of finding partial matches of internal relations across maps (which may be possible for similar maps, at least), this leaves intrinsic features as the only available identifiers of common terms.

In state space semantics, however, finding shared middle terms non-relationally isn't an option, since all representations, even within the one network, are typed relationally or holistically. Moreover, as Millikan tells her story, the representations in question seem to resemble synchronically, so to speak. That is, all the components of the maps are present at the same time. Whereas, as noted many times, state space maps are diachronic entities; they are instantiated bit by bit over time, and so are never available at any one instance for combination. In sum, Millikan's overlaying of maps looks beyond state space representation.

In any case, even if state space representation could work in this way, perhaps it still wouldn't account for the promiscuous and systematic combination of representations we assume to exist in human cognition. As candidates for combination, maps are limited in their potential partners- that is, partners have to be about much the same thing for the combination to make sense. To meet something like Evan's generality constraint, a representation scheme has to contain representations that easily combine in any combination- something that a scheme of different maps of different domains looks unsuited to doing.

Another candidate for connectionist combination is the idea of activation states being sent to different networks to be combined with or consumed by different weight structures. Clark and Thornton imagine this sort of thing as a kind of search for a fit between new problems and old solutions (as captured in old successful networks). If the weight structures themselves contain a resembling structure, then perhaps Cummins' notion of resemblance driving cognition can be realised within this version of the combination framework[51]. However, again, this isn't the clean, uncomplicated, productive and systematic, compositional combination most theorists of mental representation have in mind. The representations in question naturally come in clumps- in this case as old solutions, or perhaps old maps of a problem. In the next chapter, I will argue that it is hard, in general, for resemblance representation, and its second order variety in particular, to constitute the simple (go anywhere, any time) representation of particular objects and their properties.

---

[50] Again, this is a Fodorian point about identity in representations across holistic representational schemes.
[51] In conversation with Jon Opie

The benefits of representing intrinsically in combination

One of the benefits of combining representations is that the resultant combination might describe or picture some kind of structure in the world –as is putatively the case with a phrase structure diagram, a superposed activation state in Berg's or Elman's networks, the state spaces of those networks, a map of the London Underground or an English sentence such as 'The yellow house sits on top of the green hill.' Different sorts of representation combination are possible, then. As we have noted, the combination can be synchronic or diachronic, explicit or implicit; and the component representations, structured or simple, holistic or atomistic, context-dependent or decontextualised.

I summarise, now, some benefits of simple, discrete, decontextualised, atomistic components in synchronic combination- in other words, the benefits of using a representation scheme analogous, in some ways, to written language and of not using, in particular, second order isomorphism in combination. In the end, I think this argues against resemblance representation being the only kind of representation in the human brain.

In the previous section we made mention of the intuition that less structured representations are potentially more easily combined- simple representations allow a system to represent and combine as much structure as it needs to since name-like representations provide a system with simple building blocks. Second order isomorphism, on the other hand, ultimately has to represent structure as similarity between structures. This ties representations to each other and makes them hard to use outside of their original domain – whereas, as Clark and Karmiloff-Smith (1993) argue, if the representations are decontextualised (or atomistic) then this allows for easier integration with new knowledge and integration in systems using other formats.

Clark and Karmillof Smith also argue that decontextualised representations allow systems to cope better with novel structures. They claim that novel structures make more sense if they can be represented as novel combinations of identifiable parts. For example, a bizarre sentence like 'The universe spoke.' can be represented as a

combination of 'universe' and 'speak (past tense)' and not as a novel whole similar, to some degree, to other wholes the system has dealt with in the past.

As we noted in chapter 2, connectionist representations like those in Berg's processor and Elman's networks are treated by the system as wholes that the system hopefully has a place for in its state space landscape. Pinker's (1999) work on the past tense has tried to show that expecting a network to be able to treat novel wholes merely in virtue of detecting a similarity with past wholes is not the strategy the brain likely uses, for example, in dealing with novel verbs like 'to outgorbechev', which just don't resemble much else in the language. This is not to say a system using resembling representations cannot use part-whole relations, but again, if the resemblances involved are a second order kind, then component representations may not be broken down into the sort of constituent structure that is helps a system deal with novelty. Again, this is the difference between first and second order resemblance representation of constituent structure. First order resemblance of constituent structure looks better placed to meet the requirement of causal efficacy. Though even, then, an atomism in component representations may be required.

Markman and Dietrich (2003, p106) make a related point about access to real structure, namely, that 'people are good at accessing the commonalities and differences that emerge from comparisons' whereas, scalar distances between concepts are not the same thing as the commonalities and differences between concepts. Thus, if resemblance representation were merely the representation of overall similarity between objects, this could be a problem. According to Markman and Dietrich a system has to treat concepts as composed of discrete subparts that are recoverable during comparisons. For example, they note Landauer and Dumais' (1997) network that passed a TOEFL (English) test for synonyms using vector space similarity measurements but could not pass an antonym test. The problem lay in the fact that antonym tasks require sensitivity to internal structure, to the details of commonalities and differences.

A class of important objects with internal structure is the mental representation of rules in cognition. Clark and Karmiloff-Smith claim that decontextualised, explicit representation allow for example-independent knowledge of rules:

Only explicit rules have the genuine systematically manipulable components that make radical flexibility possible. Our argument is thus meant to do for (some of our) knowledge of rules what Fodor and Pylyshyn (1988) attempted to do for the data structures upon which the rules operate (Clark and Karmiloff-Smith, ibid, p504).

I haven't placed much emphasis on the representation of rules as opposed to objects in the world, given that most rules in the brain may be tacitly represented.[52] We can note, here, though, that this advantage of discrete representation-combination in rules, concerning systematicity in rule formation, is one that Newell (1990) uses to explain learning. Newell's basic idea is that learning is a kind of search and search through the search space of possible solutions is easier if old solutions are represented in parts that are easily extractible and recombinable. Accordingly, Newell's preferred representational format, production rules, allow for the swapping of components, or having components altered or replaced. In this way, they are adaptable to new contexts. In general, the search for new rules is easier if the old rules are explicitly structured in this fashion. This may be a reason, then, to prefer Newell's version of the idea of searching using old solutions to the one I described, above, involving networks that were successful in the past. That is, the same lessons for second order resembling structures above, might apply, here: it helps to have access to the parts of old solutions as much as to the old solutions themselves – and in a way that allows the use of those old parts in new solutions.

Markman and Dietrich make a similar point about discreteness in composition as it applies to genomes, thought of as a sort of rule, and thoughts. They argue that, in natural selection,

[d]iscrete genes were needed for a deep reason: only discrete genes can be combined in a such a way that new organisms are produced which are different enough for natural selection to have something to operate on, but are alike enough to be able to successfully mate and produce viable offspring (p 109).

---

[52] However, I claim that the same is unlikely to be true for the representation of properties. In other words, if representations are distinct from tools, it can't be 'tools all the way down', as Dennett has it.

Since 'each gene carries a fixed amount of information for constructing some protein', a gene's effects are presumably limited (leaving individuals similar to each other and able to mate) and selectable despite a gene's combination with new genes in any genome. In other words, genes allow for a kind of productive systematicity within a species. In some ways, we can treat organisms as solutions to problems. If we think of an organism's DNA, in particular, as a solution to the problem of survival, then it is good thing for natural selection that these solutions come in discrete recombinable parts. This gives natural selection something to tinker with, swap around, and choose between.

Markman and Dietrich note that composition in cognition may work the same way as composition in genotypes and phenotypes:

> … only discrete (cognitive) representations can be combined in such a way that cognitive processes have different thoughts to operate on, but where the thoughts are similar enough to insure the coherence required for rationality (p 109).

The idea is that shared components make for useful differences between thoughts. This is especially true if the components are recoverable. Just as individual genes are recoverable from successful genotypes and able to be passed on in reproduction, so are concepts, according to Markman and Dietrich, recoverable in concept combination,

> the original concepts, once combined, are nevertheless recoverable. If concepts are represented in some continuous, fluid substrate then this ought not to be possible…. The fact that constituent concepts are easy to extract from combinations suggests that they do not combine in the manner that fluids do (p 109).

The advantage of this state of affairs is that common components of different thoughts/ complex representations can be matched, and so the relations between thoughts made easier to access and use in inference. The question for resembling representations is, again, do they contain components that are easily extractable? This will be an easier matter if they are identifiable by their intrinsic qualities and not their relational properties as is the case with the divisions in network state spaces.

Markman and Dietrich point out that continuous representations have regions but not parts. What about resembling representations? My claim is that, contrary to what Horgan and Tiensen claim about effective syntactic structure, if a representation only has regions, as state spaces and activation states with superposed structure seem to do, then that representation is not suitable for representing structures such as linguistic structure, or role-argument structure, in general.

In general, discrete, atomistic parts seem a problem for connectionist representation. For example, as I noted above, when Millikan writes about mental maps being laid over one another to produce a new map, she seems to have in mind representations in which shared components can be isolated and matched, if not be retrieved. It is hard to imagine how these matched components might *not* be discrete within their host structure. Yet, Millikan entertains the possibility of maps being combined in superposition in weight states (1995, p 105). This suggests she doesn't share this intuition given the nature of superposed representation - this is not altogether surprising since she is a use theorist of syntactic type, after all. It also means Millikan does not have a first order resemblance in mind for her maps, but that she does have a synchronic mapping in mind. That is, in a weight state, all the components of a representation are present at the one time. However, how shared components of different maps are matched in superposition is especially hard to imagine. As I argued in chapter 2, Millikan's use of the map metaphor here is difficult to follow, and her use of weight states as representations only makes it more so.

In summary, then, representation combination looks a problem for resembling representation either for reasons of excess internal or external baggage. If the resemblance is second order, then, both, combination within the one representation and between representations looks unlikely to deliver the benefits to cognition that representation combination is normally taken to. In the following chapter, I pursue this charge further in relation to two roles for representation combination in cognition, namely, tracking changes in individuals and tracking cognition itself.

Chapter 7. Tracking

This chapter asks if name-like representation offers advantages over resembling representations for a system tracking change in the environment and its own cognitive states.

Keeping track of constancy within change

Human behaviour seems to reveal a tracking of individual entities and their properties. We seem to perceive, conceive and remember particulars and, also, particular *qua* the possessors of particular properties. We can ask ourselves, 'Is that the same one as last time?' What does this reveal about human representation? Can maps or resembling representations track individuals and properties and keep those properties separately represented- or, are symbols required? For instance, if maps were to vary continuously we may have a problem: Markman and Dietrich argue that continuous representations treat 'the entire input streams as a unified whole that varies continuously' (2003, p 106). Whereas, the whole point of tracking is to *not* treat the environment, or the object tracked, as a unified whole; some parts of the environment need to be picked out amidst change as constants. This seems to suggest that some representation of a changing environment itself needs to remain constant.

