Connectionism, Cognitive Maps & the Development of Objectivity

Ronald L. Chrisley
School of Cognitive & Computing Sciences
University of Sussex, United Kingdom
Email: ronc@cogs.susx.ac.uk

This paper originally appeared in *Artificial Intelligence Review* **7** (1993), pp 329-354; and in Niklasson, L. and Bodén, M., eds (1994) *Connectionism in a Broad Perspective*, London: Ellis Horwood, pp 25-42.

Abstract

It is claimed that there are pre-objective phenomena, which cognitive science should explain by employing the notion of non-conceptual representational content. It is argued that a match between parallel distributed processing (PDP) and non-conceptual content (NCC) not only provides a means of refuting recent criticisms of PDP as a cognitive architecture; it also provides a vehicle for NCC that is required by naturalism. A connectionist cognitive mapping algorithm is used as a case study to examine the affinities between PDP and NCC.

Keywords: cognitive architecture, cognitive map, concept, concept learning, connectionism, content, context-sensitivity, development, generality, intentionality, representation, non-conceptual content, parallel distributed processing, sub-symbolic computation, systematicity.

1 Introduction

Cognitive scientists are quick to point out that there are phenomena which resist analysis and explanation if one is restricted to the mechanistic, physicalist approaches of the traditional natural sciences. Such phenomena are "intentional", exhibit "aboutness", are "other-directed", etc., and as such must be explained in normative terms. Traditional cognitive science offers a naturalistic means of explaining such intentional phenomena, by appealing to representations and their attendant duality of vehicle (the physical manifestation of the representation, or,

bluntly, syntax) and content (the significance of the representation, the way it presents the world as being; bluntly, semantics). This appeal is made in order to explicate both how it is that physical systems can be intentional, and how intentionality can be a part of the physical order.

However, there are several kinds of phenomena that appear, prima facie, to be intentional, but that also resist analysis and explanation by the traditional methods and tools of cognitive science. First, there are the intentional phenomena that seem to be *pre-objective* in that they are prior (in some vague notion of "prior") to "objective" adult human cognition: the mental life and behaviour of infants and animals, for example (the notion of pre-objectivity is clarified in section 2). Next, there are the phenomena surrounding real-time perception and action integration. And there is the very notion of change within intentional systems, be it evolution, morphogenesis, development, learning, or a change of conceptualization. [1] Call these aspects of intentional systems the *recalcitrant phenomena*.

If one further assumes that the traditional techniques of cognitive science are the only ones available for (scientifically) explaining intentional phenomena, then a dilemma is forced: either these recalcitrant phenomena are, despite appearances, non-intentional; or they cannot be (scientifically) explained at all. Some might (as I do) find neither of these options acceptable; what is one to do?

I wish to reject this dilemma by rejecting one of the two premises that forced it. The premises are:

1. that traditional cognitive science cannot explain these recalcitrant phenomena; and

2. that traditional cognitive science is the only means of scientifically explaining intentional phenomena.

This rejection will be effected by sketching out an alternative intentional account that can accommodate the recalcitrant phenomena. Which premise it is that one will see me as rejecting depends on how strong a reading one gives "traditional". On the strong reading of "traditional", the approach I am taking is sufficiently radical to be considered a rejection of traditional cognitive science's monopoly on scientific intentional explanation (2). On a more catholic understanding of what traditional cognitive science comprises, the approach I advocate here is an extension of the basic notions of content and representation; it was the appearance that cognitive science could not account for pre-objectivity, development, etc., (1) that was mistaken.

But we need more than just *any* account of the recalcitrant phenomena; we need an account that will explicate how these phenomena are related to the (more or less) conceptual, objective kinds of intentionality with which cognitive science has traditionally been concerned. I think the best account of the relationship between these two types of intentionality is itself given in intentional terms. Thus,

what becomes central in cognition is an intentional understanding not of objectivity and pre-objectivity independently, but of the *development* of objectivity: intentional processes that, in general, increase the objectivity of the contents in a system. [2]

2 Content, systematicity, and objectivity

The intentional explanatory strategy that dominates cognitive science typically understands psychological states in terms of attitudes (belief, desire, knowledge, intention, etc.) toward contents (that there is a door ahead, that $2 + 2 = 4$, etc.); such attitude/content pairs are appealed to in intentional psychological explanation. For example, one might explain why an agent opened a door (i.e., show that the agent's opening the door wasn't just an accident; if circumstances had changed slightly – if the door were one foot over to the right, say – the agent would still have opened the door) by claiming that the agent *intended to open the door*; one could explain the possession of this intention as being the result of the agent's *desire to be in the next room* and its *belief that opening the door will help one get into the next room.* I will call explanations that appeal to such attitude/content pairs *content-based explanations*, or *CBEs.*

Although its extreme simplicity might make the above example of content-based explanation more caricature than characterization, I hope that it at least suggests how attitudes and contents can figure in intentional explanation. Unlike much AI-inspired cognitive science research, where the emphasis has been on the attitudes in general and the attitude of knowledge in particular, the focus in this research is on content. The proposal is that a solution to the problems of explaining the recalcitrant intentional phenomena lies in a proper understanding of the notion of content, and how it relates to the explanation and design of intelligent systems.

2.1 Non-conceptual content is pre-objective, unsystematic content

I join others [Crane, 1992, Cussins, 1990, Davies, 1990, Evans, 1982, Haugeland, 1991, Peacocke, 1992] in arguing that a distinction should be made between conceptual and non-conceptual contents. [3] For several reasons, much of the work in AI has concentrated by default on the case of conceptual content, but there is reason to believe that understanding non-conceptual content is essential to understanding (and therefore to designing) intentional systems in general; see [Cussins, 1990] .

Some ways of representing the world (contents) are objective or near-objective, some are not. A way of representing some aspect of the world is objective if , e.g., it presents that aspect of the world as something that could exist while unperceived. Strawson [Strawson, 1959] maintained, as did Evans [Evans, 1982]

after him, that at least in the case of thinking about spatio-temporal particulars, truly objective thought is manifested in the possession and maintenance of a unified conceptual framework within which the subject can locate, and thus relate to any other arbitrary object of thought, the bit of the world being thought about.

If this is a correct understanding of objective thought, then it has important implications for the understanding of pre- (or non-, or sub-) objective representation. Pre-objective representation involves contents that present the world, but not *as* the world, not as something that is or can be independent of the subject. The Piagetian case of an infant's perceptual/motor interactions with its environment before attaining the stage of object permanence is a plausible example of a subject that is employing a pre-objective mode of thought. The infant tracks an object (thus suggesting that there is some intentional relation between the infant and the object) when it is perceptually occurrent, yet when the object is occluded from view, the infant loses interest in the object (perhaps dropping it), and is in fact startled if the obstruction is removed to reveal the object. Since the very notion of an *object* essentially involves the notion of somehting that can exist even though occluded, the infant is not thinking of the object *as* an object. The contents of the infant's thoughts concerning the object do not present the object as something objective, as something that could exist while unperceived. According to the Strawsonian/Evansian line I am taking, then, the infant's lack of objectivity must be manifested in the lack a unified framework of thought: the infant is unable, in general, to locate objects in such a framework. I will call such pre-objective contents *non-conceptual contents* (*NCC's*); *conceptual contents*, on the other hand, are objective.

It is a consequence of this way of understanding the conceptual/non-conceptual distinction that conceptual contents will necessarily be systematic, will meet Evans' Generality Constraint [Evans, 1982, page 104]. That is, for any conceptual contents (ways of thinking of properties) F and G, and any conceptual contents (ways of thinking of objects) a and b, if a subject knows what it would be for a to be F and for b to be G, then it must know what it would be for a to be G and for b to F. This is a direct consequence of the necessity, for there to be objective, conceptual thought, of a unified framework within which to locate all properties and particulars. Non-conceptual contents, on the other hand, are not systematic. In the case of NCC, the mode of thought is pre-objective; a unifying framework is not present, and there are, therefore, properties and particulars which cannot be related in the proper way. The idea of non-conceptual content, then, implies that one can represent the world with proto-concepts that do not universally recombine with all other possessed proto-concepts.

