Reclaiming Symbols: The Role of Symbols and Representations in an Enactive Account of the Mutual Scaffolding of Concepts and Experience

Joel Parthemore*, Anthony Morse+ and Tom Ziemke+

*PAICS Research Group
Centre for Research in Cognitive Science
Department of Informatics
University of Sussex, Brighton, UK
J.E.Parthemore@sussex.ac.uk

+SCAI Lab
School of Humanities and Informatics
Högskolan Skövde
Skövde, Sweden
Anthony.Morse@his.se, Tom.Ziemke@his.se

This is a draft paper only. Please do not cite.

1 Introduction

An enactive account of symbols will differ from a standard account of them, just as an enactive account of concepts will differ from other, non-enactive theories of concepts. Words like "symbol", "representation" and "concept" come with a lot of baggage. Therefore this paper begins with a careful definition of terms, to make it clear what should be retained and what should not.

Section Two sets forth an agenda for reclaiming symbolic and representational language from where some might banish them, to give them a critical role within a theory of concepts. The idea of concepts as associations of sensory-motor contingencies is a powerful one but, arguably, it is incomplete. An enactive theory of concepts shows how associationist and symbolic accounts of concepts can come together in one, comprehensive account.

Section Three builds on this structure to offer a model for the mutual scaffolding of concepts and experience, within a context where neither ultimately takes precedence and both are, for the conceptual agent, constantly in play, each giving rise to the other. This notion of mutual scaffolding is a simple but, one might hope, powerful idea. Before closing the paper will give some consideration to how the model described here might be implemented within an extension of the SEER-3 robotic system.[1]

1.1 Enaction

Enaction is one of those terms-of-fashion that mean different things to different people. So Alva Noë (Noë, 2004) specifically states that when *he* uses enaction he does not mean to entail what writers of the autopoietic

[1] . . . As described in (Chrisley and Parthemore, 2007).

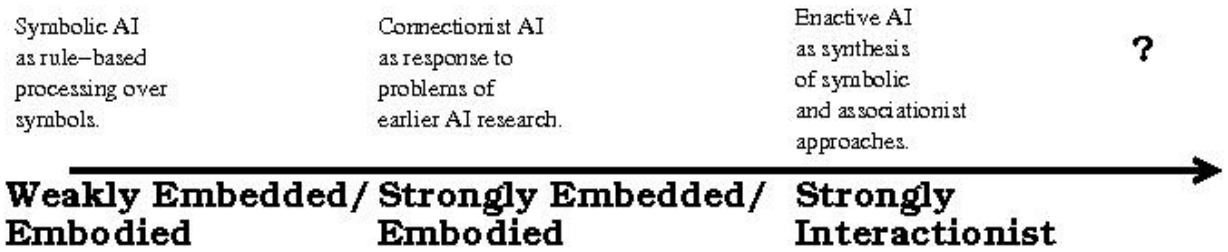

Figure 1: The progression of thought in cognitive science.

tradition mean by the term. If Noë can be seen as the more conservative end of a spectrum, then the other end may best be epitomized by Evan Thompson (Thompson, 2007), for whom mind and life are at the least continuous if not ultimately the same thing (under the autopoietic slogan "cognition is life"). In between are people who find Noë's sensory-motor account incomplete on various grounds but who might not want to take on all of the assumptions that autopoiesis is often understood to entail and who balk at the vaguely defined concept of "second-order autopoiesis" that is supposed to describe system organization at the level of multicellular organisms.

Minimally an enactive perspective of cognition is taken to include but go beyond providing an account of how the cognitive agent is embedded in its environment and embodied in a particular physical form. It sees a first-person account of cognition (as of science in general) to be not just a useful but a necessary part of an overall account. It typically emphasizes the continuous (e.g., the continuity between agent and environment) over the individuable and discrete. It takes an agent/environment, internal/external distinction to be both conceptually necessary and, at the same time, meaningful *only with respect to an observer.*

In the context of the history of cognitive science, enaction can be seen as part of a broader movement away from disembodied and "purely" symbolic accounts of cognition[2] that treated e.g. agent as independent from environment, sensory input as independent from motor output, mind (software) as independent from brain (hardware), cognition as independent from life, syntax ("grammatical" structure) as independent from semantics (meaning). As such, the enactive perspective as currently conceived, or as presented in this paper, may just be a step toward something else.

1.2 Concepts

As Stephen Laurence and Eric Margolis have pointed out(Margolis and Laurence, 1999), when one is asking what concepts are, not much can be taken for granted, even as to whether concepts are best described as abstract objects or as abilities. If concepts *are* objects, then they might typically be described as "mental representations" or "sub-propositional components of thought". But some writers, from Evans (Evans, 1982) to Noë (Noë, 2004) have taken concepts to be, at least in the first instance, abilities. Evans offered his Generality Constraint on conceptual mental content as a good place to start, being something that pretty much everyone who talks about concepts can agree on.

> When we say that a subject's understanding of a sentence, *'Fa'*, is the result of two abilities (his understanding of *'a'* and his understanding of *'F'*), we commit ourselves to certain predictions as to which other sentences the subject will be able to understand; furthermore, we commit ourselves to there being a common, though partial, explanation of his understanding of several different sentences. If we hold that the subject's understanding of *'Fa'* and his understanding of *'Gb'* are

[2]Of course even the most GOFAI-based models of cognition are embedded and embodied in a weak sense: computers as symbol-processing systems with no connection to their environment and whose physical instantiation is irrelevant are at best an idealization, a useful fiction. Further, as this paper will argue, no symbolic system is "purely" symbolic in the sense that this is commonly taken to mean.

> structured, we are committed to the view that the subject will also be able to understand the sentences *'Fb'* and *'Ga'*. (Evans, 1982, p. 101)

This is to say, if one can understand "the car is blue" and "the sky is large", one should understand as well "the car is large"and "the sky is blue". Evans' Generality Constraint implies two properties of concepts: systematicity and productivity.[3] Systematicity means that the same concepts can be systematically deployed across different contexts. Productivity means that a finite number of concepts make possible a potentially infinite number of complex constructions.

To put this a different way, conceptual mental content is mental content that abstracts away from the immediate perceptions or the immediate experience of the moment. It has, as it were, one hand in the past and the other in the future, an idea that Lawrence Barsalou(Barsalou et al., 2007) has talked about. If this is right, then paradigmatic conceptual mental content is mental content *not* of the moment; paradigmatic non-conceptual mental content is content strictly of the moment, without connection to any other moment, remembered or anticipated.[4] To the extent that conceptual mental content can be individuated then, the individuated entities are concepts.