Botterill and Caruthers (1999) enlist Horgan and Tiensens' so called 'tracking argument' as part of a case for a mentalese of sorts. It as presented as an augmentation to the standard arguments for mentalese from systematicity and productivity in thought. Like those arguments, it works as a transcendental argument in the Kantian sense (of deducing the categories from the unity of the self): there is a fact about human cognition that only X could explain- in this case X is a symbol system.

In more detail, the fact requiring explanation is that 'humans (and other intelligent creatures) need to collect, retain, update and reason from a vast array of information, both social and non social.' This information is information about particulars. Its use includes tracking the mental states of others across different situations and over time, and maintaining the more mundane awareness of the location and properties of every day objects- either new or familiar, present or not present. Tracking is also evident in

the execution of complex skills such as those involved in playing sport, or in 'online', 'response' cognition, in general.

Pinker (1997, pp117-118) makes a similar claim for tracking as a cognitive natural kind: '[m]any animals have to play shell games and thus keep track of individuals. One example is the mother who has to track her offspring, which may look like everyone else's but invisibly carries her genes.' Another case is the task of the predator when a stampede of moving stripes turn a zebra herd into a shell game (well meaning biologists who mark individual Zebras make them easier to track and hence sign their death warrants).

So what concepts or representations underlie this ability to keep track of change - this ability to perceive constancy in change? Perhaps (and this is the expensive explanation), it is a matter of having (a representation for) the concept 'individual'. Pinker is not sure about the hunting lion's concept of 'individual', 'of course, we don't know that hyenas or lions have the concept of an individual; perhaps an odd man out just looks more appetising' (1997, p 118). Again, being able to track an object using sensory information may not be something we necessarily need to posit concepts or perhaps even representations to explain- for example, Rodney Brook's wall tracking robot tracks a wall merely by continually veering to the left and bouncing off the wall upon contact. Similarly, plants seem to track the sun with a minimum of representation, if any. If food, mates, and shelter could all be found so simply, even low road representation might be a luxury. The cricket phonotaxis referred to by Clark and Wheeler, in Chapter 1, is an example we used earlier of tracking without obvious representation of the individual being tracked- though this may amount to a case of detection, at least.

However, as we consider more complex tracking, it becomes harder to keep representations, if not concepts, out of the picture. Pinker (1997) reminds us of the almost transparent case for the importance of the conception of, if not the representation of, individuals in higher cognition: conceiving of and remembering individuals is fundamental to love, agreements, revenge, determining justice etc. You just have to know which 'one' you're thinking about. You won't get through the

vicissitudes of love just by tracking a scent or veering left (then again, you never know…).

This brings us to tracking in the social world. The social intelligence hypothesis for cognitive evolution is that the evolution of human intelligence, and of human belief/desire psychology, was driven by the tracking needs of social life. If we mix in a bit of representational theory of mind, we could argue that the selection pressure to track individuals, their relationships to one another, and their mental states drove the brain into the development of richer representational resources including the ability to represent the representations of others.[53] Group life, merely in its simplest mammalian form, provides the brain with a rich environment just asking for representation. Combinatorial complexity in the domain of genetic, social and physical relationships doesn't take long to rear its head. As Sterelny points out, tracking relationships between third parties is an extra matter to representing others' relationships to oneself. Clark and Thornton might describe the problem as a type 2 problem- a relational problem that requires representing the relata first.

Sterelny (Forthcoming) counsels caution in assigning high-powered representational machinery to explain social trackings, in particular, of other's mental states. A theory of mind, or just a representation of the mental states of others may not be required in order to track the psychological states of others. Representation of what we might call situation- response pairs for an individual, eg, 'baring of the teeth' and 'aggressive behaviour' might suffice to track a mental state such as anger obliquely.

Horgan and Tiensen, on the other hand, wish to crank up the representational machinery to explain the tracking phenomenon in general. For example, they argue that playing basketball is not just exhibiting behavioural dispositions or even modelling the physical properties of an evolving scene. Propositional knowledge about teammates, opposition players, the state of the game, tactics etc needs to be updated and coordinated. In order to use tactics or make complex judgements etc,

---

[53] One point to make, here, is that if social life did require the development of decoupled, multi-cue sensitive representation, in this way, it might also be the true that these, in turn, drove the development of social life (more parties etc.) So, a bootstrapping story is in the offering here. (Jon Opie in conversation)

more is required than perceiving the location and motion of individuals. Properties applicable to these individuals need to be remembered and integrated with new information as things change, eg, as people get angry, tired, or injured.

Pinker's moral from just the minimal tracking of individuals is that,

> the networks of the mind have to be crafted to implement the abstract notion of the individual- analogous to the role played by an arbitrary memory location in a computer.

Computationally,

> what does not work is a pattern associator restricted to an object's observable properties, a modern instantiation of the Aristotelian dictum that 'there is nothing in the intellect that was not previously in the senses' (1997, p116).

We can note, here, that Pinker must have in mind an environment so translucent as to not afford tracking via a detection system of the kind we described in chapter 1; that is, a system that operates via just the detection of the presence of a certain feature in the environment. Meanwhile, Botterill and Carruthers go a step further than Pinker and conclude that

> There seems to be no way of making sense of this capacity [to keep track] except by supposing that it is subserved by a system of compositionally structured representational states.... formed from distinct elements representing individuals and their properties, so the later may be varied and updated while staying predicated on one and the same thing (1999, p196).

In other words, Botterill and Carruthers have in mind propositional representation involving something like easily combinable terms and predicates. Likewise, Horgan and Tiensen argue that given the unlimited number and range of trackings any system must make:

the only way a cognitive system could have the vast supply of potential representations it needs is by producing the representations that are 'needed'. And this means it must have a representation system exhibiting productivity (p 82).

Horgan and Tiensen take this productivity to rest on a combinatorial system of terms and predicates. We can note that Newel (1990) pushes the same need for productivity in criticism of analogue representation schemes. He argues that an open-ended world requires an open ended representational medium capable of making things up on the fly. Newell's concern with specialised representation schemes supports Horgan and Tiensens' position on terms and predicates (though not on superpositional processing). According to Newell (1990, p 61), an analogue approach to representation will lead to more and more representational schemes - as problems multiply, so do aspects of the world requiring representation. Consequently, a designer of intelligence shouldn't rely on specialised materials with specialised dynamics- it would be better to use a neutral, stable medium that registers variety through a set of independently variable parts. This is close to my point in previous chapters about the limitations of holistic representational schemes: they don't provide independently variable parts- they may have parts that vary, but not parts that vary independently.

So, 'open ended representation' goes hand in hand with what Horgan and Tiensen term 'co-reference' or 'co-predication'. The brain needs tokens of types in order to reliably represent or capture different things about the same object, i.e. to co- refer, and also, to represent the same thing about two different objects, i.e. to co- predicate. This is how constancy and change are tracked at the same time. Context neutral, independently variable representations would seem to make more sense, here, since, presumably, the world serves up the same objects and properties in different contexts- so it's hard to see how analogue schemes of representation are going to allow that degree of freedom of representation and combination given that their representations are tied to the actual experiences from which the analogues were produced[54].

---

[54] 'Productivity' needn't imply the faculty for unlimited trackings and judgements, it can just, as Horgan and Tiensen use it, mean something like the automatic facility for applying any new predicate to any term the system possesses- and the automatic facility for applying any old predicates to any new term. (1996, p74) This is clearly the case in predicate logic (their example).

Continuous representations in particular, then, look unlikely to fit the bill. In virtue of being composed of regions and not replaceable parts, tracking via the productive combination of terms and predicates looks beyond them. The question, as asked in the previous chapter, for resembling or analogue representations is similarly, 'Are we talking 'regions' or 'parts'?' Certainly the volumes in state spaces, for example, look like regions- they don't have distinct boundaries, and they are not easily extractable with their identity and constituent structure in tact.

Consider a possible connectionist (and second order resemblance) account of a tracking representation of sorts: namely, that middle layer activation states, conceived of as network responses to the same individual, track changes in that individual; that is, change in the individual is represented by change in the activation states' multidimensional properties. For example, Cottrell's face recognition network (see Churchland 1995, Forthcoming) might track a face over time as it ages and perhaps becomes more androgenous by a shifting of its middle layer activation responses to that face towards the androgyny epicentre of its state space.[55]

In one sense, this solution is more efficient than the classical one because it requires only one representation rather than two- not one for the individual and one for the property. It maybe, then, that the system actually only tracks properties- for example grammatical role or age/androgyny.[56] In that case, we might say this sort of

---

[55] Ramsey (1997) is unhappy about even treating such states as representational. Hidden or middle layers do share some features with indicators or detectors. They can be discriminating in their responses, if graded. And following Bechtel, activation states can be a learned response to the environment. But as Ramsey sets things up, indicator states are meant to have the function of indicating. So that their indicating has previously come to shape the architecture of the system. The net then should have shaped itself around the indicating skill. But in a network - world dynamic, the correspondence that might lead us to think there is an indicating going on is more of a by-product than a structuring cause. It is part of the structuring process not a cause of it. The correspondence arises out of the process of structuring the network, i.e., learning . So, we cannot really say that the co-occurence is put to use by the network (pp 55-56).

Von Eckardt (2003) disagrees with Ramsey. Her point is that training is the improvement of proto-indication, during which the hidden unit acquire the function of indicating. (footnote 7, p437) But that does not stop her doubting that this would show any representation was at work anyway. For her, indication and what she calls teleo informational semantics is compatible with minimal representation , or what we call low road representation.

[56] Philip Gerrans, in conversation

representation doesn't capture the deep ontology of the situation, namely, that the world contains both objects and terms, both faces and age.[57]

This network may well track the aging process – but what if we also want to track another feature of a face through time, such as facial expressions ? This is probably going to require another network, and the troublesome part is going to be co-predicating the two sorts of changes across both networks to the one individual. In both networks, the assignment of change to the one individual is implicit, since the network primarily tracks changes in properties. It is hard to imagine, then, how two implicit assignments of properties to an individual can be co–assigned to the same individual across networks.[58] Co-reference even to the one individual might be a problem for this model of representation. How do we know that two related activation states are representing change in the one object and not two slightly different objects? In general, Pinker's reservations about identifying objects by their properties seem to apply, here.