For example, consider an infant which cannot, as before, think of a particular object (a glass, say) as existing unseen, but *can* represent its mother as being behind, out of view (on the basis of hearing her voice or feeling her arm, say). The contents of such an infant will violate the Generality Constraint, since the infant may be able to think (something like) *glass in front of me* and *mother*

behind me but not *glass behind me*. The infant's contents are not fully objective, and are therefore non-conceptual. To ascribe conceptual content to the infant in this case would mis-characterize its cognitive life, and would not allow prediction or explanation of the infant's behaviour.

2.2 The graded nature of content: degrees of objectivity

This example brings up another, important aspect of content: it can be *more or less* conceptual, *more or less* objective; meeting the Generality Constraint can be a matter of *degree*. Thus, there is not just the simple distinction between one level of content which is non-conceptual, and one level which is conceptual. Rather, there are numerous degrees of conceptuality/non-conceptuality. This notion of graded abilities of infants is at least as old as Piaget; but what is being discovered now is that the development of objectivity has a richer texture, worthy of its own intentional account, than even Piaget imagined. For example, Hood and Willatts [Hood and Willatts, 1986] have observed that some (i.e., five-month-old) infants that lack our full notion of object-permanence *can* represent objects that are unperceived, but only under certain conditions (i.e., in the dark). This is just one instance of an intermediate stage between no notion of object permanence at all, and the full, objective notion which we take ourselves to possess. The development of objectivity isn't just one single, boot-strapping, mechanistic jump from a non-intentional system to a fully conceptual one; rather, it is a complex intentional phenomenon which needs explanation, and which constrains the possible explanations of the relatively objective form of cognition that it yields.

Much of this discussion concurs, at least in part, with a recent survey of object-permanence in infants, by Harris [Harris, 1989]. At one point, he considers the well-known "A not B" error in infants: an infant that has previously found hidden objects at a place A will search at A, even when it *witnesses* the act of hiding the object at B. Harris examines the failure of the various attempts that have been made to capture the content of the mental states of such infants. He finally suggests that these contents cannot be given a conceptual explanation. Why not? Because they fail to meet the generality constraint, *and do so to varying degrees*:

> These very orderly data create a problem for any purely conceptual or cognitive interpretation. Consider, for example, the idea... that the infant does not appreciate that an object can be in only one place at a time, and fails to rule out A as a possible hiding place. Are we to say that the 8-month-old baby can rule out A for 3 seconds but no longer, the 9-month-old baby for 5 seconds but no longer, and so forth? If the baby has come to understand that an object can only be in one place, *why does it not apply this knowledge to delays of any length?* [Harris, 1989, page 115, emphasis mine]

After suggesting a particular account of the data which does take non-systematicity seriously, he remarks:

> As soon as we talk in these terms, we are beginning to talk about the efficiency with which the baby processes the information, rather than simply proposing conceptual rules that the baby either understands or does not understand. We are admitting, at the very least, that a conceptual interpretation will not work.

Thus, Harris recognizes that conceptual accounts of the infant's mind won't do, since an account must take seriously this failure to meet the Generality Constraint. But then Harris (understandably) concludes that this means that intentional accounts in general won't do; a non-intentional, mechanistic account (he specifically considers the proposals made in [Diamond, 1988]) is the best we can do. Although incorrect, his conclusion is understandable, since the option of *non-conceptual* intentional analysis is not in common cognizance. One of the purposes of this paper is an attempt to redress this lack of awareness. I reiterate: one need not give up intentional explanations of the recalcitrant phenomena, providing one analyzes them in terms of non-conceptual content.

3 Content & vehicle: The NCC/PDP connection

In the cognitive scientific explanatory scheme, contents need vehicles. Naturalism requires that we show how our intentional and non-intentional characterizations of an organism cohere; the hypothesis of cognitive science is that we should do this by finding a computational characterization (vehicle) which marches in step (see section 3) with our non-conceptual intentional account (content) [Cussins, 1987, Fodor, 1985]. There is reason to believe that such a computational architecture will differ substantially from classical, purely symbolic notions of computation, which are typically meant to correspond to objective, conceptual contents.

There are those few who not only maintain that there is a significant difference between symbolic representation and the kind of representation one finds in parallel distributed processing (PDP) [4] models of cognition, but who also connect this difference with a difference in the kind of content carried by each of these forms of representation (e.g.,[Haugeland, 1991, Cussins, 1990]); I count myself as among this number. The general idea is that whereas symbolic representations carry conceptual, systematic content, composed of elements corresponding to objects and properties, PDP representations carry non-conceptual, non-systematic content that is composed of elements that do not correspond to orthodox objects and properties. The previous section tried to give you some idea of what non-conceptual content is, and why cognitive science should bother with the idea.

This section discusses why one might think parallel distributed processing and non-conceptual content are a natural match. The next section tries to make precise some of these affinities by citing examples involving a particular connectionist cognitive mapping network: the Connectionist Navigational Map.

In a way, some authors (e.g., [Fodor and Pylyshyn, 1988, Davies, 1990, Ramsey et al., 1991]), although they draw some conclusions that are radically opposed to the ones proposed here, have done a lot of this section's work for me. Their arguments against the compatibility of PDP and the propositional attitudes implicitly assume that all content is conceptual content. "[Conceptual] content", the arguments go, "can only be carried by representational vehicles that do not lack property X[5]. But PDP vehicles lack property X. Therefore, PDP representational vehicles cannot carry [conceptual] content." Rather than concluding from these arguments that PDP is uninteresting and irrelevant to cognitive science, one can instead conclude that these arguments establish PDP (and NCC) as a radical alternative to the orthodox: if PDP representations do not carry systematic, conceptual content, then (since they must carry *some* kind of content) they must carry some other, non-systematic kind of content. *Non*-conceptual content.

From the point of view being developed here, then, PDP and NCC need each other:

- PDP needs NCC. If there is no NCC, and if the arguments that attempt to show the incompatibility of PDP and conceptual content are right, then, despite appearances to the contrary, PDP representations carry no (explicable) content at all. Thus, PDP cannot be a cognitive architecture: it cannot naturalize intentional characterizations.

- NCC needs PDP. As stated before, naturalism demands that a characterization of an organism at the level of content, which we cannot directly understand to be manifested by a non-intentional characterization of that organism, be shown to "march in step" with a characterization (in cognitive science, this characterization is computational) of that organism that we *can* understand to be so manifested. One of the principal attractions of Fodor's Language of Thought approach to cognition is the Representational Theory of Mind (RTM) which underlies it [Fodor, 1987]. RTM attempts to explain how intentional phenomena could be realized in the physical world by showing how the physical (characterized in terms of classical computational states) and the intentional (characterized in terms of conceptual semantics) march in step. To be naturalized, then, NCC will require a computational architecture with which it can march in step. Perhaps PDP is such an architecture; there might be others (in which case, Cussins [Cussins, 1990, page 431] is right: PDP needs NCC more than the converse).

However, it is not wise to rely too heavily on these arguments that others

have provided, as they establish too much: that PDP could *never* carry conceptual content. Since I want to explore the possibility that PDP could provide the architecture for an explanation of the development from NCC to objective, conceptual representation, the full conclusions of these arguments should be avoided if possible. My position is that, unlike classical cognitive architectures, PDP architectures can be appropriate for both non-conceptual *and* conceptual contents – depending on the stage of development of the architecture. I thus make a distinction between architectures that are *necessarily* systematic, and ones for which particular configurations result in *contingent* systematicity. The basic insight is that there are non-systematic intentional phenomena, so cognitive science should favour the latter type of architecture, which can accommodate both types of intentionality.

Even if this position is taken, there remains an alternative, related argument against PDP as a cognitive architecture to be found in Fodor and Pylyshyn's paper. This argument does *not* assume that systematic, conceptual contents can be carried only by representational vehicles that are *necessarily* systematic, but allows that *contingently* systematic vehicles can do the trick. The argument takes the form of a dilemma: either PDP can attain the (contingent) systematicity required for conceptual content, or it cannot:

- Horn 1: if PDP *can* achieve systematicity, then it is uninteresting to cognitive science; it is a mere implementation of the classical architecture (structured representations and operations sensitive to that structure);
- Horn 2: if PDP *cannot* achieve systematicity, then it cannot model human competence, such as the productivity of language (e.g., the ability to understand arbitrary novel sentences).