Other properties typically assigned to concepts is that they are (at least in part) under endogenous control and that they are (at least in part) subject to rational revision. A fourth possible property, articulability, is, in the views of the authors, problematic, as it requires making a tight link between concepts and language.[5] This can be seen as having a number of unfortunate consequences:

- It denies, *a priori*, the possibility of non-linguistic agents (such as non-human animals) possessing and employing concepts, even though they may meet all the other requirements.
- It requires, *a priori*, for concepts to be primarily entities in the shared space of public discourse, as opposed to the private space of individual agents.

1.3 Enactive Concepts

As with cognition more generally, an enactive theory of concepts includes but goes beyond providing an account of how the concept-possessing-and-using agent (henceforth the "conceptual agent" for shorthand) is embedded in its environment and embodied in a particular physical form. It locates concepts not in the conceptual agent (say, as internal, *a priori* mental representations) nor in the agent's environment (say, as an external set of affordances). Concepts are enacted out of the dynamic interaction of the agent with the agent's environment and so, to the extent they can be located anywhere, they lie between the agent and the environment. Nonetheless there will be occasions when it is appropriate or indeed necessary for either agent or environment to be emphasized.

[3] . . . Though Evans does not describe the Generality Constraint this way, probably under the influence of Gottlob Frege, who, in his war of words with Benno Kerry, took the treatment of concepts as objects for *any* description to be a serious mistake: concepts were things whose nature could only be hinted at. For an excellent discussion of this in light of similar themes in Wittgenstein's work, see (Jolley, 2007).

[4] This is, admittedly, not the standard way of defining non-conceptual content, which is more frequently to oppose it to conceptual content, then attempt to provide a clear account of conceptual content: e.g., "Content is nonconceptual just if it can be attributed to a subject without ipso facto attributing to that subject mastery of the concepts required to specify it" (Bermudez, 2007, p. 55) or "The core of the idea of nonconceptual content is a modal claim to the effect that perceivers need not possess concepts corresponding to everything that they are capable of perceptually discriminating." (Bermudez, 2007, p. 69) However the authors believe it is not incompatible with that account, and some discussions with non-conceptualists bear that up.

[5] Of course many philosophers, from John McDowell to Jerry Fodor, have made just this sort of connection between concepts and language.

2 Reclaiming Symbols

People debate the value of symbolic language and whether there are or are not symbols in the brain without, much of the time, saying what they mean by "symbol". Symbols are commonly taken to be [6]:

1. Amodal, as opposed to modal (i.e., sensorily grounded).
2. Context-free, permitting a strict syntax/semantics separation.
3. Discrete, as opposed to continuous.
4. Arbitrary, so that any symbol can be substituted for any other symbol, provided it is done consistently.[7]

Taken at face value, all of these are deeply problematic.

1. Symbols have to be grounded *somehow*. This is the familiar symbol grounding problem, as reported by (Harnad, 1990). Semantics is never free floating.
2. Symbols are only ever meaningful with respect to some context. At the least there is always a shared social context in which they are learned and applied, a point that is emphasized in (Harvey, 1992). Remove them too far from their context of origin, and they really will be meaningless.
3. Symbols change over time. [8] At what point does a symbol cease being one symbol and become another? Also symbols never stand on their own but exist in relation to other symbols, some of which may be quite similar. At what point are two symbols not two different symbols but the same symbol?
4. The relationship between the form (syntax) and the meaning (semantics) of a symbol often is seemingly *not* arbitrary. So for example, the symbol "3" has three prongs.

Serial computers are often cited as the paradigmatic symbol-processing machines in the sense of the four properties above. "All" they are doing, it is said, is blindly applying preset rules to strings of "meaningless" symbols. In what sense, after all, does the computer understand what it is doing?

Yet this is precisely the point that Terry Winograd and C.F. Flores make (Winograd and Flores, 1986): even in the case of computers, the strict separation of syntax and semantics is based on an ideal, not on reality. The human agent is necessary to give meaning to the signs being manipulated by the computer, and without the human agent as part of the process, the computer isn't doing symbol manipulation at all, even in the impoverished sense of number crunching. We treat computers as idealized machines, completely separated from their environment, "unable to make a mistake", for the same reason we treat symbols as amodal/discrete/context-free/arbitrary, likewise idealizations: because we find it conceptually convenient.

2.1 Enactive Symbols

A lot of the difficulties with symbols go away once one treats these four properties as idealizations rather than absolute requirements: to wit, a symbol is a symbol *to the extent* that it meets these requirements. In that case there will not, probably, be a hard-and-fast distinction between symbols and non-symbols but a continuum, from the less interpretably symbolic to the more interpretably symbolic.

[6] A longer but related list can be found in (Harnad, 1990). Note that Harnad's list should probably be seen as more demanding than this one. Therefore if this list raises difficulties for symbols, so does his.

[7] So if you consistently uses the word "blue" to mean *apple*, then over time anyone familiar with you will in time adjust to your peculiar usage.

[8] Compare this with the discussions of conceptual change in e.g. (Woodfield, 1994).

Press people, and will anyone claim that symbols can be interpreted outside of *any* context, or that symbols are *always* arbitrary? Probably not. If this is right, then the difference between GOFAI symbolists and nouveau AI associationists (or connectionists) appears to be a difference of emphasis, albeit one that, still, hides difference of substance. As this paper will attempt to establish, the real question may be not, are there symbols in the brain?, but rather, for the symbolists, can an account of cognition be given *solely* in terms of symbolic language; and, for the associationists, can an account of cognition be given without *any* resort to symbolic language?[9]

The answer, as this paper will suggest, may be that symbols and symbolic language are necessary (*pace* the strict associationists) but not sufficient (*pace* the strict symbolists). Indeed, it will be the position of this paper that an enactive account of concepts bridges the apparent gap between the associationist and the symbolic accounts of concepts. In between the clearly symbolic level and the clearly associational level, some further account is called for that is not beholden to associations or to symbols. This is the same goal sought by Peter Gärdenfors in his conceptual spaces theory of concepts (Gärdenfors, 2004), and indeed, it will be argued, his conceptual spaces theory may well be a good platform from which to build an enactive account.

A further issue with symbols needs to be addressed. Inman Harvey, among others, has long argued that it is in the nature of a symbol, as a form of representation, that it is *not* a symbol in the absence of an agent to give it meaning. (This, again, is why the computer is not, on its own, doing symbol processing.) Failure to acknowledge the role of the observer in the act of representing can lead to a lot of confusion. And yet, "the underlying assumption of many is that a real world exists independently of any observer; and that symbols are entities that can 'stand for' objects in this real world – in some abstract and absolute sense. In practice, the role of the observer in the act of representing something is ignored."(Harvey, 1992, p. 5) Again: "The gun I reach for when I hear the word *representation* has this engraved on it: 'When P is used by Q to represent R to S, *who is* Q *and who is* S?"' (Harvey, 1992, p. 7, emphasis original)

With these thoughts in mind, and attempting in general to sharpen the definition, the properties of symbols can be restated as follows:

1. Modally grounded, but in such a way that the links back to the modal grounding may be difficult or impossible to reconstruct.