We can note that while the middle layer activation states in Cottrell's face recognition network are semantically unstructured, at least in any understandable way, there are also connectionist networks where the middle layer activations states are taken to be semantically structured. Ramsey (1992 p265) makes this distinction using the Rummelhart et al's 'room schema' model where individual nodes in an activation state have specific semantic content- such as 'bed' or 'carpet'.

We might think that such activation states might represent different individual rooms depending on which nodes were activated most- eg, the system could represent 'the room with the big bed' versus 'the room with lots of cupboards'. However, such solutions run up against the problem of particularity when two objects have identical features. To use Pinker's example, when two identical chairs swap position in an

---

[57] Gerard O'Brien, in conversation

[58] Horgan and Tiensen might argue that in a network such as Bergs's processor or an Elman network, a middle layer activation state might contain something like predicate and name components smeared in superposition. The terms and predicates, in their version of a language of thought, may well exist independently in input and output states but their combination is superpositional. Horgan and Tiensen may consider superpositional processing to be what happens after naming representations have done their stuff. That is, they make a claim for identity via intrinsic syntax just for input output representation identity. However, I have argued consistently that such representation of structure, in this case the structure of objects and their properties, cannot do the work demanded of them.

empty symmetrical room we still perceive that something has changed (1997, p 115). Pinker argues that a measurement 'of the coexistence of features' doesn't seem to capture the oneness of the object these features qualify. For example, 'we can represent vegetableness [by microfeature activation]... but not a particular vegetable' Pinker asks, how could two particular vegetables be represented? Presumably not by making the relevant nodes twice as active (1997, p115).

Anyway, as Pinker points out, it seems unlikely that an individual might be represented as a very specific subclass, identified by its specific collection of qualities. It certainly would be a laborious way to perceive individuals - to have to token all their qualities or enough until individuation fell out.

Pinker offers the old fashioned representationalist solution to these problems- something must collect an individual's features:

> we are back to arbitrary labels for memory slots, as in the despised digital
> computer! (p 115).

Alternatively, in a network, 'one could give each individual [chair, for example] its own unit, or give each individual the equivalent of a serial number, coded in a pattern of activation units.'

Either way,

> [t]he moral is that the networks of the mind have to be crafted to implement the
> abstract logical notion of the individual, analogous to the role played by the
> arbitrary labeled memory location in a computer. What does not work is a pattern
> associator restricted to an object's observable properties... (p 116).

Pinker is referring to networks like Rummelhart et al's room schema, but I think the same moral might apply to unstructured representations like those in Cottrell's face recognition network. Although properties are captured in a different holistic way in

those networks (the preferred stimulus of each middle layer node is not a simple property), basically it is still the case that properties, only, are being captured. In contrast, according to Pinker, a digital computer might track which individual it's dealing with because it knows which memory location it got its information from, or the name of the individual- these two possibilities amount to the same thing. This is organising via names- names of one thing or another, either objects or locations in memory.

In summary, names or labels, of some sort, seem required by a system in order to track change, or rather constancy in change. The value of such representations lies in their constancy, over time and in combination. In new combinations they provide for the representation of change and constancy at the same time. Perhaps names might struggle to qualify as representations on low road criteria such as Markman and Deitrichs', in virtue of being only minimally informational. However, even were it the case that such 'organisers' were not informational about the world, they nonetheless might help the brain to deal with particulars in the world. On those grounds, they would count as (referential) representations for a role theorist like Dennett or Newell. I propose to consider then as non-resembling, low road representations- not forgetting that they may also qualify as representations as referring 'cold indicators' (see chapter 4).

Tracking in analogical reasoning

Dennett (1996) makes a case for labels, or names, in an internal kind of tracking, in problem solving. However, he thinks these labels come from without, from language and culture. In this section, I appropriate his arguments for labels, in general, and ask why his considerations shouldn't apply to a mental language?

Although it may seem a bit worryingly high road, according to the classical model of cognition, tracking individuals and their properties is propositional representation, using representations of the form P(x). As we've seen, the tracking, or detection, of properties by itself doesn't require such basic sentential structure. The same may

apply to internal tracking or organisation, but I will argue that nonetheless it requires something like names.

In his 'Reflections on Language and Mind', Dennett makes a case for words, or internally represented words, as facilitators of intelligence. In fact, Dennett outlines an interesting role for arbitrary-form representation in general.

> For the moment, set grammar (and logical form) aside and think of words just as isolated objects--images or labels, not parts of sentences obeying the relevant rules (1996, p 5).

In discussing the cognitive power language offers a mind, then, Dennett wishes to focus on labelling and naming and not on combination (and probably not composition[59]). He argues that words act as 'labels' that allow complex sets of processes to be delineated and patterns of such processes to be made apparent or accessible. Clark (1997, forthcoming) makes similar points about the virtues of representations *qua* objects, namely, that mental objects, perhaps words, allow higher-order and abstract relations to be recognised between the objects themselves.

Accordingly, and according to Dennett, we 'don't have to tie the good idea of inference to logical form'. Predicate calculus representations might have transparent logical form and are thus easy to manipulate in a logical fashion- but Dennett is warning against an internal implementation of such a formal system. He claims that people hardly ever state arguments fully and explicitly and we don't know if unconscious cognition takes place by symbol manipulation, either. Hardcastle (1997) puts it this way: human reasoning probably does not progress in a step by step manner even though public versions of it may do so.

We don't have to tie inference to logical form, says Dennett, because we have another model of how representation may contribute to intelligence that makes use of mental objects:

---

[59] Unless the co-tokening of labels counts as a kind of loose composition.

> The thought processes exhibited ... (by Copycat, by the chimps [who solved certain abstract problems]) are familiar human thought processes, and they are not logical arguments; they are (roughly) processes of competitive concept-building.

Competitive concept building, then, is conterposed, by Dennett, against formal inference. 'Progression' in thought is not concatenated representations being 'read' somehow as well formed formulae and this, in turn, causing logically related formulae to be 'written'. Thought, or its implementation, is more like a distributed flow of representation appearances and disappearances- where the order of appearance, or activation, is not determined by the internal syntactic structure or logical form. Instead, semantic relatedness is implemented in the tacit knowledge of the system, or the causal powers of the consumers of these representations. However, the process relies on discrete symbol-like representations.

In the above quote, Dennett refers to Copycat- Hofstadter et al's model of analogical reasoning. I will briefly describe (some of) Copycat in order to illustrate how representations might interact in a manner that doesn't rely on logical form. But I will also argue, Copycat, nonetheless, makes good use of name-like representations- and in a manner that it not consistent with a resemblance semantics.

Copycat approaches analogical reasoning by building something in between structural models and structural descriptions of the objects of the analogical reasoning task. The objects in question are letter strings. The analogical task is to map the relationship between a given pair of strings onto a third string so as to produce a new pair, analogous to the first pair. For example, Copycat might be asked: what is to 'ijk', as 'abc' is to 'abd'? The terms of the descriptions are label-like symbols, though Copycat's architecture is not 'symbolic' – it has a parallel architecture that includes a long term memory called the 'Slipnet', which looks something like a localist connectionist network. The slipnet contains Copycat's conceptual 'Platonic' types, such as 'letter', 'successor', 'leftmost' and 'triple'. Tokens of these types appear in a working memory, called the 'Work Space', along with the letter strings themselves. Within the Work Space, competing structural descriptions are built of letter strings using these symbols. The symbols attract little builders called *codelets* which more or

less implement production rules of a sort, eg if two adjacent letters have the label 'same', attach the 'group' symbol.

Gradually, descriptions of letters feed into descriptions of letter groups, and visa versa, until a description of the entire letter string is built. For example, three adjacent letters, eg, 'abc', might come to be described as containing two *successors* and then as constituting a *triple*. Within Copycat, different mappings and partial mappings of the letter strings compete at the same time; hence, the competitive concept building, Dennett refers to[60].

I have three points to make about representation in Copycat, as described so far, that illustrate points made in previous chapters:

1. Obviously representations like 'successor', 'group' or 'letter' in Copycat are not isomorphic to the relationship of 'succeeding' or 'being a group'. The use of Cummins' terminology of 'targeted structures' would be okay, here, except that the structures in question are *described* and not *shown*. For example, in the model, 'successor' is a name, not an isomorphism.

    Of course, the second order isomorphism card is always available; Copycat's whole conceptual scheme might be isomorphic to the domain of letter string structure. However, as I argued in previous chapters, it would be nice for the theory if representations like 'successor' were themselves structured given that it and similar representations are significant representations in the scheme of things- and of the sort that we are trying to account for in our theories of representation. A theory of form and content is not that appealing if it rarely applies to the concepts or representations we are interested in.

---

[60] Given the question, what is to 'ijk', as 'abc' is to 'abd'?, Copycat might conclude that the answer is 'ijl'. This would follow if 'abc' and 'ijk' were targeted as 'triples', if 'd' were targeted or mapped as a 'successor' to 'c', and if the 'successor' concept was applied to 'k' in 'ijk' to produce 'ijl'. There is not much competition between structural descriptions, but in other cases the competition is more serious. For example, what if Copycat were asked 'what is the analogous change from 'aabc' to 'aabd' for 'ijkk'?'. It might 'question' whether the 'successor' function, which it detected in use in the example pair, should be applied, in 'ijkk', to either the rightmost letter 'k' or the rightmost group, namely, 'kk'. In Copycat, this question is fought out by the representations 'group' and 'letter' according to biases in the system that favour some analogies over others.

We can also start to wonder if any larger patterns made out of our non-resembling representations play a decisive role, anyway. The generation of solutions may reflect a background compositional or functional isomorphism between the world of letter strings and the representation of those strings, but perhaps that is an emergent feature of Copycat's workings and not a driving force behind those workings. It just doesn't seem to be the case that Copycat's whole conceptual structure is being targeted at anything on the occasion of every tokening of a representation in the Work Space and, as such, is being used as a representation. This is related to the point, above, that non-resembling representations seem to do causal work, whether it is in virtue of any relational content or not. In their node-like form they are able to usefully concentrate a content, even though they may refer in virtue of being linked to a set of operations implemented beyond their boundaries.

2. The representations in Copycat represent just the kind of abstract properties, eg, 'successor-hood', that representations of arbitrary form might be thought to usefully capture and make available for higher order abstractions. Having them represented in a context neutral way seems to be the key in Copycat – by that I mean, that their form remains constant as they take part in different descriptions or form different relationships with other representations. Indeed, their relationships with each other are actually designed to change- as we shall see, the connections between representations in the Slipnet slip or tighten as their importance to a mapping alters.