Some might wish to reject Horn 2, and deny that human behaviour is ever such that it must be understood to be the product of systematic mental processes; I need not make such a bold claim here. Rather, I concede that PDP's success as an architecture for *all* of human cognition requires that it be able to construct systematic representations: PDP must be able to *attain* (contingent) systematicity. [6] Otherwise, we would have an unattractive, inexplicable schism in our understanding of cognition: the (necessarily) non-systematic, PDP-modeled phenomena, incomensurable with the (necessarily) systematic, classically-modeled phenomena. But the best theory will be the one that does *not* leave the transition from non-systematic subjectivity to systematic objectivity a mystery, but explains it. The existence of developmental data provides an extra, crucial constraint on our cognitive architectures. The point being made is not just that one should prefer architectures whose existence can be explained to those which are miraculous; one should also prefer, ceteris paribus, architectures for which there is a *better* explanation of its development to those with a *worse* explanation.

The advantage of a non-conceptual approach to cognitive architecture is that the explanation of systematicity *is itself intentional*; in a classical system, such an explanation, if there is one, must remain non-intentional and mechanistic, with all the attendant disadvantages in terms of explanatory power, generalizability, etc.

Thus (to get back to the dilemma), Horn 2 cannot be denied; it is Horn 1 which should be resisted. The supporter of PDP in general could do this in a number of ways (although only the first response is appropriate here):

1. First and foremost: agreement in a special or limiting case is not the same as implementation; a PDP architecture that meets the classical constraints in the special case of systematic adult human behaviour is no more irrelevant to cognitive science than a quantum theory that meets Newtonian constraints in the special case of macroscopic conditions is irrelevant to physics (after Smolensky [Smolensky, 1988]);

2. Even if PDP's contingent systematicity *did* render it a strict implementation of classical architectures, that would still allow it to facilitate the construction of theories with explanations and predictions considerably different from theories built on other classical architectures. To be a member of the class of classical architectures, one need only meet two very general constraints: possess structured representations and structure-sensitive operations. Thus, theories based on PDP architectures would be just as relevant to cognitive science as those based on architectures like SOAR [Rosenbloom et al., 1992] or ACT* [Anderson, 1983], which also respect the classical constraints. There is no reason to believe that just because two architectures meet the classical constraints that all the theories specifiable in terms of one architecture will also be specifiable in terms of the other. If PDP is mere implementation theory because it meets the classical constraints, so are SOAR and ACT*. But surely Fodor and Pylyshyn don't want to claim that SOAR and ACT* are therefore irrelevant to cognitive science;

3. Also, "mere" implementation theory isn't that irrelevant to cognitive science. As said before, naturalism demands that the content-involving intentional level be shown to "march in step" with the computational architecture. A computational architecture which met this demand would be of reduced value if it could not be shown to cohere with a non-computational, physiological characterization of the organism whose psychology we were striving to understand. In this respect, not all implementations are created equal: we should prefer the one that can be understood to be instantiated by the organism;

4. Last (though least likely), PDP *might* be able to achieve systematicity

(model our generative, productive, etc., behaviour) other than by meeting the classical constraints, *pace* Fodor and Pylyshyn's claim that theirs is the only game in town.

In summary, Fodor and Pylyshyn were right to admit that PDP might be able to achieve contingent systematicity, but they were mistaken in failing to realize that such a characteristic is not a detraction, but a desideratum; "It's not a bug, it's a feature!"

4 A case study of PDP/NCC affinities

With the forgoing by way of introducing the notion of non-conceptual content and the theoretical issues concerning its connection with parallel distributed processing, the remainder of this paper employs the development of a particular PDP network in order to illustrate the fact that the contents of at least some PDP representations are best viewed as non-conceptual. This is done by giving five examples of how such contents cannot be understood to be conceptual; a sixth example demonstrates (in a modest way) the transition from representations with non-conceptual contents to those with conceptual contents (the kind of transition required for the "development of objectivity" mentioned in the title). But first, the network architecture itself is described.

4.1 The Connectionist Navigational Map

The *Connectionist Navigational Map* (*CNM*; [Chrisley, 1990]) is a computational architecture being developed with the aim of providing an autonomous robot with the ability to learn and use spatial maps for navigation. One component of this architecture, the *predictive map*, allows the robot to predict what sensations it would have if it were to move in a particular ego-centrically specified manner (e.g "rotate $\pi/4$ radians to the right", "move forward 10 feet"). Of course, this requires the robot to have some kind of representation of its current location, since, in general, the mapping from actions to sensations is dependent upon where one is in the world. That is, the mapping from sensations and actions to sensations is one-to-many, since more than one place can have any given sensory signature. Thus, the spatial environment, and therefore a model of it, can be seen instead as a function from current location and current action to predicted sensations. The input consists of a state representation, or *location code*, corresponding to the current location a of the robot, and an action representation representing the move m being made. The output of the network is a vector that is supposed to be equal to the sensation vector the robot would receive from its senses if it were actually at the place that is reached by making the move m at location a.

Of course, there is more structure to space than a simple, direct mapping from locations and actions to sensations indicates (see figure 1). Specifically,

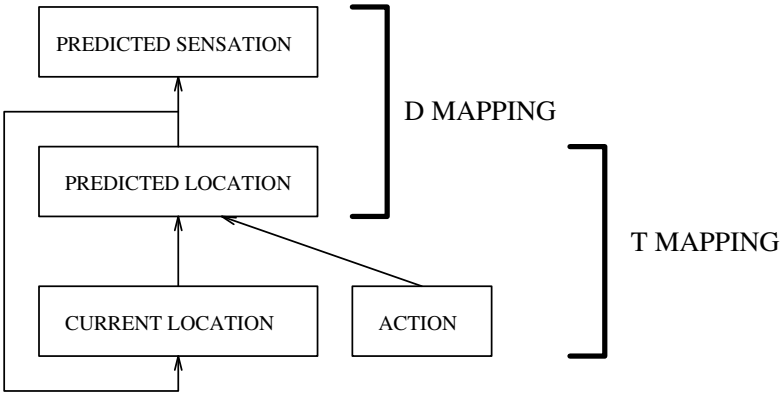

Figure 1: The PDP architecture of the predictive map (*locations* $\times$ *actions* $\mapsto$ *sensations*) formed by composing a topological mapping T (*locations* $\times$ *actions* $\mapsto$ *locations*) with a descriptive mapping D (*locations* $\mapsto$ *sensations*). Arrows indicate directed, full inter-connection between layers of units.

location and action determine a new location, which itself determines the sensations of the robot. Thus, it might be easier for a robot to learn (or a theorist to analyse) a predictive map if its structure reflects this regularity of the spatial environment. The predictive map of the CNM is a composition of two mappings: a topological mapping T (from locations and actions to states) and a descriptive mapping D (from locations to sensations). In actual use, the location output of the T mapping, after a given action, is used as the location input to the T mapping for the next action. [7] Thus, if a constantly north-facing robot considers moving forward and then moving right, it can use the map to predict what sensations it would have after those moves by calculating $D(T(T(a,$**move-north**$),$ **move-east**$))$, where a is a location representation corresponding to the robot's initial location before the actions, and **move-north** and **move-east** are action representations with the intuitive interpretation.

The predictive map can be realized in a hybrid system, with the topological mapping realized symbolically, and the descriptive mapping realized by a PDP network. It is argued in [Chrisley, 1991] that this kind of hybrid structure is attractive for the purposes of engineering a working system; *however*, if the primary motivation for constructing the system is to understand pre-objective representations, and how a robot can make the the transition from pre-objective to objective representations of space (as it is in this paper), then a uniform, sub-symbolic architecture for both mappings should be employed.