2. Consisting of a sign (syntax) - meaning (semantics) dyad (*per* Wittgenstein[10]) where, to some practical extent, the sign can be distinguished from the semantics and the symbol from any particular context of interpretation. The extent to which the symbol will be recognizable as a symbol will be the extent to which it abstracts away from any particular context of interpretation.

3. Along with this, a practical extent to which the symbol is individuable (discrete, as opposed to continuous with other symbols or with a non-symbolic background).

4. Also along with (2), an apparent arbitrariness between sign and meaning precisely in relation to the extent to which the symbol has abstracted any from any particular context of interpretation. Note that "true" arbitrariness is not required, only that the relation between sign and meaning, if any, has been lost.

5. An observer: someone for whom the symbol is symbolizing, and someone the symbol is symbolizing to. (These could be one and the same.)

[9] . . . Aside from the trivial sense, in which language will be necessary for communicating the account. The question is whether symbols and symbolic language play a necessary *intrinsic* role in the account. That is to say, symbols are the *vehicle* of the theory; but are they also part of the content?

[10] As Wittgenstein pointed out, much confusion can arise from confusing the symbol with the sign. "The sign is the part of the symbol perceptible by the senses. Two different symbols can therefore have the sign... in common – they then signify in different ways.... In the language of everyday life it very often happens that the same word signifies in two different ways – and therefore belongs to two different symbols – or that two words, which signify in different ways are apparently applied in the same way in the proposition." (Wittgenstein, 2001, 3.32)

Possessing the first four properties but lacking the fifth, a certain object can be symbol-like, but it is not properly a symbol. Through the rest of the paper, such objects will be referred to as "symbol-like".

2.2 Representations

"Representation" is typically used co-extensively with "symbol". Symbols, it is said, *represent*, and representations are *symbolic*. Nonetheless some useful distinction can be made.

A symbol can be defined, again, as a sign-semantics dyad. In principle, the relationship between the two can, but as we have seen need not be, arbitrary.[11] In particular, the structure of the sign is not supposed to be relevant.

In contrast, the relationship between a representation and its represented must be non-arbitrary, if the thing is to be an effective representation: i.e., there should, up to some point, and at some level of abstraction away from details deemed irrelevant in the representational context, be a structural isomorphism between aspects of the representation and aspects of the represented. A painting of a dog will make a poor representation of a waterfall unless and until the viewer of the painting can establish the (presumably hidden) isomorphism. Note that as Nelson Goodman famously pointed out, resemblance cannot yield representation(Goodman, 1976); rather it is the case that representation yields resemblance.

Representations may simply be symbols with the requirement for relative arbitrariness relaxed. Remember what was said earlier: that the extent to which a symbol is recognizable as a symbol is the extent to which it abstracts away from any particular context of interpretation. The further abstracted away the symbol is from the initial context(s), the less obvious its relationship back to the initial context(s) will be and the more arbitrary the relationship between form and meaning will appear. With representations, the relationship between form and meaning is still apparent. So on this account, and to the extent that representations can meaningfully be distinguished from symbols, symbols may be understood as an impoverished form of representation in which the relationship between form and meaning has for practical purposes been lost. It's not that symbols are unstructured – they can't be, if one symbol is meaningfully to be individuated from any other symbol – but that the structure is irrelevant to the interpretation. Symbols are unstructured *relative to the domain of interpretation.*

Note that this impoverishment works to the symbols' advantage. A relatively sparse structure – just enough to distinguish one sign from another sign and so one symbol from another symbol – can carry with it a great deal of information, indeed arbitrarily much.

With that distinction in mind, the requisite properties of representations can be given as follows:

1. Modally grounded, but in such a way that the links back to the modal grounding may be quite indirect.
2. Consisting of a representation (syntax) - represented (semantics) dyad where both the representation and the represented are structured entities but where the representation is more unlike the represented than it it is like the represented. The extent to which the representation is a useful representation is a combination of the extent to which it still bears a recognizable relationship back to its referent and, at the same time, the extent to which it abstracts away from particular contexts of application and so can be applied across many different contexts.
3. Along with this, a practical extent to which the representation is individuable (discrete, as opposed to continuous with other representations or with a non-representational background).
4. An observer: someone for whom the representation is representing, and someone the representation is representing to.

[11] (Harnad, 1990) gives the example of Chinese characters.

2.3 Concepts and Symbols

The extent to which something is recognizably a symbol is the extent to which its meaning can be held invariant across contexts. A mark ß that means one thing today and another tomorrow is only weakly, at best, a symbol. In the same way, the extent to which something is recognizably a concept is the extent to which its meaning can be held invariant across contexts.[12] A successful concept abstracts away all of the irrelevant detail to get at what is *most likely* to be invariant and hence re-applicable. A successful symbol does the same thing. No wonder then that at the level of self-conscious introspection, concepts look a great deal like words, which are paradigmatically symbolic.

Given these considerations, one might be tempted to think that concepts just are symbols, or perhaps a type of symbol. But that would be wrong, on two accounts:

1. Taking Harvey's lead, this paper suggests that a symbol is not, properly speaking, a symbol unless someone is using it to represent something to someone. Without that something can, at most, be symbol-like. It's not clear, however, that such a restriction should, unmodified, apply to concepts. In particular, a distinction needs to be made between what one might call the *operational* and *observational* contexts of concepts: that is, when concepts are being employed by an agent non-reflectively, without consideration of the concepts *as* concepts; and when concepts are being employed reflectively, where the object of awareness is the concepts themselves. (This could be when one is introspecting about one's own concepts or when one is observing them in others, via e.g. a folk psychology theory of other minds.) It might seem that most of the time, concepts are being employed in the first sense and not the second; and that the second is really a special case. Most of the time conceptual agents just get on with using concepts, not thinking about them.