3. Objects in the task domain, eg, 'abc' and 'ijk' etc have the very type of intrinsic syntactic structure that the classicist has in mind and they are represented as such by Copycat using concatenated symbolic representation. Each component of the letter strings and their description is discrete and identifiable by its syntactic properties. Copycat depends on having the representations of the relationships of order and grouping between letters discrete and recombinable.

So, as Hofstadter admits, Copycat has some symbolic properties. However, one feature that is *not* language-like is the absence of predicate-term parings. As we noted, Dennett is impressed by the way Copycat builds complex structural descriptions via processes of parallel activation and inhibition of its explicit representations and the parallel activation of procedures that consume and produce those[61]. In some ways, then, Copycat runs on 'symbolic gestalts'- the choice of symbol-like representations used to build descriptions in the Work Space is determined in part by the changing links between the entire set of representation types in the Slipnet, Copycat's long term memory.

The gestalt-like spread of activation in the Slipnet is partly mediated by a measurement of the 'conceptual distance' between concepts, which itself can change as Copycat searches for new conceptual relationships. Closely related concepts such as 'letter' and 'group' tend to activate each other more easily than less closely related concepts would, such as 'predecessor' and 'successor'. However, these influences are not fixed. Were, for example, the deeper concept[62] 'opposite' to be strongly activated, 'predecessor' might, in fact, 'slip', as it is termed, into 'successor'. For example, instead of the command *replace the leftmost letter with its successor* coming into play, Copycat might replace the leftmost letter with its *predecessor*. In this way, Copycat illustrates a pleasing fluidity as conceptual shifts permeate through the entire system.

But having acknowledged a parallelism in Copycat, I note that Copycat relies crucially on labelling. Not only are concepts named, but so are the links between concepts, which are themselves concepts. Hofstadter explains how labelling is essential to the kind of conceptual slippage we referred to above- in Copycat's long term memory, the Slipnet,

---

[61] Composition via inference still lurks perhaps, but it is a loose kind of composition which never takes place as concatenation but as spreading contemporaneous activation. Composition is built in, in a sense, in the links between concepts.

[62] Copycat uses a range of competing representations of different 'conceptual depth'- which means something like 'saliency in the letter string world'- the less salient a property is, the deeper it is. Conceptual depth is tacitly represented in Copycat's organisation, i.e., deeper representations unite less deep representations and take longer to decay or de-activate once activated. The activation of nodes standing for deeper representations then feed back to the activation of nodes for shallower concepts.

there are a variety of *link types* and for each given type, all links of that type share the same *label*; ... each label is itself a concept in the network; and ... every link constantly adjusts its length according to the activation of its label. ... If concepts A and B have a link of type L between them, then as concepts L's relevance goes up (or down), concepts A and B become conceptually closer (or further apart). ...An example of a label is the node *opposite*, which labels the link between nodes *right* and *left*, the link between the nodes *successor* and *predecessor*, and several other links. If the node *opposite* gets activated, these links will all shrink in concert, rendering the potential slippages they represent more probable (p 214).

This feature of Copycat certainly illustrates Clark's point about layers of abstract property detection being facilitated by name-like representation. For example, the relationships between 'right' and 'left' and 'successor' and 'predecessor' are usefully connected to the more abstract representation 'opposite'. Dennett's 'labels-as-trackers' theme is also well served,

In order to engage in these processes [of competition], however, one must be able to keep track of the building blocks, and tracking and recognition are not for free. Our concepts are clothed in re-identifiable words for the same reason the players on a sports team are clothed in uniforms of the same *familiar* color so that they can keep track of each other better... (Dennett, ibid, p 6).

Though Dennett is referring to words as the uniforms in question, presumably the same benefits of delineation and co-ordination would accrue to a system using word-like *mental* representation.

Dennett also makes the point that better concepts captured by better names would make for better structural descriptions and hence better analogies. If Copycat, for example, could capture the concept 'repair' with a label, then Copycat might come to 'see' the equivalence of the following pairs:

abcdjf → abcdef; ppypp → ppppp.

In this analogy, two very different strings on the surface are understood as sharing the same abstract property of 'being in disrepair'. Whether abstract concepts are just representations or not, Copycat helps us see how they might be put to use if they can be captured by a name-like representation. Although 'repair' will need to be integrated into Copycat's functional structure, it seems to be a useful tool in itself, qua name, in advance of its position in any larger functional system.

What makes the role of these representations in Copycat *qua* nodes, or names, more salient is that the wiring facts between them, in the Slipnet, can alter according to the activation of other representations and yet they remain, in an important sense, the same representation. However, although representations are node-like in Copycat, Hofstadter talks about conceptual clouds:

> Although it is tempting to associate a concept with a pointlike node, a concept is better identified with this probabilistic 'cloud' [of link revisions] or halo centred on a node and extending outwards from it with increasing diffuseness. As links shrink and grow, nodes move into and out of each other's halos (to the extent that one can speak of a node as being inside or outside a blurry halo) (p 215).

Hofstadter considers a parallel, here, with connectionist distributed representations:

> In distributed systems there would seem to be halos, since a concept is equated with a diffuse region, but this is somewhat misleading. The diffuse region representing a concept is not explicitly centred on any node, so there is no explicit core to a concept, and, in that sense, no halo. But since slippability depends on the existence of discrete cores, there is no counterpart to slippability, even in distributed connectionist models (p 215).

This importance of an identifiable centre nicely captures the problem that distributed representations have with maintaining identity as they slide into each other. Lacking an inner core, they struggle to represent change in an entity without representing another entity altogether. Copycat, on the other hand, achieves this by keeping its representations node-like.

Hofstadter takes a half way position on the structure of concepts:

> We [Hofstadter and Mitchell] believe there is a sub-cognitive, supra-neural level at which it is realistic to conceive of a concept as having an explicit core surrounded by an implicit emergent halo. P216

Hofstadter, then, like Cummins wants concepts to be bigger than single representations. Even so, the 'explicit core' he refers to looks a lot like a symbol and Copycat won't work without them. The thing about names or labels is that they maintain their syntactic and semantic properties in different combinations. Although there isn't concatenation in Copycat, there is contemporaneous activation of a multitude of representations and implementation of the type-token distinction. As a computational model of reasoning, Copycat makes a good case for the organisational power of labels. My question for Cummins is, then, surely this kind of organising is useful work for a representation to do- even if, it turns out (and I'm not sure it does) that some of the semantic properties involved are role properties?

This brings us back, of course, to the question of what type of semantics are consistent with the notion of a representation. Dennett is adamant that all that matters in a case like Copycat is what the symbol-like representations do for the system or the system does with them. He asks,

> how could a brain's central system merely writing Mentalese in itself count as thinking? How could that do any work, how could that guarantee understanding?

Dennett answers his own question…

> Only by enabling something. What? Enabling the multitudinous items of information that are variously distributed and embedded around in the brain to influence each other, and ongoing processes, so that new, better informational structures get built (usually temporarily) and then further manipulated. But the manipulanda have to manipulate themselves (ibid, p 4).

For Dennett, it's what a representation does that's important- not so much what it 'says', in virtue of its intrinsic properties. It's not surprising then, that Dennett should see content as a matter of role and not a matter of the properties of the role player. We might point out here, that, Dennett, here, is also extolling the virtues of certain intrinsic properties for representations, namely, symbol-like identifying physical features which we might think counts as part of an explanation of intelligent behaviour. So, for Dennett, these must count as useful features, but not semantic features.

This provides us with another way of describing the disagreements between Dennett, Cummins and Fodor. Fodor and Cummins are also not arguing that just the tokening of representations explains understanding, or cognition. Obviously, for Cummins, just to begin with, there is the process of applying a representation to a target. But Cummins and Fodor *are* arguing that representations help explain cognition because of their semantic properties. Cummins thinks these are intrinsic to the representing vehicle and Fodor at least thinks they are 'pre-use', or something a representation brings to a system. The whole dispute, in part, comes down to whether a representation's useful properties have to be semantic properties. If they do, then a symbol's defining features count as semantic properties, which Cummins thinks is absurd- leaving resemblance representation the only game in town. (I tried to show in chapter 4 that it may not, in fact, be entirely absurd) If, on the other hand, not all a representation's useful (pre-use) features have to be semantic features, then it is easier to make a case for non-resembling representations. If none of a representation's (pre-use) features have to be semantic properties then it is even easier.

As for the tracking argument, Copycat may provide some reason to consider the possibility that the tracking of particulars in the world does not involve propositional representation. Although, that said, Copycat serves more as a model of reasoning (concerning the properties of static objects) than of the cognisance of change. In any case, it provides further support for the idea that name-like representation serves a useful role in cognition.

An alternative approach to tracking: effects don't have to reflect their causes

In the section above, I considered a mapping process which relied on representational uniforms- abstract properties in a domain were mapped using discrete, syntactically identifiable labels that helped the mapping system keep track of its own mapping. Before that, I considered the mapping of individuals and their properties over time and called that mapping 'tracking'. In both cases, symbolic representations seemed required.

What reason could we have for *not* positing such 'clean' representings of things and properties as Pinker, Caruthers, Botterill, Horgan and Tiensen (though not in combination), and Markman and Dietrich suggest? We have been relying on several sources of evidence: our behaviour *vis a vis* objects and their properties, our ability to reason about objects and their properties, as in analogical reasoning, and our experience of understanding the world as populated by things and their properties. Concerning the latter two, someone like Dennett would argue that words are the only symbols required to explain this higher order thinking[63]. But what about our everyday

---

[63] Perhaps language also helps us track the world. Symons (2001, M&M) wants to have words as the only representations in the game. He argues we should ignore the inference from compositionality and systematicity in cognition to any 'inner cause theory'. 'Cognitive systems are responsive to representational systems in appropriate ways without themselves being representational systems' (p. 538). He uses, like Dennett, Quine's elephant bush metaphor- there are lots of ways of achieving the same behavioural shape; the art shapes the overall pattern, not the details: ' We can recognise that cognitive systems exhibit the correct pattern of behaviour in response to the environment, without thereby assuming that an internal mechanism is responsible for this systematicity' (p 538).

The weight here is on the systematicity in the language environment the human brain develops in and comes to master. The representations that drive systematicity are in the world not the brain- they are words and signs (p 523). The self-organising dynamics of the brain and world apparently implement Quine's social art theory of language. The art shapes the student in his own way to the desired proficiency.

Thus, in the case of tracking the partition of the world into objects and their properties, language shapes our skilled behaviour towards them with a little help for natural selection. Tough luck for chimps then.