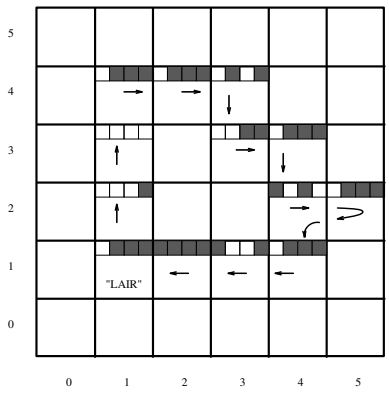

Figure 2: The region of the "grid world" used in the simulations. Only the details for the locations on the used route are shown. Arrows indicate the direction of travel along the route. The four-bit binary vector at each location indicates the description or sensation vector associated with that location.

4.2 The experimental setup

Although the CNM is intended for actual robots moving in real space, it can be of use here in making concrete some of the connections between NCC and its computational vehicles. The experimental situation used here is a deliberately impoverished one: a developing (learning) agent moving through a simulated "grid world"; the part of the world simulated has only 36 cells or locations (6 by 6). Each location has a 4-bit vector associated with it, which can be understood to be the sensations the agent has when at that location (see figure 2). As in its normal use, the CNM is to provide a means for this agent to improve its navigation of its space (and thus increase the objectivity of its contents concerning that space) through sensory prediction. This situation is grossly impoverished with respect to real-world cognitive map building, but the simplifications are not only acceptable, but necessary, given the nature of this discussion: *explication* of the affinities between NCC and PDP.

The agent has only four actions available at any location, those of moving into each of the adjacent locations (orientation is not modeled: the agent can be thought of as always facing north). It is assumed that the developing agent has some adult or teacher, not providing the agent with input/output patterns to be learned, but actually guiding the agent through the territory along a route that starts and ends at some privileged location, called the *lair*. [8]

The adult traverses the route, taking the developing agent along. The developing agent stores the sequence of actions taken and the sensations that result. Note that sometimes the same action yields different sensations, and that different actions sometimes result in the same sensations. When the agent returns to its lair, it iteratively learns the route, not by actually moving, but by reviewing [9] the day's events in the following manner:

- First, generate a training set:

    1. Assume some arbitrary representation or code for the initial location (the "lair"). Store this code, and the code for the first action taken, as an input pattern; store the sensations that were observed after taking that action as the target output for that pattern.
    2. Propagate the current input pattern into the T mapping.
    3. Use the output of the T mapping, along with the next action on the remembered route, as the next current input pattern. Store this pattern into the training list as an input pattern, as well as storing the next remembered sensation as the target output pattern for that input.
    4. Go to 2 until finished with the remembered route.

- Then, learn with the current training set: adjust the weights of the T and D mappings according to the gradient of the error (difference between target and actual outputs).

- After some period of learning with one training set (in the simulations described, the time was 10 epochs) , a new training set is created, in the same manner as before (steps 1-4), and the agent trains with the new set (for another 10 epochs).

In the simulation used for the examples, this was repeated 18 times, at which point the network had learned the route. That is, starting with the code for the lair and the initial action taken, the T mapping would produce a new code that not only yielded the correct predicted sensations via the D mapping, but which also, in conjunction with the representation for the next action taken, produced a code via the T mapping which could itself yield both the right sensation vector and location code, and so on, iteratively.

The T mapping was realized by a network which comprised 8 inputs (4 for the location code, and 4 for the action code; action codes were the 4 unit vectors, one for each of the 4 directions the agent could move), 4 hidden units, and 4 outputs (a location code). The D mapping was implemented by a network with 4 inputs for a location code, and 4 outputs for a sensation vector at that location. A modified version of Fahlman's "Quickprop" algorithm [Fahlman, 1988] was used for learning (although the network was simply recurrent, back-propagation through time was not necessary).

In analysing the representational contents of this network, the activity patterns of the location, action, and sensation vectors were the vehicles chosen for consideration. This bypasses several very important issues (Don't weights serve as representational vehicles in the CNM? Shouldn't the representational vehicles include aspects of the network's environment?, et al.) which cannot be considered here.

4.3 CNM representations cannot be analysed conceptually (examples 1-5):

Of course, it is relatively easy to find something that cannot be analysed as carrying conceptual content; one only has to pick up a stone. But that's because a stone can't be analysed as carrying *any* kind of content. The examples that follow are meant to show why a particular PDP network that is, unlike a stone, intuitively a representational, content-involving system, cannot be understood conceptually, but is best analysed as containing representations with non-conceptual content.

4.3.1 Example 1: Perspective-dependent representations

A good way to see the connection between one kind of systematicity and objectivity is to examine the limitations of the CNM T mapping. It is a consequence of the non-systematicity of T mappings that are learned just for the sake of navigating a few routes that their place representations will not be objective.

Let's forget, for the moment, about the particular route mentioned in 4.1, and consider the nature of T mappings in general. For example, suppose that the T mapping is non-systematic in that not all representations are mapped to four new, distinct representations. Suppose that only a few (enough to provide competence on the particular routes travelled) locations around the lair are systematically represented in the T mapping. This situation is illustrated in figure 3. Location codes are represented by circles: e.g. the circle marked "L" denotes the location code L which the agent is using to represent the lair; the code $T(L,$**move-east**$)$ is denoted by a circle that is pointed to by an arrow from the east (that is, right) side of the circle marked "L", etc.

It might be that this non-systematic topology is a *direct product* (albeit sub-optimal one) of the CNM's attempt to account for the data it has already encountered; perhaps there is something about what is in the space, like a pit at the place ten units east of the lair, which makes this sub-optimal, non-systematic topology quite practical in that context. Or it could be that the non-systematicity in regions not close to the lair is a mere *by-product* of developing a systematic T mapping for the area near the lair (perhaps like flattening a lump in a carpet merely moves the lump to some other part of the carpet). The reasons for the lack of systematicity are irrelevant for the purposes of this paper. What is important is that we have a case where there is enough systematicity to get a notion of representational content going, but not enough to be able to speak of full Generality and objectivity.

In order to see why this non-systematic representation of space is non-objective, we have to notice a few constraints on objective space in general:

1. If one moves (in a very abstract sense of "moves") in a straight line from a location a, one will be located at a location $b \neq a$;

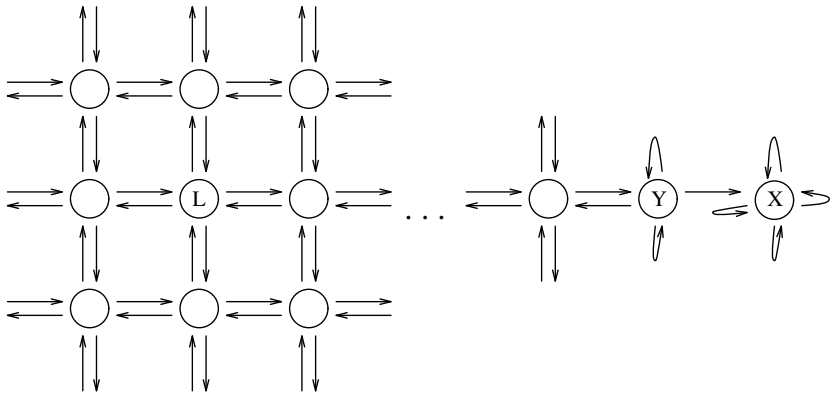

Figure 3: An example of local systematicity but global non-systematicity in the T mapping. Each circle represents a location code; "L" indicates the code that is being used to represent the lair; arrows indicate the T relation between codes (see text). In such a case, some, if not all, location codes (such as the one marked "X") will not represent any objective place.

2. Moving in distinct directions from the same place will take one to distinct locations

3. For every simple movement m there is an inverse m^{-1} such that $T(T(a, m), m^{-1}) = a$;

The T mapping illustrated in figure 3 is incapable of representing a space that meets these constraints. For example, once one iterates enough times through the T mapping, to calculate the representation for the location one gets to by, say, starting at the lair and executing **move-east** ten times, the resulting location code, X (represented by the circle marked "X" in the diagram), is such that:

1. $T(X,$**move-east**$) = X$, which violates constraint 1;

2. $T(X,$**move-east**$) = Y = T(X,$**move-south**$)$, which violates constraint 2;

3. $T(Y,$**move-east**$) = X$, and **move-west** = **move-east**$^{-1}$, yet $T(X,$**move-west**$) \neq Y$, which violates the third constraint.