2. There is no obvious reason save stipulation to think that the content of concepts must be strictly conceptual content, any more than the content of representations must be strictly representational.[13] Indeed, on Jerry Fodor's informational atomism account Fodor (1998), *none* of the content of (most) concepts is conceptual content. This must be what he means when he talks about concepts as (unstructured) atoms: i.e., that they are *conceptually* atomic: they are atomic with respect to the domain of application. After all, if they were unstructured in *all* dimensions, then they would be indistinguishable one from another. Since the content of Fodor's atoms is not conceptual content, it must be (some form of) non-conceptual content. Likewise on Jesse Prinz's proxytypes account, which embraces the informational semantics without the (conceptual) atomism Prinz (2004, p. 164), *part* of the content of concepts is conceptual – i.e., concepts decompose *in part* into other concepts – but Prinz allows for other kinds of content as well.[14]

It is the position of this paper that an account of concepts cannot solely be given in terms of concepts any more than an account of representations could be given solely in terms of representation. Concepts, representations, symbols all need to be grounded. Unless meaning just comes by nomic relations between concepts and their referents in a presupposed objective reality [15], then some account must be given of how objective meaning arises from e.g. subjective experience. This is Adrian Cussins' position Cussins (1990), and he believes that, in order to provide that account, concepts require both a conceptual and a non-conceptual specification. Without necessarily embracing Cussins' CCC model (Connectionist Construction of Concepts), this paper adopts that view.

[12] This implies a continuum between the conceptual and the non-conceptual, which conceptualists like McDowell are eager to deny and non-conceptualists like Adrian Cussins (e.g. in Cussins (1990)) are eager to embrace. Thanks to Ron Chrisley for making that observation.

[13] Take a painting of a waterfall as a representation of a waterfall. Many of the aspects of the painting – e.g., the type of canvas used – will have nothing to do with representing the waterfall.

[14] Indeed specifying the content of concepts solely in terms of conceptual content raises problems that threaten paradox. For example, what does one make of the concept of "concept" itself, since any specification of the contents of the concept will make use of concepts that presuppose it?

[15] But consider Harnad: "The standard reply of the symbolist... is that the meaning of the symbols comes from connecting the symbol system to the world 'in the right way.' But it seems apparent that the problem of connecting up with the world in the right way is virtually coextensive with the problem of cognition itself." Harnad (1990, p. 340)

So, concepts are both *more* than symbols in that they require a non-symbolic, non-representational account in addition to the symbolic/representational one; and *less* than symbols, in that symbols involve someone representing something to someone – someone to recognize the symbols *as* symbols – and concepts do not. On the one hand, the conceptual agent – say, the philosopher or neuroscientist – who is offering an account of concepts cannot step outside her role as an observer, so when she offers an account of concepts she does so as an observer, at least in part, of the concepts she is accounting for. At the same time, she requires means to talk about contexts in which there logically is no observer – e.g., when a conceptual agent employs concepts non-reflectively – without, tacitly, opening the door to homunculi.

When the content of concepts is being expressed conceptually, then the properties of concepts will look like this[16]:

1. Interpretable as conceptually atomic symbols (e.g., lexical concepts as represented by the word "dog") or conceptually structured representations (e.g., Prinz's proxytypes).

2. Modally grounded (in fact, as the paper will shortly consider, sensory-motor grounded) , but in such a way that the links back to the modal grounding may be difficult or impossible to reconstruct.

3. Consisting of a sense-reference dyad where the extent to which the concept is recognizably a concept is the extent to which it usefully abstracts away from any particular context of interpretation.

4. Along with this, a practical extent to which the concept is individuable (discrete, as opposed to continuous with other concepts or with a non-conceptual background).

5. An agent to possess and employ the concept, whether or not the agent is aware of employing or possessing the concept.

2.4 Symbols and Associations

More needs to be said about how symbolic and associationist account of cognition in general or conceptual mental content in particular come together, and what it means in practice to talk, as (Harnad, 1990) does, of synthesizing the two approaches, and including associations and symbols, connectionist approaches and symbolic approaches, within the same account. [17] As he notes (and many others have noted as well), symbolic accounts seem particularly well suited to what it often referred to as high-level cognition: linear, self-conscious thought, logically structured, seemingly language-like. [18] Associational accounts seem particularly suited to what is often referred to as low-level cognition: parallel, unconscious or even subpersonal cognition directly tied to sensory-motor engagement.

Put another way, associationist methods and symbolic methods each have things they do well, things they do poorly. Symbolic methods are poor at handling things like typicality effects, priming effects and conceptual fuzziness. Associationist methods have their own set of problems. "In particular, although, like everything else, their behavior and internal states can be given isolated semantic interpretations, nets fail to meet the compositeness... and systematicity... criteria listed earlier: The patterns of interconnections do not decompose, combine and recombine according to a formal syntax that can be given a systematic semantic interpretation." (Harnad, 1990, p. 338) The result, as Fodor and Pylyshin [ref] have noted, is not infrequently a de-emphasis of the importance of these properties to structured thought.

[16] Note that this is a very broad and permissive definition of concepts, compared to e.g. the very strict requirements that McDowell places on concepts. On this account many non-linguistic animals are also likely to qualify as conceptual agents. For arguments and empirical results supporting this conclusion, see e.g. (Newen and Bartels, 2007).

[17] It should by this point in the paper be clear that the role that is being set out for symbols is close to the role Barsalou has consistently given them and continues to give them in a number of recent papers, including (Barsalou et al., 2003) and (Barsalou, 2008) and is not necessarily incompatible with the avowedly "anti-symbolic" account of enactive concepts offered by Vittorio Gallese and George Lakoff (Gallese and Lakoff, 2005), though they are keen at points in that paper to stress the completeness of a sensory-motor account, which is precisely what this paper is denying. See 3.1.

[18] This is the cognitive level at which Fodor's language of thought hypothesis and informational atomism seem most comfortably at home.

The underlying question is not whether a symbolic account can account for the emergence of associations or whether an associational account can account for the emergence of symbols: both approaches are up to the job. As Harnad writes, "Connectionist networks can be simulated using symbol systems, and symbol systems can be implemented using a connectionist architecture, but that is independent of the question of what each can do qua symbol system or connectionist network, respectively." Harnad (1990, p. 338)Fodor is well known for allowing that the symbolic language of thought (LOT) may be implemented, in the case of human brains at least, on a connectionist-style foundation.

2.4.1 Theories of Meaning

Behind the two approaches are two competing theories of meaning.[19] On the one account, "semantic facts are somehow constituted by nomic relations." One possesses the concept DOORKNOB[20] if one is a reliable tracker of doorknobs. Likewise "... the fact that DOG means *dog* (and hence the fact 'dog' does) is constituted by a nomic connection between two properties of dogs; viz. *being dogs* and *being causes of actual and possible DOG tokenings in us." Fodor (1998, p. 73, emphasis original)* This, in a nutshell, is informational semantics. Some form of direct realism is presupposed.