So, if theorists like Symons and the supporters of emergence have a point then the tracking we see in the world, in our understanding of a football match, may not map back even to sub-personal trackers- homunculi that track. This is to deny even Cummins' notion of an intender- since Simons are denying representation they are also denying the consumers of representations.

It really could get that bad, at least according to Clark. Relying on functional decomposition as a strategy for understanding cognition may itself be problematical. Clark (2001) suggests we ask ourselves: can we be sure that cognition is fully and essentially decomposable into jobs or tasks?

physical interaction with the world? Should we also be wary of reading too much of the nature of our behaviour into the nature of the brain states that make that behaviour possible. Perhaps the only object-like states involved in tracking behaviour are the tracked objects themselves, much like some argue that the only word-like representations are the ones that exist in books and the soundwaves that surround us.

Dennett's concern, here, (eg, in 'Real patterns') is with reasoning from product, i.e., behaviour, to the means of production, i.e., cognition. The worry is that we cannot assume from the nature of the patterns in behaviour that the internal workings responsible for that behaviour are similarly structured. Wittgenstein's influence on Dennett is clear here, 'But why should the *system* continue further in the direction of the centre? Why should this order not proceed, so to speak, out of chaos?' (Zettel, 1981. 608. Quoted in Davies, 1981)

 In *Real Patterns (in his 1998)*, Dennett splits the distinction between rule following phenomena, eg, as in planetary motions, and rule consulting phenomena, eg, as in computer software, with a third sort of patterned phenomena. He describes a pattern-like behaviour that is,

> preserved under selection pressure: the regularities [are] dictated by principles of good design and hence homed in on by self-organising systems (1998, 112).

Pattern, then, can be a 'statistical effect of vary many concrete minutiae, as if [shaped] by a hidden hand, an approximation of the 'ideal' order' (1998, p111).

This is Wittgenstein with an evolutionary gloss, I suppose. Except, Dennett is not claiming that the order in question arises from chaos but from a multiply factorial complexity, which the order in the behaviour belies.

Secondly, are we then committed to a parallel physical decomposition of the brain- are these jobs, or functions are separately carried out by different parts (or aspects) of the hardware?

Clark raises the possibility that instead 'jobs' are shared by physical parts and moreover in a manner that denies the possibility of functional and algorithmic decomposition.

So is Clark saying the jobs disappear? That would be a high price to pay, since computation relies on there being specific computations to compute. Are the jobs messier than we thought? If so then does the same hold for the representations- either that they are messier or they don't exist?

Dennett is concerned here, in particular, with Fodor's representational account of beliefs, rather than tracking, per se- namely, Fodor's contention that the patterns discerned in folk psychology should be taken as reflecting patterns in the brain (Dennett, 1998, p111). The tracking arguments I have presented here have not been necessarily concerned with such high road cognition as beliefs states, but Dennett's argument might still apply.

To motivate his distinction between emergent and specified patterns, Dennett describes two such ways of generating an actual (bar code-like) pattern[64]. The main point of the exercise for us to note is that, in the generation of one of the patterns, not one component (i.e., one dot- except perhaps the mean dot, see footnote) of the pattern is specified or represented, in any way.

Dennett makes use of the patterns in Conway's Game of Life to make similar points about emerging patterns that don't reflect their 'emergence base', so to speak. In the flashing squares of the Game of life, patterns at a higher level of description, such as the 'glider' pattern, have an ontology that seems unconnected to the square world 'below' them. For example, at the level of the individual square there is no motion, no complex shapes and no reference to any part of a 'glider'- nor is there in the algorithm that determines the 'on' or 'off' state of the squares. The glider pattern is analogous to Dennett's emergent pattern, above- both consist of an order of sorts that is the result of an 'unconnected' order. Is the same true about our tracking behaviour and the machinery behind it? That is, is there no obvious connection between the directedness of our behaviour *vis a vis* particulars and its underlying machinery?

---

[64] The first pattern is produced by a simple algorithm which more or less specifies the length of the bars bit by bit (i.e., it is a bit map with a bit of noise thrown in). In contrast, the second set is not produced by anything which represents the patterns, but by using a normal distribution of dots around some spaced points. The points act as means for the distribution of the dots and are spaced so that a bar code type pattern is produced. When the second set of patterns are tidied up by a simple contrast enhancer algorithm they take on the same look as the first set of bar code patterns (especially the less noisy).

Dennett also claims that there may be real indeterminacy about which sorts of patterns really exist in the world, including the brain?[65] Again, could this indeterminacy be true of the tracking of human beings? Just as Dennett thinks there are patterns in human behaviour that can be described in various ways, using different belief desire stories, could the tracking we see in animal and human behaviour be similarly multiply described?

So, could it really be true that we don't know *what is being tracked by what* in everyday 'deal-with-the-world' animal and human cognition? At one level of description, things look pretty simple- the footballer is tracking the ball and the other players, the lion is tracking his prey, the monkey is tracking the gaze of his rival. Surely these descriptions are capturing something about our cognitive states.

Sterelny, as we saw, raises the possibility that the psychological states of others may be tracked obliquely in a species. It may not be the case that the unity, or oneness, of the entity being tracked should be reflected in the unity of the tracking representation, i.e., that there is one representation for each thing tracked. Actually, what Sterelny calls 'robust tracking' involves tracking something via detection of several of its properties, as compared to detecting something via only one of its properties. Indeed detecting an object doesn't necessarily count as representing an object, on Sterelny's account. But perhaps the tracking in dispute, in this section, is what Sterelny would call tracking via *multi-purpose, or decoupled* representations- that is, representations that don't have immediate, single, effects; but stand for things without making, anything in particular, necessarily happen.

For theorists like Dennett or Clark, the tracking is more likely multi factorial, involving many partial representings and detectings of things and properties[66] - also involving 'representings' and 'detectings' that are never combined in a simple

---

[65] His position is 'different possible patterns and no truth of the Matter' as compared with Davidson's - 'different patterns about same underlying reality eg, 0 degrees Celsius and 32 degrees farenheit'p114

[66] As we noted above, the embedded camp provides examples of intelligence in relation to objects in the world 'falling out' from less representation of the object than we might think- eg wall following behaviour without representations of walls, and ball catching behaviour without trajectory calculations (it apparently turns out that outfielders track ball trajectory by tracking/maintaining the angle of their own gaze at the ball (Clark 1997)).

'subject-predicate' format. Perhaps, for example, tracking is a process involving many (possibly superposed) activation states. Each activation state might hint at change in a particular direction for a property or even a combination of properties (perhaps in 'as if' composition). With many of these states feeding into behaviour and with the world feeding back into those states, via constant sensory updates, unified directed behaviour might emerge. Although it may be hard for us to imagine how, the whole system will be self-organising, shepherded by development and evolutionary processes.

In response, I think the earlier sections of this chapter should remind us of the weaknesses of the kind of representations involved in Dennett and Clarks' picture, especially as a description of sub natural language mental representation. That is, there seem to be limits to feature representation only. There seem to be limits to continuously varying representations as opposed to multiple, discrete representations related to co-reference and co-predication across change. There also seem to be limits to relying on special purpose representational schemes that only track in a given context. Ignoring those apparent limits and calling a system 'self-organising' might just amount to a way of not explaining how organisation, or pattern, in cognition occurs. The classical view of a productive and systematic representational scheme may lack imagination, but it gives us some indication of how tracking might work. Perhaps a Rodney Brook's type robot might show otherwise in the future, but it will need to do a fair bit more than merely track walls.

Until we can put more meat on the bones of the notion of tracking as a self organising system, the tracking argument can be seen as an inference to the only explanation, or to the only explanation we can make much sense of. In the end, it is only an inductive argument. The skill of tracking may not reveal the presence of mental names or stand-ins in the processing that underlies the tracking. But here are two more reasons why it might: firstly, inference to the simplest explanation might do some work for the classicist here. Inferring from product to process is simplicity itself. Clark (2001), himself, seemed to express this intuition when he argued that a neo classicism was inevitable given that the world itself was neoclassical. Dennett may worry about

cognitive wheels[67], but on the other hand, wouldn't a representation system with identifiable names seems a pretty obvious trick for nature to latch onto.

Perhaps there is room for a Dennettian reply along the lines that simplicity is actually less likely than complexity - nature tends to be more tricky than simple, simple is a top down sort of thing etc, simple building blocks don't lie around waiting for nature to put them together. Familiar Dennett moves. But how do we reconcile this with the simplicity of the explanation - we make thoughts by putting bits of thoughts together, and we track things by using re-identifiable bits of representational stuff to stand for things and their properties? As we saw with Copycat, using uniforms to keep track of a system's own tracking is a good idea.

Secondly, Dennett and Clark both make great play over the usefulness of the object like qualities of words. In their stories, words, *qua* internalised mental objects turn out to be very useful in an internal kind of tracking and organisation. My question is, how would Dennett justify restricting his own tracking argument to just the organising properties of words? If labels are so useful, why would nature wait around for public language to arise to hit on this good trick? Clark and Karmiloff-Smith make this kind of argument against Dennett by pointing to stages in child development that precede the use of words. As we saw, they argue for a pre-lingual 'objectification' of skill, in the sense of making knowledge object like- at least that seems the direction 'representational redescription' seems to take us.

In summary, the dynamicists and logical behaviourists have work to do if they are to treat concepts merely as abilities and abilities as intentional but emergent properties of minimally representational systems. The representationalists have more obvious

---

[67] Dennett (1984) worries that classical AI stories that contain symbol talk might be proposing a whole bunch of what he calls cognitive wheels. Cognitive wheels are things, like wheels, which nature is unlikely to have used in evolution because of engineering constraints- how could an axle system get up and running, for example? (Although there are some borderline cases of wheels in nature, apparently.) In particular, symbolic and language-like mental representational schemes, with their compositional semantics, might not be likely to have evolved. In any case, for Dennett, what theoretical entities in cognitive science should do is give us some idea about the processes behind our introspective common sense awareness of our minds. In Dennett's mind, a symbol system is unable to explain the common sense idea that conscious thoughts often seem to be relevant to the problem at hand.

weapons in their armoury- including, importantly, the symbol-like representations of the classicist.

Chapter 8. How resemblance could drive cognition

Weight State Semantics

In this chapter, in spirit of the pluralist approach to representation taken in this thesis, I describe another account of resemblance representation. In doing so, I also enlarge on the notion of representation as map. We might term this approach 'Weight State Semantics' and characterise it as the synchronic version of the diachronic resemblance in State Space Semantics.