In such a case, the T mapping breaks down to such an extent that one cannot interpret the resulting location code, X, as a representation of the place ten units east from the lair. [10] Thus, the network cannot, as things stand, represent the location in question *at all*. This alone is sufficient to establish that, in such a case, even the location codes that *can* be understood to represent locations (i.e., those codes near the lair code, L) do not represent objective, conceptual locations, since truly objective locations are part of an arbitrarily extendible space, while these locations are not. These codes, unlike X, *can* support a sizeable, though limited,

region of intentional spatial activity; for these codes, unlike for X, there is enough systematicity to determine a referent. But there isn't enough systematicity for that referent to be represented *as* an objective, perspective-independent place, a part of a purely objective space. Thus, the location codes, even those near (and including) L, carry non-conceptual content. Similarly for the action vectors: truly conceptual, objective movement should be arbitrarily iterative.

The fact that the violation of these constraints implies a non-objective representation of space can be shown in the following. The only way an agent with such a CNM configuration *could* represent (at this stage of development of its T mapping) the place ten units east would be by actually *moving* to that location. Upon moving east ten times, its location code for its current location would, by hypothesis, be of no use (e.g., it would not allow any successful predictions of the result of any action the agent might take). Keeping the T mapping fixed, the agent would have to co-opt the T mapping for its new locale, by using L to represent not the lair, but the current location, and altering the D mapping accordingly. The location codes would have to undergo a change of referential significance (of a kind that would be impossible within the context-invariant restrictions of a classical architecture; see 4.3.3, below). But this means (assuming that the non-systematicity of the T mapping crops up when iterating **move-west** as well) that the agent will no longer be able to represent the space around the old lair; *after* the change, the lair will be just as unrepresentable, mutatis mutandis, as the current location was *before* moving.

Thus, the agent's ability to represent a place is dependent on the agent's *perspective*: in this case, its actual proximity to that place. Like the infant without the concept of object permanence, the agent would lack the notion of *location* permanence, to some *variable* extent. The CNM can represent places (i.e., it doesn't treat all qualitatively identical places as the same place, but individuates locations using some notion of non-locally-observable, relational properties). But since *objective* places do not disappear when one moves away from them, the network cannot be representing places *as* objective, perspective-independent places. Thus, the contents of the location codes of the network at such a stage in its development must be pre-objective, non-conceptual.

Another, brief way of putting the point: the relations between objective places are independent of both what is located at those places, and of our abilities to move between those places, whereas a non-systematic T mapping typically is not independent of these. [11]

4.3.2 Example 2: Non-systematicity

In some absolute sense, the above example of perspective-dependent representations already demonstrated the non-systematicity of such representations. But it demonstrates an asymmetry between the places that it can represent and those it cannot, whereas the Generality Constraint (cf. section 2) was meant only to

impose systematicity among contents that the agent can actually entertain. Nevertheless, the CNM also exhibits this kind of non-systematicity among the places it *can* represent.

Perhaps the easiest way of seeing this is to consider the fact that the network's choice of codes to use in the T mapping will constrain the possibilities for the D mapping. For example, suppose that the T mapping has very similar codes for two places. Now one of the main features of PDP representation is that all of the representation vectors exist within a common relational space. The ability for PDP networks to generalize in an interpolating (and extrapolating) fashion stems from the fact that the requirement of producing a particular output vector for a given input vector will severely constrain the possible output vectors for any nearby input vector. Thus, networks cannot, in general, assign arbitrarily dissimilar outputs to arbitrarily similar inputs. This means it will be difficult for the D mapping to assign different sensory properties to similar location codes. There will be some predicational combinations (of sensory properties to locations) which will be difficult or impossible, violating the Generality Constraint. Thus, PDP's generalization ability is, in this sense, inherently non-systematic, as it violates the independence of subject and predicate.

4.3.3 Example 3: Context-sensitivity I

A third way of motivating the link between PDP and NCC focuses on the methodology of content ascription in both symbolic and PDP architectures. In symbolic architectures, the compositional nature of the semantics encourages theorists to make their primary semantic ascriptions on the atomic level. They assign (conceptual) contents to the atomic symbols; familiar recursive rules then determine the content of any molecular representation from the contents of its atomic constituents, together with their mode of combination. So **block-57** and **block-23** are assigned particular blocks as referents, **is-above** is assigned the "above" relation, and **is-above(block-57, block–23)** is assigned the value *true* if the first block is above the second block, and $false$ otherwise. The procedure is to assign a semantic value to the atomic symbols[12], and then, given the compositional rules which determine the semantics of complex expressions, try to build a system that uses these symbols so that the conceptual ascriptions, to both the atomic and molecular symbols, are warranted. The compositional relations are built-in and guaranteed; what has to be achieved is a complex function of operation, both internal and interactive with the environment, that bestows on the atomic symbols the content optimistically assigned to them.

In contrast, the primary locus of PDP content ascriptions is the propositional level. For example, consider the following system. An autonomous robot receives sonar input from the 360° ring in its two-dimensional plane of action. Its perception and action are coordinated by a PDP network which has learned, not only to associate different sonar signatures with obstacles, but also to associate with

the input an appropriate avoidance behaviour depending on both the type of signature (person vs. chair) and the position of the obstacle signature in the input ring (immediately to the left, straight ahead ten feet or so, etc.). In giving an intentional analysis of this network, a theorist might note that a certain pattern in the hidden units is always present when the network classifies the input as *chair that is in front and within three feet*. The hidden unit pattern in question is assigned the content *there's a chair less than three feet in front* or something similar. But there is no guarantee that there is any isolable part of this hidden unit representation that corresponds, in all situations, to the concept *chair* or the relation *being less than three feet in front*. For example, a hidden unit vector with the content *there's a chair over ten feet behind* might be completely *orthogonal* to the one about the chair being close and in front. The situation is the converse of the one for typical means of content ascription to symbolic systems: the content ascription is warranted, but the structure of the representation is in question. Thus, even if a hidden unit pattern in a more advanced PDP system were ascribed a content that involved a *particular* object, such as *the chair bumped into two minutes ago is now behind*, there would be few, if any, constraints on what hidden unit pattern is carrying the content concerning the chair that the robot bumped into two minutes ago. In fact, it might be that there is *no* context-invariant part of the hidden unit pattern that corresponds to the chair at all. The object/property decomposition is not built-in.

It is this difference in content ascription, I believe, that is responsible for two different kinds of context-sensitivity in PDP representations that makes PDP especially suited for NCC. I visualize these two kinds of context-sensitivity as functions from (sub-total [13]) representations to (sub-propositional) contents, and call them *one-to-many* and *many-to-one* context-sensitivity, respectively.

The first kind of context-sensitivity is manifested in the fact that the same (sub-total) representation will carry many different contents, depending on the context. As said before, the fact that the locus of the primary content ascriptions for PDP representations is at the propositional level means that there is no built-in semantic relationship between propositional and sub-propositional representations, and therefore there is nothing to prevent the same sub-total representation from being a constituent in two quite different representations, even two that have no overlapping content: *the chair is ahead* and *the person is behind*, say. In such a case, the content of the sub-total representation must be different in the two contexts, since there is no sub-propositional content common to both contexts that could be assigned to the representation. Note that this context-sensitivity is ruled out for symbolic representations, since the atoms are typically assigned context-invariant contents, and the contents of molecular symbols are determined from these atoms in a context-independent manner.

In our case of an agent learning with the CNM, this kind of context-sensitivity could arise if the same code were used for two different places simultaneously. That is, a case where, say, there are two regions of the the space that are suffi-

ciently similar that the CNM can use the same representation for two places, one in each of the similar regions. This will result in a lack of systematicity. If two place representations are not co-referential, then they must have different content (since content determines reference). If these contents were conceptual, and met the Generality Constraint, then it should be possible to arbitrarily combine these contents with predicational contents; it should be possible for the network to have a D mapping for one of the representations that is different from the D mapping for the other. But since the representations are identical in this case, there is no such possibility. The contents of the representations are not independent of each other, so they are not representing the places as conceptual, objective places; they must be representing them non-conceptually.