On the other account, semantic facts are constituted not by nomic relations with a mind-independent world but by associations between nodes in a(n) (artificial or natural) neural network at one level (in particular, "dynamic patterns of activity in a multilayered network of nodes or units with weighted positive and negative interconnections" Harnad (1990, p. 337)) and associations between experiences at another. Meaning is mind dependent. Some form of anti-realism is often taken to be implied.

The approach of this paper, as should shortly become clear, is that both theories of meaning are needed. The problem with the nomic account is not the nomic account *per se* but the assumption that it is (or even can be) a complete account, and the pitching of the nomic relations at completely the wrong cognitive level: namely, at the level of recognizable concepts as opposed to rudimentary proto-concepts.

2.4.2 The Continuity of the Conceptual and the Non-Conceptual

But the issue is more than any of that, and to some accounts at least, goes to the heart of the divide between conceptualists and non-conceptualists, which cuts across e.g. the divide between symbol-favouring cognitivists and association-favouring connectionists.

The difference between the two is not, in the main, over whether there is non-conceptual mental content: most conceptualists allow for *that*. The difference is not even, at heart, over whether there is non-conceptual content *of experience*, though the argument is often (mistakenly, Bermudez believes, and the authors are inclined to agree that he is right) framed in just that way.

The *real* question is whether conceptual mental content is continuous with non-conceptual mental content or whether they are distinctly different kinds of content: in other words, whether being conceptual or non-conceptual is just a matter of degree. Is there a distinctly conceptual level of cognition and a distinctly non-conceptual level, or are such discrete levels merely useful fictions? The enactive perspective has already led us to be skeptical of the agent/environment divide; what will it tell us about this one?

Remember that concepts are a lot like symbols; they are typically described in symbolic language, e.g. both concepts and symbols are recognizably concepts and symbols to the extent that they abstract away from the particulars of context. At the same time, concepts are also associational, certainly at the level of lexical concepts and linguistic metaphor. Non-conceptual mental content, in contrast, is, not surprisingly described

[19] A full account of meaning is, of course, beyond the remit of this paper.

[20] Note the convention here of putting concepts in all capitals, meanings in italics and words *qua* words in quotes.

in the non-symbolic language of images and associations.[21] So if conceptual mental content is continuous with non-conceptual mental content, that might also imply that symbolic language/methods/etc. are continuous with associational language/methods/etc. Symbols might also be associations or associations symbols, or the same structures might have aspects of both. Symbols might be continuous with sub-symbolic entities.

Likewise if conceptual mental content and non-conceptual mental content are distinct, then that might imply that symbolic language/methods/etc. and associational language/methods/etc. are distinct. A mental particular might be (at least in the first instance) symbolic or associational, but it could not be both. Mental particulars on one level of cognition might be symbolic, on another level associational, but the two levels would be distinct.[22]

2.4.3 Symbols and Associations: A Hybrid Account

How might a hybrid account work? Again, Harnad offers some thoughts:

> . . . Connectionism can be seen as a complementary component in a hybrid nonsymbolic/symbolic model of the mind, rather than a rival to purely symbolic modeling. Such a hybrid model would not have an autonomous symbolic module, however; the symbolic functions would emerge as an intrinsically dedicated symbol system as a consequence of the bottom-up grounding of categories' names in their sensory representations. Symbol manipulation would be governed not just by the arbitrary shapes of the symbol tokens, but by the nonarbitrary shapes of the icons and category invariants in which they are grounded. Harnad (1990, p. 335)

Amodal, context-free, discrete, arbitrary symbols must be grounded in something that does not meet those qualities:

> Symbolic representations must be grounded bottom-up in nonsymbolic[23] representations of two kinds: (1) iconic representations, which are analogs of the proximal sensory projections of distal objects and events, and (2) categorical representations, which are learned and innate feature-detectors that pick out the invariant features of object and event categories from their sensory projections. Harnad (1990, p. 335)

Note that Harnad's symbols are grounded in *categorical representations* (i.e., symbol-like structures) and in *iconic representations* (i.e., associational structures).

How should associations be grounded? As the topic of Harnad's paper is the symbol grounding problem, grounding associations is not so much his concern, even though his stated goal is a hybrid symbolic/connectionist account. But some connectionists, notably Barsalou but also e.g. Gallese and Lakoff (2005), have looked at how connectionist networks might be grounded in something partly symbolic, or symbol-like: i.e., localist networks, where meaning is not just distributed through connections and weights but is also partly local to individual nodes (and partly, as e.g. in the neural theory of language [NTL] model described by Vittorio Gallese and George Lakoff, in small "functional clusters" of nodes).

The problem, as always with localist networks, is to identify the appropriate level of granularity for localized meaning.[24]The temptation may be to set the granularity too high and in the interests of understandability

[21] E.g., Fodor (2006) talks of "iconic" (associational, picture-like) vs. "discursive" (symbolic) representations and says that "'iconic' and 'discursive' are mutually exclusive modes of representation; that a representation is either entails that it's not the other."

[22] Compare what Harnad says of symbolists: "According to proponents of the symbolic model of mind... symbol-strings... capture what mental phenomena such as thoughts and beliefs are. Symbolists emphasize that the symbolic level (for them, the mental level) is a natural functional level of its own, with ruleful regularities that are independent of their specific physical realizations." Harnad (1990, p. 336)

[23] Perhaps, given the symbol-like nature implied by "categorical representations", he might better say sub-symbolic.

[24] "In the caricature, each concept—say, the concept of your grandmother—is represented by one and only grandmother—is represented by one and only one neuron. If that neuron dies, then you lose the concept of your grandmother. No localist ever proposed such a theory,and nor do we." Gallese and Lakoff (2005, p. 468)

make the localized meaning recognizably conceptual, as opposed to e.g. proto-conceptual. The granularity should probably be set as low as possible, the localized meaning well below the level at which concepts begin to look recognizably like concepts.

In a hybrid symbolic/associationist account, something of the symbolic account will go *all the way down*, so that even at the most basic subpersonal cognitive level, there will be some tightly circumscribed sense in which the system can be meaningfully interpreted, by an observer, as doing symbol manipulation. Likewise something of the associationist account should go *all the way up,* so that even at the most abstract level, furthest removed from immediate perception, there will be some minimal context dependency, no matter how impoverished.