O'Brien and Opie have recently argued that the place to look for the structure in a network is not in its activation states, or the state space they describe, but in its weight states. The representational truth of connectionism is more likely to be that the net itself is isomorphic to its domain (see O'Brien 1999). For instance, consider that in Cottrell's face recognition network, two activation states, that we might take to represent two faces, may be similar to each other in physical structure in a related way to the faces in question (given the discriminations that the network has been trained to make). If we take this relationship between the activation states as a resembling structure, then it is the diachronic resembling structure of a multi part representation that is never instantiated in full at any one time, since the network is only capable of one activation state at a time. On O'Brien and Opies' account, this structure, between representations, only exists because the network, itself, embodies a model of human faces (or a subset of faces) synchronically. The real model is in the network itself. This is the resembling structure that is put to use, and its use is reflected in the diachronic structure we find between activation states.

To show this, the first task is to find structure in a network's weight system and then to find a correspondence between that structure and the network's target domain. Having done that, we may be in a position to make a more intuitive case for causally efficacious and explanatorily powerful content as resemblance.

I will very briefly give some indication of how O'Brien and Opie think this might be possible. Although, when represented as a vector, a weight state looks like a list,

within the network, itself, we can immediately distinguish between the 'fan out' connections that leave each input node and terminate across the middle layer and the 'fan in' connections that originate across the input layer and converge or 'destinate' at each middle layer node. Thus, for any network with 'n' input nodes and 'm' middle layer nodes, there will be 'n' fan out structures of length 'm' and 'm' fan in structures of length 'n'. The task then is to find some abstract structure either within each fan or between fans that can count as resembling the network's task domain. Without going into the details, O'Brien and Opie have claimed to do just that for a network that solves a colour discrimination problem- that is, they claim the network represents the structure of various light waves with a resembling structure of its own. The structure is claimed to be found in the fan in structures (represented as vectors), each of which (i.e., each fan) resembles one category of input waveforms.

Whatever the details here, the thing about the structure of an actual network, as opposed to a state space or the activation states that a state space represents, is that it exists completely in any one instant. It is much more tool-like than a state space- it is more like a real map- because it is applied to the input rather than being the result of the input being applied to the network.

Let's look at this tool analogy more closely. A consumer of a map can apply an input to the map, or visa versa, so that a *current position* representation is incorporated into the map. Something useful might then happen in accordance with what the map says about the current input. When O'Brien and Opies' network receives a colour input and the input is channelled through the weight structure, this may be taken to amount to the application of the map to the input, or the integrating of the input into the map. We can read the resulting activation state as a point in map space equivalent to the 'you are here' arrow, as Churchland does, or, perhaps better, we can read the activation state as something useful already happened as a result of the coming together of the map and the input. That is, the activation state's position in state space just tells us what useful thing the map did when it came in contact with the input, that is, when the input vector was multiplied by a weight vector. The real map, the one doing work, is not the state space but the weight state and it has already done the work by the time the activation state is produced.

This account fits with Cummins' auto-bot example. In effect, when the auto-bot's wheels pull the card through the car they are asking the card, or the map, in what direction should we go next? The part of the card that the steering axle is in contact with amounts to the system's representation of where the car is in the maze, for example, as in the section heading due north. This all fits well with the map reading metaphor- the car reads the map by having the card pulled past the steering axle. There is also something reminiscent, here, of the Watt Governor and the way the spindle arm might be said to resemble the speed of the flywheel and be read by the input valve. However the important thing is that, in both cases, as in the network case, the map, or resembling structure, automatically produces an answer when it comes in contact with an input. So, it looks like the map is doing causal work according to its structure in the act of being applied to an input. In the network case, the resultant activation state is then consumed by the next set of weights as the overall task of the network requires.

Should we be suspicious of weight states as representations?

In chapter 1, I left the representational status of weight states uncertain. I noted that they partially meet Markman and Dietrichs' notion of a representation in virtue of mediating in a purposeful way between input and behaviour, but, also, that they failed to play the informational role that the authors saw as typical of representation. It is certainly true that weight states are slow, gradual changing representations, but that makes sense for a map of something that doesn't change very quickly. *Qua* resembling devices, they don't represent constant change in the way, for example, the angle of spindle arm might resembles the change in speed of the fly wheel in the Watt Governor. That said, weight states are not fixed either- a network can always learn more about its domain.

I also noted that the notion of the consumption (or interpretation) of a weight state was unclear- in some ways, the weights themselves seem to be the consumer systems in the network. *Qua* maps, the answer, here, might be that a cognitive system, or the network as a whole, consumes a weight state when it inputs a problem into it and

takes note of the answer. I explain this in more detail below in considering Ramsey's objections to weight states as representations.

We can also note that it is not easy to imagine the easy combination of weight states, even as Millikan imagined the overlaying of maps to produce new information. This may be a failure of imagination, but it needn't rule out weight states as representations, in any case, since, as I claimed in chapter one, such combination needn't be considered a necessary feature of a representation (though it does suggest limitations for this kind of representation). I also noted that the application of different kinds of inputs to a weight state might be considered as different combinations of representations- in this case, representations as input and representations as knowledge.

We can remember that Ramsey considers two reasons to posit representations; either representations have contents that push cognition around, or representations reveal the functional organisation of the cognitive system. As for the second criterion, according to Ramsey, connectionist networks don't do much for task decomposition for the simple fact that connectionist models of cognition are not algorithmic (1997, p 53). (They don't break a task into subtasks. This is reflected in the trial and error ascent to competence in training. For MacLelland and Rummelhart, for example, a single settling of a network into a particular state was never meant to amount to a series of computational steps.[68])

It seems true that the standing knowledge of a network would be in the weight structure of a network, if it were anywhere – after all, that's where learning has its effects. According to Ramsey, typically with knowledge in classical systems 'causally distinct structures encode commands for specific stages of the computation' (1997, p 58). Whereas, Ramsey claims, in connectionist networks, 'there is no level of analysis at which we can say a particular weight encodes a particular command or governs a specific algorithmic step of the computation' (p 48). This is, after all, what

---

[68] Indeed compression, or, doing everything at once, has been advertised as a strength of connectionism- compression gets you out of brittleness and potentially, some have claimed, even the frame problem. Brittleness is avoided because the failure of one step doesn't hold up other steps, and the frame problem might be avoided if all the knowledge in a network (or maybe even a network of networks) is accessed at once.

you get with superimposed 'knowledge'. So, there is no heuristic value in understanding weight structures as knowledge structures, networks are not rule following.

Haugeland (1991, p 86), considering weight states as know-how, asks

> Does it even make sense to regard the embodiments of a system's abilities or *know-how* as *representations*? Why not take them, rather, as just complex dispositional properties- acquired and subtle perhaps- but, for all that, no more representational than a reflex or an allergy?

According to Ramsey, if a system is a rule following system, as opposed to a rule describable system, then it must be possible for the rules to be activated separately. It must be possible for different rules to fire at different time. In a network, no part of the weight state ever operates in isolation. No parts of the knowledge are isolable because no part of the weight states structure embodies just the one rule and comes into play only when that rule is required. Consequently it makes little difference to posit representations within the weight state and the same causal story can be essentially told without them, (pp 50-51).

Marin Davies (1991) imputes knowledge of a rule to a system whenever any systematicity in input-output relations results from a causal systematicity. That is, when a system does the same kind of thing to instances of the same kind of input in different contexts. This is not to say that the system contains an 'explicit syntactic encoding of the rule that is known' that it always uses in the presence of the same type of input- 'the standing condition of knowledge of a rule can be realized just as well by the presence of a component processor...'

Would this reading of rule use render a weight state as a collection of (superimposed) rules or not? Since, every input gets exactly the same weights applied to it- one size fits all- perhaps the answer is 'no'. No rule is selectively and consistently applied to any particular problem. The analogue connectionists might claim that the one domain is dealt with by the one class of models (defined by their state space divisions). But

this wouldn't meet Davies' requirement, since we couldn't admit the whole domain to be an input- brains may deal with whole domains but never all at once.

However, we need to ask if these same criticisms of weight states as representations apply to a weight structure as maps? Surely, it is the nature of maps that they are holistic- in the sense that it is impossible to use only one part of it at a time. They are not like algorithms in this regard. If this is the case, then map talk is not task decomposition talk in the first place and so shouldn't be criticised on the grounds of not referring to the same kind of representations.

In any case, it looks like the weight structure as map may qualify on Ramsey's first ground for representation- that representation bring causally potent content to cognition. Along such lines, Haugeland goes on to make a defence of weight states as representational. He suggests that in a weight structure, some representation, not of the task at hand, but the task domain might be found. Consider, for instance, a network that allows one to recognise faces such as the face of a friend. His idea is that while no two views of a face are ever exactly alike, there is something these 'face' experiences are experiences of, even if that thing is never actually experienced - a certain feature of the environment.

> Clearly such a feature could never be detected on a given occasion; yet adjusting to it in *absentia*, as a means to correct re-identification, would be of great value. Accordingly, whatever incorporated the ability to recognize those faces could, by our account, be deemed a representation of that feature (1991, P86).

Haugeland may have had in mind a *standing in* type of representation than a resembling type of representation, though, there is no reason to believe that this couldn't be a case of standing in by resembling. In any case, the same move away from task decomposition to task domain representation is apparent.

One advantage of this idea, as O'Brien and Opie argue, is that the content of such domain maps, *qua* physical structure, seems to do causal work. A weight state's content lies in it intrinsically. So it seems to meet both Ramsey's requirement for

representation and Cummins' requirement for explaining cognition and not being explained by cognition.

Millikan seems to a have had a similar view of weight states as maps in mind, perhaps as multiple maps:

> Much of our thinking may occur in media more like maps or like models. …such models would have to be very abstract indeed, the mathematical isomorphisms between representation and world structures being far from tangible. I have already suggested that certain very abstract patterns found in neural nets could be intentional icons. Nor is there any reason why the results of superpositional storage of information in neural nets should not be considered to yield conclusions of inference (p 104).

Millikan describes maps being overlaid on top of each other, making use of shared terms, in superpositional storage; that is, in what we are calling weight structures. If feasible, this would meet the objection about combination raised above, but in any case, why shouldn't the network considered as just the one map count as representational by itself, in the absence of any combination with other maps?

Because abstract maps are such cumbersome, complex entities, it looks to be the case that the representational space that is created by a set of recursively combinable representations and the relations between them may not be available to weight state maps. But that only looks significant when one is attached to a 'use' theory of content- when it's the distinctions between representations and their uses that count, and not the intrinsic properties of the representations themselves. Once we are dealing with the intrinsic content of a representation, as would seem to be the case here then other representations don't matter so much.