In the context of other networks, this one-to-many type of context sensitivity can imply non-systematicity in a different way. Since the content of a given sub-total representation depends on the context of the propositional representation in which it finds itself, it is possible that the sub-total representation might warrant the assignment of a particular content (e.g. *is green*) only in the context of propositions involving food, say, and not predators (to use a slightly more biological example). Thus, the way that the system is representing *is green* does not meet the Generality Constraint, since the system cannot represent the proposition *the predator is green* even though it can represent other propositions involving greenness and predators. Thus, the content of the representation in question must be non-conceptual. Not that we didn't already suspect, or even know, that PDP representations are not necessarily systematic; but this provides another way of seeing why non-guaranteed systematicity, and therefore NCC, is so fundamental to PDP representation: it is bound up with the means of content ascription used for them.

4.3.4 Example 4: Context sensitivity II

It's a familiar point that very often, one has to change oneself in order to maintain a stable link with a particular aspect of a changing environment. If one wants to continue to point to an airplane that is moving across the sky, then one will have to move one's arm, head, and torso. This basic idea underlies the second, many-to-one kind of context-sensitivity: several different sub-total representations being assigned roughly the same (conceptual) content. This is the phenomenon illustrated by the now-familiar coffee example described by Smolensky [Smolensky, 1988] . There the same conceptual object, coffee, is represented by different hidden unit vectors, depending on whether the coffee is being thought about in the context of a cup, can, or spilled on the floor. Just as I have to alter my bodily configuration in order to continue to indicate the airplane as its position changes, so also the network must alter its hidden unit activations if it is to continue to represent coffee as the context changes.

This phenomenon is again a direct result of the fact that the locus of primary

content ascriptions to PDP representations is at the propositional level. Since the representation/content relation is fixed only at the propositional level, the sub-total representation required for representing a particular sub-propositional content will, in general, vary from propositional context to propositional context, depending upon the other particular sub-propositional contents to be represented. This yields the many-to-one relationship.

The upshot of this second type of context-sensitivity is that typically, a network will not be able to adopt an adequate representation of a particular sub-propositional content in any arbitrary propositional context. The same constraints that require a change of representation will sometimes outstrip the representational capacities of the PDP system. Therefore, the content of such PDP representations cannot be represented in all contexts, as (we have seen) a unified, conceptual framework requires; the content in question must be non-conceptual. Although there is no reason why a symbolic representational system could not employ a many-to-one relationship (though it would be difficult to dream up ways to exploit the relationship), the fact that the choice of symbol to use is not dictated by the context means that this case of context out-stripping representational capability could never occur; conceptuality is guaranteed (as long as the context-invariant semantics for the atoms are justified, which they probably never are). [14]

In the context of the CNM, such examples are common. Just consider a network that is simultaneously learning two routes that happen to intersect at a few points. There is nothing that necessitates that the network use the same code for the place in both contexts; as a matter of fact, it might facilitate learning the sequences to have the representations distinct. Since the representations have to play different roles in the different route contexts, finding one representation that can do both might be difficult (but see 4.4). [15]

4.3.5 Example 5: Sub-symbolic computation

Another reason for linking PDP with NCC falls out as a result of attempting to discover what it is that makes PDP representation an interesting representational genus, an alternative to symbolic representation. This is a question which Haugeland [Haugeland, 1991] has addressed. He argues that one might look in one of three possible places in order to find what it is that sets PDP representation apart:

1. the syntactic properties of the representations;
2. the relation between the representations and what they represent;
3. what the representations represent.

I would add a fourth place to look:

4. the contents of the representations. [16]

I disagree with Haugeland that 3 is the place to look, since I feel that PDP and classical systems could very well be representing the same things, even if they are doing so with very different contents. And the connectionist community has, in general, avoided purely syntactic (type 1) distinguishing criteria (since symbolic representations can also be, e.g., continuous); rather, most settle on *distribution* as the important aspect of PDP representation, and the best recent analysis of distributed representation, by van Gelder [van Gelder, 1991], sees distribution as superposition, a type 2 criterion. [17] Perhaps contents are merely a relation bewteen representations and the environment; perhaps 4 is subsumed by 2: I don't know. But it does seem that the standard account, along the lines of 2, that is offered as a way of distinguishing PDP representation from other types, is inadequate. For example, one can construct symbolic cognitive architectures out of distributed, superpositional representations [Touretzky and Hinton, 1988, Smolensky, 1987]. The fact that one can use distributed representations to construct *non*-classical architectures only raises the question: what is it that unites the set of representations (that may or may not be a subset of the set of distributed representations) that are characteristic of viable, non-classical cognitive architectures? I think it is by answering *this* question that we can see what it must be about PDP-based architecture, *in addition to* distribution (for I do think that distribution of a sorts is *necessary* for PDP to be interesting), that makes it a significant alternative to classical cognitive architecture.

Even if one admits that classical, symbolic representations can be distributed, it must be conceded that they would be of a very particular kind of distributed representation, a kind for which:

1. there is a conceptualized domain of objective values (objects, properties/relations, propositions, etc.) to be represented;

2. there is a scheme for assigning to certain configurations of the representational elements (symbols) these conceptual values as their contents; and

3. the computational operations of the system only manipulate the elements in groupings that correspond to these symbols.

Understanding symbolic representation in this way, we leave open the possibility of sub-symbolic representation: representations for which, relative to an appropriate conceptual semantic interpretation (1), and despite the fact that there might be identifiable symbols (configurations of elements that correspond to a conceptual object or property) (2), there nonetheless exist computational operations for which there is no conceptual interpretation for them, nor for the sub-propositional alterations that such operations make to the representations on which they act. For example, there might be no conceptual-level interpretation

for an operation that changes the weights of the CNM predictive map after one epoch of learning. This is not to say that there is *no* level of description above the level of talk about numeric weight changes which can account for what's going on after each epoch of learning in the predictive map. Such activity is indeed intentional, with representational significance, and therefore has a characterization beyond that of mere mechanism. It's just that this characterization must be non-conceptual, for the reasons given.

Another way of seeing that the operations of the CNM are sub-symbolic, and therefore admit of non-conceptual interpretations, is to attempt to think of the D mapping as a predication: the assertion that something has the property of stimulating the agent in a particular way. Changes in the D mapping weights are therefore changes in what is predicated. But to what are these predications being applied? As we saw in the first example, they aren't objective, conceptual places. Thus, the change in predication cannot be understood as the retraction of some objective, conceptual predications and the assertion of others. The change has non-conceptual significance only.

The claim is that it is *this* property of PDP representations, that they are sub-symbolic, that they can underwrite a non-conceptual understanding of a system, which distinguishes them for the purposes of cogntive architecture construction. It is this sub-symbolism which sets PDP representations apart as an alternative to other forms of computation, at least for my (perhaps narrow) concerns.

Now we are in a position to see why distribution (at least in one sense of the term) is necessary for PDP if it is to provide an alternative cognitive architecture through sub-symbolic computation. Take distribution, not in van Gelder's sense of superposition, but in a content-based variation of an earlier notion [Hinton et al., 1986]. Call it "dual distribution": 1) several elements are involved in carrying each content and 2) each element being involved, at some point, in carrying several contents. [18] For purposes of illustration, consider the case of a system that has both conceptual and non-conceptual content. The possibility of sub-symbolic operations requires that the conceptual contents be represented by sets of elements (that is, it requires *many-to-one* distributed representation [19]); if all representations of conceptual contents (symbols) were atomic, then there could be no way for an operation to make a representational change that has no conceptual interpretation. Now it may seem that sub-symbolic representation does not, strictly speaking, require the converse: that each element be used in representing more than one content (i.e., that it does not require one-to-many distributed representation). But if this were the case, then sub-symbolic operations, in addition to not having any conceptual interpretation, might not have any meaningful interpretation at all. To see why, consider the following. If the system employs only many-to-one distributed representation and no one-to-many distributed representation at all, then any given element (or a given value for a variable) is involved only in the representation of one particular content. The representations for any two contents will be entirely disjoint. Thus, any

representational operation that alters any proper subset of the elements of a content's representation will produce a non-content-bearing representation (which is really no representation at all). But such a system would not really provide an interesting alternative to symbolic representation, since, inter alia, any truly sub-symbolic operation would be rendered semantically void. Thus, one-to-many distribution is necessary after all, if the representational system is to have any interest. The type of representation that PDP is offering as an alternative to symbolic representation is an alternative in that it comprises dual distribution *and* sub-symbolic operations. [20]

4.4 The development of objectivity: Cognitive maps (example 6)

In example 4 we saw that having many-to-one context-sensitivity might be necessary, given how difficult it is to find one representation that can fulfil multiple roles in multiple contexts. But if a network can find such a common representation, then the objectivity and systematicity of its contents will be increased dramatically. Such a situation is depicted in figure 4.