2.4.4 A Final Word on Fodor

It has arguably been some years since Fodor could claim of his representational theory of mind (RTM), on which his informational atomism theory of concepts is based, that "RTM remains the only game in town. . . ." (Fodor, 1998, p. 23) Its slogan "no cognition without representation" (Fodor, 1998, p. 26) may seem naive in light of Harvey's analysis. However much as Fodor may continue to feel otherwise, it surely cannot be the case that RTM can offer a complete theory of mind nor informational atomism, with its focus on rule-based operations over symbols, a complete account of concepts. Yet, as the authors hope the preceding discussion has shown, there is room for RTM within a larger theory of mind and for informational atomism, or something like it, within an overall account of concepts. Particularly at the level of self-conscious thought, where attention is serial and thoughts appear propositionally structured, the idea of concepts "just being" word-like symbols may find its most natural home. Of course, what makes those apparently unstructured symbols possible may be a rich wealth of highly structured and highly interconnected content at unconscious and subpersonal levels. Surface appearances may be deceiving.

3 The Mutual Scaffolding of Concepts and Experience

The possibility of innate concepts aside, there are no concepts without experience. To borrow a page from Noë, that experience must further be *sensory-motor experience*, for the agent must be cognitively and physically engaged with its environment to experience environment or self. The standard reference here, of course, is the classic study on kittens reported in Held and Hein (1963).

Can there be experience without concepts? Remember that, earlier, non-conceptual mental content was described as content "of the moment", conceptual mental content as content "not of the moment" but abstracted away from the particulars of the moment. One can seemingly conceive of an agent, without possession of or ability to employ concepts, that still has experiences. In such an agent, memory could play at most a subpersonal role; in any case, the experienced world would, in every instance, be something new. There would be no relating to the past *as* the past or to the future *as* the future, for that would imply at least some minimal conceptual abilities.

For the conceptual agent, however, such experience completely uncoloured by concepts may no longer be a possibility. Such an agent may never experience the "now" on its own; instead, the "now" is experienced in the light of past moments and in anticipation of future ones. At least in the human case, concepts reliably shape and re-shape our experience of the world. So for example, though there is no reason to think that anyone is born with an innate concept of DOORKNOB – to borrow Jerry Fodor's example – once an agent has the concept DOORKNOB then, in most instances, that agent cannot fail to see a doorknob *as* a doorknob. Not only does that agent, in Fodor's language, become a reliable doorknob tracker; she cannot step aside from that role.

To a conceptualist like John McDowell, all experience is conceptualized[25], so the possibility of an experiencing agent without concepts cannot arise; our hypothetical non-conceptual experiential agent is inconceivable.[26] Other conceptualists might say that the extent to which something is an experience is the extent to which it is conceptualised.[27]

To the non-conceptualist, on the other hand, all experience, for the conceptual agent, is some mix of the conceptual and the non-conceptual, such that fully conceptualized and fully non-conceptualized experience will be unrealized ideals. Conceptual and non-conceptual are not strictly separated (say, conceptual as experiential, non-conceptual as sub-experiential or sub-personal) but exist along a continuum. In that case Evans' Generality Constraint can be used as a metric for deciding whether, and to what extent, some particular mental content is (more or less) conceptual.

Concepts, it seems, require experience. Experience, at least for the conceptual agent, requires concepts, which structure the very experience that structures them. Concept acquisition and application go hand in hand. It's like the question of the chicken and the egg: which comes first, concepts or experience? Something, of course, must start the process off: perhaps some small set of innate concepts (one might prefer to call them proto-concepts); perhaps some minimal set of experiences. It will be the position of this paper that both are required.

Acquiring concepts is a process of applying concepts, which may themselves change in the process of acquiring the new concepts. Experience scaffolds concepts, which in turn scaffold experience.

Of course while one can talk about concepts and experience giving rise each to the other, implying one single process viewed from two perspectives of acquisition and application (like two sides of a coin), in order to get some kind of conceptual handle on matters, it helps to talk of the two perspectives as if they really are two separate processes with slightly different rules.

3.1 Experience Scaffolding Concepts: Noë's Enactive Sensory-Motor Approach

Even if concepts (or proto-concepts) and experience (or just the capability for it) are scaffolding each other from the very beginning, still, one needs to begin one's account somewhere. Since, by our present understandings of early child development, experience takes priority over concepts (indeed, many philosophers would deny any conceptual mental content to pre-linguistic infants), perhaps that account should begin with how experience scaffolds concepts[28]; and since, as we have suggested, experience is first-and-foremost sensory-motor experience, perhaps *that* account should begin with Noë. On Noë's account, sensory-motor profiles are sets of expectations about what an agent would expect to happen were that agent to engage in various sensory-motor activities. They are derived from Gibson-style affordances offered by different objects in and properties of the agent's environment: so chairs afford sitting and ladders afford climbing. "...In effect, perceiving is a kind of skillful bodily activity" Noë writes (Noë, 2004, p. 2), and the account he offers of conceptual understanding proceeds from there.

For all that Noë's account seems right in many respects, several reservations can be raised:

- Noë's enactivism is, as noted before, at one end of the enactive spectrum, and Noë specifically eschews any connection between the way he uses the word "enactive" and how others use it.[29] If the analogy of

[25]...Which is to say fully conceptualized, although McDowell, as a conceptualist, does not accept that there are degrees to conceptualization; rather, it is an all-or-nothing proposition.

[26]See, for example, his discussion of (non-human) animal consciousness in Lecture VI of (McDowell, 1996).

[27]Susan Hurley, though not a self-identifying conceptualist, might be taken to be sympathetic to this view, in e.g. (Hurley, 2003).

[28]...Which is to say, concept acquisition, which any account save a radical nativist one must provide for.

[29]Noë's enactivism is focused quite narrowly on sensory-motor contingencies and profiles, whereas other enactive accounts (e.g., Evan Thompson's) are both more widely focused (in that their attention is not limited to the sensory-motor) and more restrictive about what qualifies as enactive.

a spectrum is the appropriate one, then the present authors lie between Noë and Thompson on that spectrum.

- Noë's account is arguably forward-looking without being backward-looking: Noë has a lot to say about where expectations take us but relatively little to say about where they come from. One might contrast his account with the sort of account that Barsalou, in the spirit of William James, would like to give: one hand always in the future, one always in the past.
- This might in part be a consequence of Noë's strong externalism. In contrast the sort of enactivism endorsed by the present authors sees either internalist or externalist perspectives, at least when taken on their own, as deeply misleading. So one of the important things to remember about affordances is that they are not *in* the environment but always relative to the perspective of an agent: one person's chair may be another person's stepladder. What is a tree stump at one moment in time may be a chair at another; and so on.
- In particular, as Anthony Morse and Tom Ziemke have written in an upcoming paper(Morse and Ziemke, 2008), Noë focuses on sensory-motor contingencies to the exclusion of any consideration of the agent's bodily states in general or emotions in particular. Without an account of emotions and, consequent to that, of motivations[30], Noë has no explanation for why some affordances in the environment are salient and others not. On Morse and Ziemke's account, rather than relating sensory input just to motor output and vice versa, both sensory input and motor output relate to somatic input.
- Noë eschews representational language (though the authors share his concern about *a priori* internal representations). As should become clear, this paper takes that to be a mistake, so that representational/symbolic language is not just a useful but a necessary part of an overall account of cognition or conceptual mental content. Per (Chrisley and Parthemore, 2007), a representationalist account – of the right kind – is, *pace* Noë, compatible with an enactive sensory-motor approach.