Of course, there may be important roles that such weight states are not going to be suited to play, by themselves, in virtue of their physical structure, such as indication and easy combination in inference and tracking. However, this just suggests that multiple types of representation and representational content may need to be

entertained by cognitive science. It also goes to show how different particular representational roles in cognition might be.

Finally, it also seems true that the resembling structure in a weight state is going to be second order resembling structure. The pertinent resembling structure within or between 'fan in' or 'fan out' structures is likely to be relational syntax, as the term is used in chapter 2, given that the resemblance in question is unlikely to be part of a first order isomorphism with a represented domain. Is this a problem? Well, the content in a weight state may be holistic, but it is structure that belongs to the one real object- an object which holds structure intrinsically. It is structure within the one present representation and not structure between representations in time. This is a different kind of relational syntax to the one we find in State Space Semantics; it is perhaps structure that is capable of doing the kind of representational work that is required to represent structure in the world.

Perhaps, then, weight state structure does meet Cummins' vision of resemblance in representation, even though not exactly in the (first order resembling) way that representation in his examples, such as in the autobot, works. Given the strengths of Cummins' position, such as the intuitive notion of causally potent content it provides, Weight State Semantics may be a useful indication of how a kind of resemblance might be realised.

Conclusion

Mereology is the study of the relationship between parts and wholes. In this thesis, I have wrestled with the relationship between the representation of parts and the representation of wholes, including the representation of parts that find themselves in new wholes altogether, as their surrounds alter. It might be said, then, that this thesis deals with the mereology of mental representation.

In chapter 1, I provided a broad outline of what makes something a mental representation. The bottom end of my low road → high road continuum of representation allows for representations that, unlike resemblance representations, don't 'say' much about the objects or properties they represent and don't, necessarily, take part in part-whole representation at all. But given that some representation of complex part-whole relations must exist (if only because representation-combination exists), in Chapter 2 I focused on the syntax of representing parts and wholes of the sort we find in language. I concluded that it is still difficult to imagine representing such structure without using the same kind of structure in the representation itself.

There seems to be a certain compositionality in language[69], in particular, related to the promiscuous interchangeability of its parts, that requires a (physical) atomism in its representation's parts. That is, an atomism of form seems to be required where form is maintained in combination, which makes the representation of the represented structure explicit. In effect, the structure of language-like representation is actually closer to the spirit of Cummins' notion of causally efficacious, contentful structure than its connectionist rival's *relational* view of syntactic structure – since some sentential structure is fairly straight forwardly resembled in the structure of the concatenated symbols.

In subsequent chapters, I turned to the mereology of content, as opposed to the mereology of syntax- and considered how 'bits' of content might need to stand in relation to each other in abstract representation and representational combination. As in Chapter 2, I argued for some 'peer pressure resistance' in representation; this time,

more explicitly an atomism in content rather than form. I tried to make the case that the role of capturing content was sometimes best played by representations that were not tied to other representations in doing so; in this way, such content could more freely take part in new combination or use, for example, in solving new problems or tracking change. In defence of self contained content, I also argued that the need in resemblance representation to refer to a non-resembling representation's position in a larger resembling structure may not amount to actually explaining the smaller representation's importance as a representation and the properties of the representation that explain its importance. That is, in relation to the role of the smaller non-resembling representation, a larger structure may be epiphenomenal. This was claimed to be the case, in particular, in State Space Semantics where activation spaces were described as causally impotent, and, moreover, incapable of bestowing complex or resemblance content on the activation states that give rise to them.

Underlying these considerations were the questions of how content can usefully exist within a cognitive system and what qualified as a useful role in cognition for a representation to play. Against Cummins, who argues that *only* resembling content can be efficacious content, I tried to show that there may be several different ways that representation and content may play a useful role in cognition. In the first place, I tried to show that *indicational* or *covarying* content was implicit content; and, moreover, that implicit content was causally efficacious content. The argument here was that form could encode content in a representation in a useful non-resembling way for a cognitive system. This might seem to be in contradiction to Chapter 2's conclusions about the inadequacies of implicit representation of constituent structure. But the difference, here, is that I was arguing for the implicit representation of objects and properties as unstructured wholes. This kind of representation is different- more of a referring type than a descriptive or mapping type- though there seems no reason why in combination name-like representations cannot also *describe* as well as indicate or refer. In fact, the structural descriptions of Copycat suggest how this might be so and, moreover, how symbols allow for useful employment of the type-token distinction in building such descriptions.

---

[69] … and thought, according to Fodor- as we know.

My second attempt to account for non-resembling content was to make a (fall back) case for the explanatory relevance of representational content as a tool for functional delineation, and organisation. Although this functionalist approach is consistent with an epiphenomenal account of content, I also suggested that roles themselves might have causal consequences over time for the representations that fill them and that Cummins' distinction between 'non-use' and 'use' content is perhaps not as clean as he supposes. Having a foot in both the intricisist and functionalist schools of representational content may seem inconsistent, but the idea was to outline the options open to opponents of resemblance-only representation. I also raised the possibility that perhaps content might exist and work in various different ways in the human brain, including via resemblance[70].

Leaving these metaphysical difficulties behind, I then focused on some putative non-resembling roles for representation in cognition, namely, 'capturing' abstract properties, combination and tracking. As I noted above, I tried to show that an atomism in form and content might be advantageous in representations playing any of these roles. If I failed to show that some of these representing roles involve the obvious causal use of a representation's content as use of resembling representation might, then, I hope I showed, nonetheless, given my account of mental representation in Chapter 1, that the fillers of these roles have intentional properties and deserve to be called representations. If this is right, then perhaps Cummins' is wrong about the role of representation in cognition[71]. In my opinion, he makes a convincing case for one type of representation in cognition, but perhaps not all types.

I now end on Clark's (2001) point that since the world is neo classical, representation will probably be neo classical as well. His point has nothing to do with the art of ancient Greece, just with the fact that the aspects of the world have an atomistic flavour to them; that despite the infinitely many layers of structure and composition that make up the world, things seem to come as *things*. Accordingly, although our brains need to represent structure, they also need to represent things just as things; and

---

[70] Markman and Dietrich (1998) make a case for plurality in representation.
[71] He may argue that these roles are roles for 'meaning for' as opposed to 'meaning'; in which case, I would need to 'de-motivate' that distinction. In general, it also follows that if Cummins is wrong about

that means a scheme of representation where representations represent without describing and represent independently, as well as co-operatively, and, moreover, independently in co-operation. This is the strength of classical schemes of representation. Classical representations are able to represent structure without being beholden to structure for their content. This frees them to move content around, in different combinations, so as to be able to represent a very wide range of syntactic and semantic structures; it also allows them to capture constancy within change by staying constant themselves as their referents change and their own combinations change.

Of course, on the other hand, there remains the terrible possibility that my reaching this 'atomistic' conclusion could just be an example of evolutionary psychology at work- this time, at work in my own psychology. In particular, my *rigid object mechanics* module, or some such predisposition-making device in my brain, may have be firing away a bit too strongly, looking for objects in the world and in the brain. If we throw in some innate bias towards essentialist thinking, then we reach my conclusion that representations are name-like, that is, objects that to the untutored mind are seemingly essentially (i.e., magically) and self sufficiently representational - atomistic in both form and content. In any event, in this thesis, I have tried to make some defence of a representational connectedness between an 'objectness' in both the world and the brain- perhaps a kind of ironic resemblance. If this was just trying to make sense of an unavoidable bias in my thought processes, I apologise and suggest some attention be paid to the psychology of philosophy as opposed to the philosophy of psychology.

---

types of useful representation, then he must also be wrong about how to explain error and cognition, as well.

Bibliography

Aydede, M. (1995). Connectionism and the languague of thought., CSLI Technical Reports ON-LINE.

Barsalou, L. W. (1999). "Perceptual symbol systems." Behavioral and Brain Sciences **22**(4): 577-660.

Bechtel, W. (1998). "Representations and cognitive explanations: Assessing the dynamicist's challenge in cognitive science." Cognitive Science **22**(3): 295-318.

Berkeley, I. (2000). "What the #$*%! is a subsymbol?" Mind and Machines **10**(1).

Blachowicz, J. (1997). "Analog representation beyond mental imagery." Journal of Philosophy **94**(2): 55-84.

Boterill, G. and P. Carruthers (1999). The Philosophy of Psychology. Cambridge, Cambridge University Press.

Bradden-Mitchell, D. and F. Jackson (1996). Philosophy of Mind and Cognition, Blackwell.

Chalmers, D. (1993). "Connectionism and compositionality: Why Fodor and Pylyshyn were wrong." Philosophical Psychology **6**(3): 305-319.

Chalmers, D. J. (1998). A computational foundation for the study of cognition, Cogprints.

Chomsky, N. (1980). "Rules and representations." Behavioral and Brain Sciences **3**: 1-162.

Churchland, P. M. (1995). The Engine of Reason, the Seat of the Soul: A Philosophical Journey into the Brain. Cambridge, Mass., MIT Press.

Churchland, P. M. (1998). "Conceptual similarity across sensory and neural diversity: The Fodor/Lepore challenge answered." Journal of Philosophy **95**(1): 5-32.

Churchland, P. M. (Forthcoming). "Neurosemantics: On the mapping of minds and the portrayal of worlds."

Clark, A. (1988). "Tightness and constituent force." Australasian Journal of Philosophy **66**(3).

Clark, A. (1989). "Connectionist minds." Proceedings of the Aristotelian Society **XC**: 88-90.

Clark, A. (1991). Systematicity, structured representations and cognitive architecture: a reply to Fodor and Pylyshyn. Connectionism and the Philosophy of Mind. T. Horgan and J. Tienson. Boston, Kulwer Accademic.

Clark, A. (1993). Associative Engines: Connectionism, Concepts, and Representational Change. Cambridge, Mass., MIT Press.

Clark, A. (1997). Being there: Putting brain, body, and world together again. Cambridge, Mass., MIT Press.

Clark, A. (1997). The presence of a symbol. Mind Design 2. Haugeland.

Clark, A. (2001). Mindware. New York, Oxford University Press.

Clark, A. and A. Karmiloff Smith (1993). "The cognizer's innards: A psychological and philosophical perspective on the development of thought." Mind and Language **8**(4): 487-519.

Clark, A. and C. Thornton (1997). "Trading spaces: computation,

representation, and the limits of uninformed learning." <u>Behavioral and Brain Sciences</u> **20**(1): 57-66.

Clark, A. and J. Toribio (1994). "Doing without representing?" <u>Synthese</u> **101**(3): 401-431.