At one point in the route learned by the agent in our example (solid lines), the trail doubles back. That is, the route involves going to (4 2) from the north, going east to (5 2), going west back to (4 2), then going south to (4 1). The thing to note is that the first time at (4 2), the network uses the code **.002 0.00 .176 .122** to represent the location, while the second time there it uses the different, but very similar code **.025 0.00 .021 .012**. Although the network, at earlier stages of development, used functionally distinct vectors for this purpose, the network has, through learning, forced these two representations together to such an extent that they are functional equivalents in the local context.

This yields a kind of generalization. Starting with the code for (4 3), the T mapping produces the first code for (4 2). But since this is functionally equivalent to the code used for the second appearance at (4 2), the network gives the correct result (in terms of both sensation *and* location vectors) when the agent moves *south* from the first (4 2) context (dotted line), instead of east, as it has always done before. A form of spatial generalization has occurred; the CNM is more than just a means of memorising a list of action/sensation sequences.

It was argued before that lack of systematicity is an indication of lack of objectivity. Thus, if it is claimed that figure 5 illustrates a case in which there has been an increase in objectivity, one might expect it to illustrate a corresponding increase in systematicity. It does indeed; systematicity is another "side-effect" of the convergence of many-to-one representations into a functional cluster. Since the "before (5 2)" and "after (5 2)" representations have merged into the same functional cluster, we now have a degree of systematicity of the form: $T(T(a, m), m^{-1}) = a$, where a =(4 2), $m =$ **move-east**, and $m^{-1} =$ **move-west**.

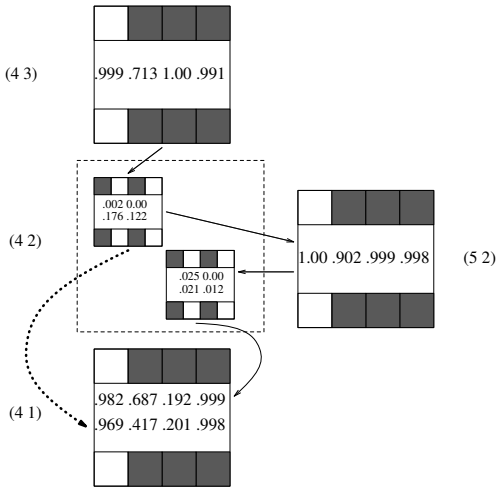

Figure 4: An illustration of how an increase in the objectivity of location representations yields generalization in the CNM. Selected locations on the route are labelled with each location's actual associated sensation vector (top four bits), the sensation vector that the agent (using the predictive map) predicted that the location would have (bottom four bits), and the code that the T mapping produced for the location (four numbers in between). Note that location (4 2) is represented by two diagrams (in the dashed box), one for each of the two ("before (5 2)" and "after (5 2)") contexts. Location (4 1) has two codes depicted: the top one is produced by travelling from (4 2) via (5 2) (solid lines); the (functionally equivalent) bottom one is the result of moving from (4 2) to (4 1) directly (dotted line).

At the beginning I argued that the development of intentionality may itself require an intentional analysis. I also cited evidence that suggests that this developmental process may have a complex, fine-grained structure, rather than being manifested in a single, inexplicable leap from mere mechanism to complete conceptuality. Thus, the appropriateness of an architecture for grounding this intentional characterization will depend upon its ability to march in step with the fine shades of increasingly perspective-independent contents that best characterize the development of objectivity. I think the example just considered suggests that this is another reason why PDP architectures, such as the CNM, and non-conceptual content are a natural match.

5 Classical cognitive architecture and NCC

In closing, it should be pointed out that this exploration of the connection between PDP and NCC will still be of value even if, in addition to PDP's suitability, and the arguments of section 3 notwithstanding, some simple variations on classical architecture also turn out to be suitable for non-conceptual content. The value will remain, since in such a situation, one might still be able to appeal to other advantageous properties of PDP to isolate it as the cognitive architecture of preference. In any case, the work that has been pursued here, of finding a match between PDP and some kind of content is a *necessary* condition for it to be a successful cognitive architecture.

6 Acknowledgements

This paper benefited not only from its presentation at SNCC 92, but also from presentations to the COGS and Psychology Research in Progress seminars at the University of Sussex. Accordingly, I would like thank the following for their comments and suggestions: Maggie Boden, Matthew Elton, Chris Thornton, Michael Wheeler; George Butterworth, Lindsay Parkin, and Kim Plunkett. Also my thanks to an anonymous referee.

Notes

[1]See [Woodfield, 1993] for a discussion of, and pointers to other work on, the philosophical problems that arise for the notion of conceptual change

[2]Thus, my motives and ends are distinct from those of, e.g., Plunkett [Plunkett, 1992] and Bates & Elman [Bates and Elman, 1992], who also seek to apply PDP to developmental phenomena, but who, I think it is fair to say, are not primarily interested in an *intentional* (e.g., content-involving) account. Rather, their work focuses on the provision of the *mechanisms* underlying developmental behaviour.

[3]I should point out that none of the cited authors characterize non-conceptual content in exactly the same way, nor does my notion exactly agree with any one of theirs. But the differences are largely irrelevant for the purposes of this paper.

[4]For terminology buffs: I continue to use the designation "Parallel Distributed Processing" or "PDP" instead of, e.g., the less passé "connectionism" because it gives a better indication of what is central to the relevance of such networks to this research.

[5]Where X is, variously: necessary systematicity; structured representations and operations sensitive to that structure; isolable, causally efficacious syntactic entities that mirror conceptual contents; or whatever.

[6]And there is work [Chalmers, 1990, Elman, 1990, Pollack, 1990, van Gelder, 1990] that *suggests* that PDP *can* achieve such contingent systematicity. Of course, it is clear that there could be a PDP implementation of a systematic architecture. But this would be as insufficient for cognitive science (in the broad, recalcitrant-phenomena-involving sense of "cognitive science" being used in this paper) as any other classical architecture, and for the same reasons. That is, it would not be able to account for the non-systematic phenomena, and it would leave inexplicable the transitions between non-systematicity and systematicity – the development of objectivity.

[7]Thus, if the predictive map is to be implemented in a PDP network, it must be a recurrent network. In the simulations discussed here, it is implemented as a *simple* recurrent network [Elman, 1990]

[8]Thus, the cognition under investigation here is at least partly social, in that it relies upon the cooperation of multiple agents, and the interaction between agents and the social technology of *routes*.

[9]One can think of this review in a literal sense; perhaps it could take place during dreaming?

[10]Furthermore, one will not be able to understand it as representing *any* objective place, nor will one be able to understand *any other* location code as representing the place ten units east from the lair.

[11]One might think, then, that the way to increase one's objectivity is to make one's T mapping of a region more and more independent of what one believes to be located in that region. This is to equate objectivity with Generality and systematicity, as I have been doing so far. But there is another view of what objectivity is: the ability arbitrarily to change one's T mapping, depending on whether one is thinking of the movement of, e.g., oneself, air, or light. That is, perhaps objectivity isn't Evans' "thinking from no point of view"[Evans, 1982, page 152], nor Nagel's "view from nowhere"[Nagel, 1986], but rather Smith's "view from somewhere"[Smith, 1992], or, perhaps better, "a view from anywhere"[Cussins, 1990, note 103, page 428]. This latter view of objectivity would seem to give even more pride of place to non-systematic architectures such as PDP.

[12]This is typically done with a large degree of wishful thinking; the comments in [McDermott, 1981] continue to be germane.