3.2 Sensory-motor Plus

One might call a refinement of Noë's account "sensory-motor plus": sensory-motor plus somatic and other bodily information (per Morse and Ziemke), plus (with appropriate qualifications) symbolic/representational language (per Chrisley and Parthemore). How might it work?

Per associationist accounts, the story begins with pattern recognition, albeit with the caveat that there is a minimal pre-existing notion in the system of what patterns are: these are the proto-concepts, and they are (for practical purposes at least) governed by nomic relations. Regularities in the perceptual stream (between one moment and another and another) are, by some somatic-based account, recognized as salient and remembered by the agent. A minimal perceptual regularity is a mapping of one point in the perceptual stream to another.

Regularities in the regularities and regularities in *those* regularities (which could be represented as associations between first lower- then higher-level maps, with respect to time on different scales) yield increasingly complex, more abstract, higher-order concepts. Concept formation, from this view, is an abstracting away from particular moments, from particular contexts. If this model of concept acquisition is anything close to being right, then as one ascends through the associational hierarchy – from associations between perceptions to associations between associations, associations between (associations between associations), and so on – the richness (dimensionality) of the referring structures will be reduced at each step, and the richness (dimensionality) of the referent structures (the target descriptive space) will be likewise increased, until the referring structures come increasingly to look like unstructured symbols, whose both sign and semantics bear no obvious relation to any particular context. The referring structures become mere pointers, but to richly structured descriptive spaces. If one inverts the associational hierarchy, then what at first looked very much like symbols will gradually lose themselves in context as their meaning becomes more and more defined by context, until most (if not all) of what we understand by symbols disappears.

[30] ...Though some authors, notably Aaron Sloman [ref.], have argued that emotions and motivations pull apart, and that motivations need have nothing to do with emotions.