Cosmides, L. and J. Tooby (1999). "Toward an evolutionary taxonomy of treatable conditions." <u>Journal of Abnormal Psychology</u> **108**(3): 453-464.

Crane, T. (1995). <u>The Mechanical Mind</u>. London, Penguin Books.

Cummins, R. (1989). <u>Meaning and Mental Representation</u>. Cambridge, Mit Pr.

Cummins, R. (1996). "Systematicity." <u>Journal of Philosophy</u>.

Cummins, R. (1996). <u>Representations, targets, and attitudes</u>. Cambridge, Mass., MIT Press.

Cummins, R. (1997). "The lot of the causal theory of mental content." <u>Journal of Philosophy</u> **94**(10): 535-542.

Dalhbom, B., Ed. (1995). <u>Dennett and His Critics</u>, Blackwell.

Damasio, A. R. (1994). <u>Descartes' Error: Emotion, Reason, and the Human Brain</u>. New York, G.P. Putnam.

Davies, M. (1991). Concepts, connectionism and the language of thought. <u>Philosophy and Connectionist theory</u>. W. Ramsey, S. Stich and D. Rumelhart.

Davis, S., Ed. (1992). <u>Connectionism: Theory and Practice</u>. Oxford, Oxford University Press.

Deitrich, E. and A. Markman (2003). "Discrete thoughts: why cognition must use discrete representations." <u>Mind and Language</u> **18**(1): 95-119.

Dennett, D. (1987). <u>The Intentional Stance</u>, MIT Press.

Dennett, D. (1996). <u>Reflections on Language and mind</u>. Sheffield.

Dennett, D. (1998). "Things about things."

Dennett, D. (1998). <u>Brainchildren</u>, MIT Press.

Dennett, D. (2000). Making tools for thinking. <u>Metarepresentations: a Multidiscinplinary Perspective</u>. D. Sperber. New York, Oxford University Press.

Dennett, D. C. (1979). <u>Brainstorms: Philosophical Essays on Mind and Psychology</u>. Hassocks, Harvester Press.

Dennett, D. C. (1984). Cognitive wheels: The frame problem of AI. <u>Minds, Machines and Evolution</u>. C. Hookway. Cambridge, Cambridge Univ Pr**:** 129-151.

Dennett, D. C. (1986). The logical geography of computational approaches: A view from the East Pole. <u>The Representation of Knowledge and Belief. Arizona colloquium in cognition</u>. M. Brand and R. M. Harnish. Tucson, AZ, USA, University of Arizona Press**:** 59-79.

Dennett, D. C. (1991). "Real patterns." <u>Journal of Philosophy</u>: 27-51.

Dennett, D. C. (1991). <u>Consciousness Explained</u>, Little, Brown.

Dennett, D. C. (1993). "Learning and labeling." <u>Mind and Language</u> **8**(4): 540-548.

Dretske, F. (1981). <u>Knowledge and the Flow of Information</u>, MIT Press.

Elman, J. L. (1990). "Finding structure in time." <u>Cognitive Science</u> **14**: 179-211.

Elman, J. L. (1991). "Distributed representations, simple recurrent networks, and grammatical structure." <u>Machine Learning</u> **7**: 195-224.

Elman, J. L. (1992). Grammatical structure and distributed representations. <u>Connectionism: Theory and Practice</u>. S. Davis. New York, Oxford

University Press.

Elman, J. L. (1993). "Learning and development in neural networks: The importance of starting small." Cognition **48**(1): 71-99.

Fodor, J. (1981). Representations, MIT Press.

Fodor, J. (1983). The Modularity of Mind : An Essay on Faculty Psychology. Cambridge, Mass, MIT Press.

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

Fodor, J. (1990). Why there still has to be a language of thought. Mind and Cognition: A Reader. W. Lycan, Blackwell Publishers Ltd.

Fodor, J. (1998a). Concepts: Where Cognitive Science Went Wrong. Oxford ; New York, Clarendon Press.

Fodor, J. (1998b). In Critical Condition : Polemical Essays on Cognitive Science and the Philosophy of Mind. Cambridge, Mass., MIT Press.

Fodor, J. and E. Lepore (1992). Holism: A Shoppers Guide. Cambridge, Mass., Blackwell.

Fodor, J. and B. McLaughlin (1990). "Connectionism and the problem of systematicity: Why Smolensky's solution doesn't work." Cognition **35**: 183-204.

Fodor, J. A. (1994). The Elm and the Expert: mentalese and its semantics. Cambridge, Mass, MIT Press.

Fodor, J. A. (2000). The Mind Doesn't Work That Way.

Fodor, J. A. (2001). "Language, thought and compositionality." Mind and Language **16**(1): 1-15.

Fodor, J. A. and E. Lepore (1999). "All at sea in semantic space: Churchland on meaning similarity." Journal of Philosophy **96**(8): 381-403.

Garson, J. (1998). "Chaotic emergence and the language of thought." Philosophical Psychology **11**(3).

Garson, J. (2001). "(Dis)solving the binding problem." Philosophical Psychology **14**(4).

Gayler, R. (2003). Vector Symbolic Architectures answer Jackendoff's challenges for cognitive neuroscience. Sydney, Joint International Conference on Cognitive Science.

Glymour, C. N. (1992). Thinking Things Through : An Introduction to Philosophical Issues and Achievements. Cambridge, Mass, MIT Press.

Gopnic, A. and A. Meltzof (1997). Words, Thoughts and Theories. Cambridge Mass, MIT Press.

Hardcastle, V. (1997). "Distinctions without differences: commentary on Horgan and Tienson's Connectionism and the Philosophy of Psychology."

Haugeland, J. (1991). Representational genera. Philosophy and Connectionist Theory. W. Ramsey, S. Stich and D. Rumelhart, Lawrence Erlbaum.

Horgan, T. and J. Tienson, Eds. (1991). Connectionism and the Philosophy of Mind. Boston, Kulwer Accademic.

Horgan, T. and J. Tienson (1996). Connectionism and the Philosophy of Psychology. Cambridge, MA, MIT Press.

Jackendoff, R. (1992). Languages of the Mind: Essays on Mental Representation. Cambridge, MIT Pr.

Jackson, F. and P. Pettit (1990). "Programme explanations: a general perspective." Analysis **50**: 107-117.

Keijzer, F. (1998). "Doing without representations which specify what to do." Philosophical Psychology **11**(3).

Kirsh, D. (1991). Putting a price on cognition. Connectionism and the Philosophy of Mind. T. Horgan and J. Tienson. Boston, Kulwer Accademic.

Lycan, W. G. (1993). "A deductive argument for the representational theory of thinking." Mind and Language **8**(3): 404-422.

Lyons, W. (1997). Approaches to Intentionality. Oxford, Oxford University Press.

Markman, A. and E. Dietrich (1998). "In defence of representation as mediation." Psycoloquy **9**(48).

Markman, A. and E. Dietrich (2003). "Discrete Thoughts: Why cognition must use discrete representations." Mind and Language **18**(1): 95.

Millikan, R. (1995). Orthography of mentalese. Dennett and his Critics. B. Dalhbom. Oxford, Blackwell.

Newell, A. (1990). Unified Theories of Cognition. Cambridge, Mass, Harvard University Press.

O'Brien, G. (1999). "Connectionism, Analogicity and Mental Content." Acta Analytica: 111-131.

O'Brien, G. and J. Opie (Forthcoming). "weight state semantics."

O'Brien, G. and J. O'pie (Forthcoming). Notes Toward a Structuralist Theory of Mental Representation. Representation in Mind: New Approaches to Mental Representation. H. Clapin, P. Staines and P. Slezac.

Page, M. (2000). "Connectionist modelling in psychology: A localist manifesto." Behavioral and Brain Sciences **23**(4): 443-512.

Perner, J. (1991). Understanding the Representational Mind. Cambridge, Ma, MIT Press.

Pinker, S. (1997). How the Mind Works. New York, Norton.

Pinker, S. (1999). Words and Rules. London, Weidenfeld and Nicolson.

Pollack (1990).

Preston, S. and F. de Waal (2001). "Empathy: Its ultimate and proximate bases." Behavioral and Brain Sciences.

Pylyshyn, Z. (1980). "Cognitive representation and the process-architecture distinction." Behavioral and Brain Sciences **3**(1).

Pylyshyn, Z. (1980). "The `causal power' of machines." Behavioral and Brain Sciences **3**: 442-444.

Pylyshyn, Z. (1984). Computation and Cognition. Cambridge MA, MIT Press.

Pylyshyn, Z. (1991). Rules and representations: Chomsky and representative realism. The Chomskyan Turn. A. Kasher. Cambridge, MA, Blackwell.

Ramsey, W. (1992). Connectionism and the philosophy of mental representation. Connectionism: Theory and Practice. S. Davis. Oxford, Oxford University Press.

Ramsey, W. (1997). "Do connectionist representations earn their explanatory keep?" Mind and Language **12**(1): 34-66.

Rolls, E. (2001). "Representations in the brain." Synthese **129**(2).

Smith, N. and I. Tsimpli (1995). The Mind of a Savant, Blackwell.

Smolensky, P. "Grammar based connectionist approaches to language."

Smolensky, P. (1988). "On the proper treatment of connectionism." Behavioral and Brain Sciences **11**: 1-23.

Speitzer, M. (1999). The Mind Within the Net, MIT Press.

Sterelny, K. (Forthcoming). <u>Thought in a Hostile World</u>.

Stich, S. (1983). <u>From Folk Psychology to Cognitive Science</u>. London, MIT Press.

Sutton, J. (1995). Reduction and levels of explanation in connectionism. <u>Perspectives in Cognitive Science: Theories, experiments and foundations</u>. P. Slezac, T. Caelli and R. Clark.

van Gelder, T. (1995). "What might cognition be, if not computation?" <u>Journal of Philosophy</u> **92**(7): 345-381.

Von Eckardt, B. (1993). <u>What is Cognitive Science?</u> Cambridge, Mass, MIT Press.

Von Eckardt, B. (2003). "The explanatory need for mental representations." <u>Mind and Language</u> **18**(4).

Wheeler, M. (2001). "Two threats to representation." <u>Synthese</u> **129**(2): 211-231.

Wheeler, M. and A. Clark (1999). "Genic Representation: Reconciling content and causal complexity." <u>Brit. J. Phil. Sci.</u> **50**: 103-135.

Wheeler, M. and A. Clark (1999). "Genic representation: Reconciling content and causal complexity." <u>British Journal for the Philosophy of Science</u> **50**: 103—135.

Wittgenstein, L. (2001). <u>Tractatus Logico-Philosophicus</u>, Routledge.