[13]By sub-total representations I mean sets of hidden units that are strictly less than the set of all hidden units in the system. All I mean to achieve by using this term is an emphasis that I am not talking about representations with complete, propositional contents (which in simple systems might not vary their contents across contexts), but rather constituent sub-patterns of such representations.

[14]Of course, there can be a many-to-one relationship between symbolic representations and *referents*; one *can* represent, in a production system say, the same block with either of the representations **block-57** or **The block X such that is-above(X, block-23)** (assuming that block 57 is the only block above block 23). But this would not be a many-to-one relationship between vehicles and *contents*; the contents of the two representations are different, even though their referents are the same.

[15]Why isn't this case the same as the one in the previous note, i.e., a case of many-to-one

vehicle/reference, but not many-to-one vehicle/content? Perhaps the best way to view these CNM representations is as having distinct contents, such as *place X in the context of route A* and *place X in the context of route B*? There is not enough space here to fully refute this, but let me just say that if one went all the way down this slippery slope, there would be no possibility of valid inference. For example, consider the inference from *the wall is grey* and *the wall is in front* to *there is something which is both grey and in front.* This is of the form $P(a), G(a); \exists x : P(x) \wedge G(x)$. By the reasoning being used against the many-to-one vehicle/content view of the CNM, one would have to conclude that the wall is being represented by distinct contents, since it is being represented in two different contexts. But this would mean the inference would be of the form $P(a), G(b); \exists x : P(x) \wedge G(x)$, which is invalid. We do not, in general, want to individuate contents as finely as the contexts in which they appear.

[16]Haugeland does talk about the "content" of a representation, but by this, he means what it represents, and not what I have been taking "content" to mean in this paper.

[17]What is superposition? Roughly, if R_1 represents i_1 and R_2 represents i_2, and R_1 and R_2 use the very same representational resources, then the representations of i_1 and i_2 are superposed; but the details are irrelevant here. The point is that superposition is a relation between representational vehicles and what they represent and is thus a criterion of type 2.

[18]This is different from superposition in that (at the least) it does not require representations to represent more than one item *at the same time.*

[19]This is not to be confused with many-to-one context-sensitivity, discussed earlier.

[20]van Gelder argues briefly (and, I think, successfully) that dual distribution is not sufficient for an interesting representational alternative. Something more, I claim, is needed. The notion of superposition that van Gelder puts forward certainly seems *compatible* with sub-symbolism, but I do not know if it is *necessary* for it, as dual distribution is.

References

Anderson, J. (1983). *The Architecture of Cognition.* Harvard University Press, Cambridge.

Bates, E. and Elman, J. (1992). Connectionism and the study of change. Technical Report 9202, Center for Research in Language, University of California, San Diego.

Chalmers, D. (1990). Syntactic transformations on distributed representations. *Connection Science*, 2:53–62.

Chrisley, R. (1990). Cognitive map construction and use: A parallel distributed processing approach. In Touretzky, D., Elman, J., Sejnowski, T., and Hinton, G., editors, *Connectionist Models: The Proceedings of the 1990 Connectionist Models Summer School.* Morgan Kaufmann, San Mateo.

Chrisley, R. (1991). A hybrid architecture for cognitive map construction & use. *Artificial Intelligence & the Simulation of Behaviour: Special Issue on Hybrid Models of Cognition.* No. 78.

Crane, T. (1992). The non-conceptual content of experience. In Crane, T., editor, *The Contents of Experience.* Cambridge University Press, Cambridge.

Cussins, A. (1987). Varieties of psychologism. *Synthese*, 70:123–54.

Cussins, A. (1990). The connectionist construction of concepts. In Boden, M., editor, *The Philosophy of Artificial Intelligence*, pages 368–440. Oxford University Press, Oxford.

Davies, M. (1990). Thinking persons and cognitive science. *AI and Society*.

Diamond, A. (1988). Differences between adult and infant cognition: Is the crucial variable presence or absence of language? In Weiskrantz, L., editor, *Thought Without Language*. Oxford University Press, Oxford.

Elman, J. (1990). Finding structure in time. *Cognitive Science*, 14:179–212.

Evans, G. (1982). *The Varieties of Reference*. Oxford University Press, Oxford.

Fahlman, S. (1988). Faster-learning variations on back-propagation: an empirical study. In Touretzky, D., Hinton, G., and Sejnowski, T., editors, *The Proceedings of the 1988 Connectionist Models Summer School*, pages 11–20, San Mateo. Morgan Kaufmann.

Fodor, J. (1985). Fodor's guide to mental representation – the intelligent auntie's vade-mecum. *Mind*, 94(373):76–100.

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

Fodor, J. and Pylyshyn, Z. (1988). Connectionism and cognitive architecture: A critical analysis. In Pinker, S. and Mehler, J., editors, *Connections and Symbols*. MIT Press, Cambridge.

Harris, P. (1989). Object permanence in infancy. In Slater, A. and Bremner, G., editors, *Infant Development*, pages 102–121. Lawrence Erlbaum, Hove.

Haugeland, J. (1991). Representational genera. In Ramsey, W., Stich, S., and Rumelhart, D., editors, *Philosophy and Connectionist Theory*. Lawrence Erlbaum Associates, Hillsdale, NJ.

Hinton, G., McClelland, J., and Rumelhart, D. (1986). Distributed representations. In *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, volume 1, pages 77–109. MIT Press, Cambridge.

Hood, B. and Willatts, P. (1986). Reaching in the dark to an object's remembered position: Evidence for object permanence in 5-month-old infants. *British Journal of Developmental Psychology*, 4:57–66.

McDermott, D. (1981). Artificial intelligence meets natural stupidity. In Haugeland, J., editor, *Mind Design: Philosophy, Psychology, Artificial Intelligence*, pages 143–60. MIT Press, Cambridge.

Nagel, T. (1986). *The View from Nowhere*. Oxford University Press, Oxford.

Peacocke, C. (1992). Scenarios, contents & perception. In Crane, T., editor, *The Contents of Experience*. Cambridge University Press, Cambridge.

Plunkett, K. (1992). Connectionism and developmental theory. *British Journal of Developmental Psychology*, 10:209–254.

Pollack, J. (1990). Recursive distributed representations. *Artificial Intelligence*, 46.

Ramsey, W., Stich, S., and Garon, J. (1991). Connectionism, eliminativism, and the future of folk-psychology. In Ramsey, W., Stich, S., and Rumelhart, D., editors, *Philosophy and Connectionist Theory*. Lawrence Erlbaum Associates, Hillsdale, NJ.

Rosenbloom, P., Laird, J., Newell, A., and McCarl, R. (1992). A preliminary analysis of the SOAR achitecture as a basis for general intelligence. In Kirsh, D., editor, *Foundations of Artificial Intelligence*, pages 289–326. MIT Press, Cambridge.

Smith, B. (1992). The owl and the electric encyclopedia. In Kirsh, D., editor, *Foundations of Artificial Intelligence*, pages 251–288. MIT Press, Cambridge.

Smolensky, P. (1987). On variable binding and the representation of symbolic structures in connectionist systems. Technical Report 355-87, Department of Computer Science, University of Colorado.

Smolensky, P. (1988). On the proper treatment of connectionism. *Behavioral and Brain Sciences*, 11:1–74.

Strawson, P. (1959). *Individuals*. Methuen, London.

Touretzky, D. and Hinton, G. (1988). A distributed connectionist production system. *Cognitive Science*, 12:423–466.

van Gelder, T. (1990). Compositionality: A connectionist variation on a classical theme. *Cognitive Science*, 14:355–384.

van Gelder, T. (1991). What is the 'D' in 'PDP'? In Ramsey, W., Stich, S., and Rumelhart, D., editors, *Philosophy and Connectionist Theory*. Lawrence Erlbaum Associates, Hillsdale, NJ.

Woodfield, A. (1993). Do your concepts develop? In Hookway, C. and Peterson, D., editors, *The Proceedings of the 1992 Royal Institute of Philosophy Conference on Philosophy and the Cognitive Sciences*, volume 34, Cambridge. Cambridge University Press. Supplement to *Philosophy*.