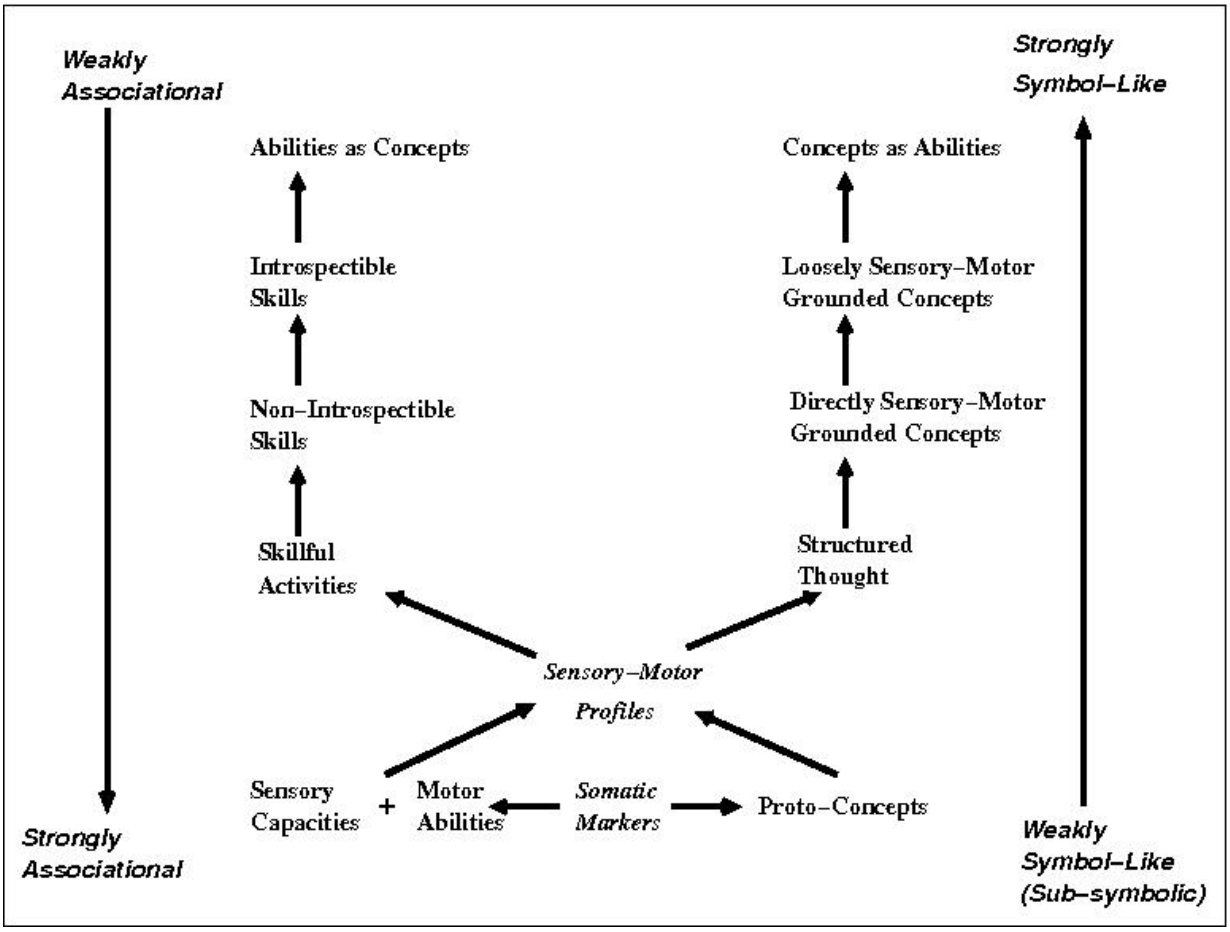

Figure 2: Extending the Sensory-Motor Model

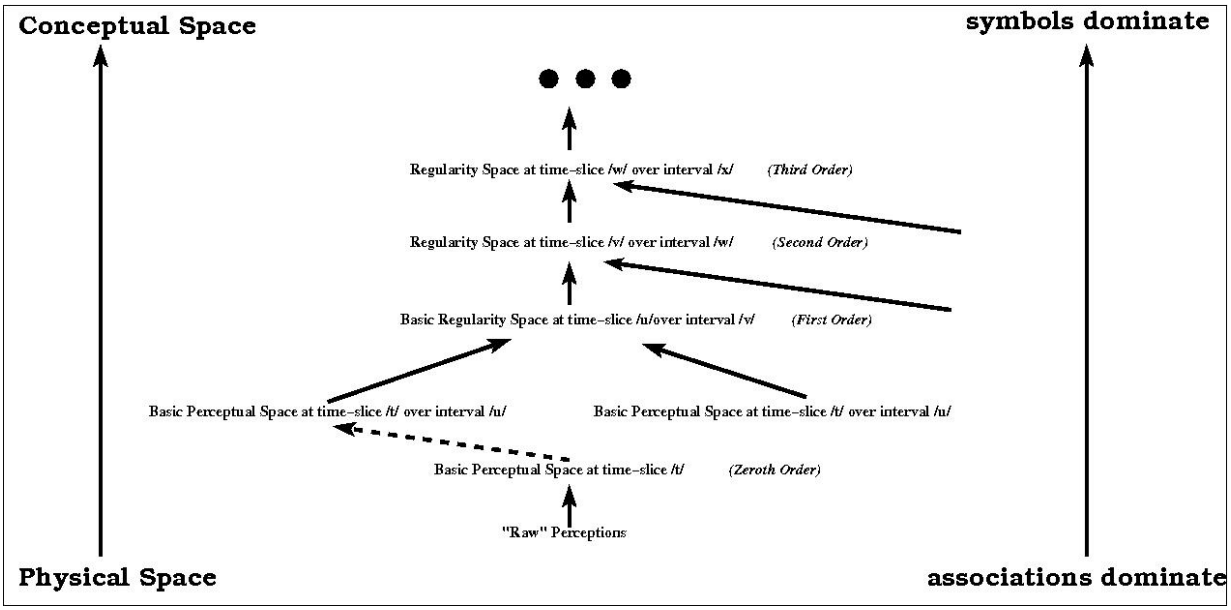

Figure 3: Concept Acquisition: The Associational Story

A concept then is or could be described as a synchronized pattern of higher-order[31] association between some aspect of the mental world of the agent and some matching aspect of her experienced environment. It is both abstracted away from and structurally isomorphic to its referent in perception.

If this does not sound much different from traditional concept empiricist approaches, or modern ones (e.g., Prinz's proxytypes account), it's not. Enactive accounts, which ground concepts and cognition in sensory-motor engagement, owe much to concept empiricist accounts, which ground concepts in percepts.[32] It's also not so different, in a way, from classical definitionist accounts, which likewise tended to start from concept empiricism, using perceptual primitives to build concepts, which could then be used to build more concepts, which could in turn be used to build yet more (more complex, more abstract, etc.) concepts.

Enactive accounts of concepts, however, go beyond concept empiricist accounts: not only are concepts not static entities, they are not, ultimately, to be found in the agent (nor in the environment) but rather dynamically enacted out of the interaction of the agent with her environment.

3.2.1 How It Could Work, Associationally

Given some set of "raw" perceptions – representationally, a set of (subpersonal) perceptual spaces – indexed by the moment of perception, at interval t, over some period of duration u, one can derive some minimal regularities from them: say, to borrow an example from robotic vision, recurring pixels at the same or similar locations, or sudden localized changes in pixels. Those regularities describe a space of their own, but it's no longer a perceptual space, strictly; rather it's one step removed: a space of regularities, albeit very, very basic ones. Call it a regularity space.

Given some other set of perceptions indexed by the moment of perception at the same interval t, over some period also of duration u, one can derive some regularities from them as well. They, also, will describe a regularity space.

[31] ... In the sense of recognizably being a concept to the extent that it is abstracted away from the particulars of context.

[32] Indeed, for all of the resemblance to concept empiricist accounts of concept grounding in Gallese and Lakoff (2005), it is perhaps surprising that they do not note the relationship.

Given these two "first-order" spaces of regularities, one can then compare them and others like them just as the sets of perceptual spaces were compared. But t (as the minimal, individuable unit of time) is no longer significant; rather u takes its place. And one is no longer comparing perceptual spaces but regularity spaces. A new value v needs to be introduced, which takes the place that u had before, as the duration value.

Likewise one could compare "second-order" spaces of regularities to derive "third-order" spaces, and so on through to nth order spaces, limited only by practical boundaries (e.g., available time, energy). As one goes along, two things happen: abstract concepts (or general principles, to look at it another way) are derived from specific instances/encounters, and parts (simple concepts) come together into wholes (complex, composite concepts). In similar fashion general abilities are derived from more specific ones, and simple abilities come together into complex, composite abilities. At some point one gets line and blob detection[33], at some other point object recognition and persistence. This description implies a hierarchy. But whether one thinks of these as discrete stages or as a continuum might simply be a matter of which perspective is most useful.

3.2.2 How It Could Work, Symbolically

Given some small set of sub-symbolic proto-concepts – and it could be quite small, and quite basic, limited to, say, a proto-concept of object, a proto-concept of action or event, and a proto-concept of property – one can assemble them building-block style into more finely delineated proto-concepts, then into simple and highly context-dependent "low-level" concepts, and eventually into complex and highly context-free "high-level" concepts. These conceptual building blocks are different from their children's toy counterpart because the blocks do not just fit together locally (one block on top of another or next to another) but distally (e.g., *this* block here is actually the same as *that* block there).

The smaller the set of initial building blocks, the more uniformly structured the resulting complex structures will be; and, as Prinz has noted[34], a uniformly structured theory of concepts is, all other things equal, to be preferred. Likewise the more basic the initial blocks and the fewer the assumptions that are built into them, the wider the range of complex structures that can subsequently be built. Ideally anywhere one examines the conceptual network, the building blocks and the rules (for composing or decomposing them) should be the same.[35] If concepts are similarly structured at *any* level, then one has *scale invariance.*

Applied to the regularity spaces they are derived from and deeply intertwined with, these proto-concepts and concepts can be used at many different levels of cognition to ask questions [36] like:

1. What is here?
2. Is *this* here?
3. What if *this* were here?

3.3 Concepts Scaffolding Experience

Of course if the account given so far, of concept acquisition, was *all* there was to be said, then concepts really would be static entities, with no means for update or obsolescence. But concepts are not some exotic species of things to be collected, like a lepidopterist's butterfly collection. They have no meaning unless at the same time they are being acquired they are being applied. Concepts are at least as much skillful abilities as they are expressible knowledge.

[33] . . . Which could, also, simply be hardwired in by evolution.

[34] "If concepts are structurally uniform (or uniformly unstructured), a uniform theory of concepts is easier to achieve." (Prinz, 2004, p. 94)

[35] The rules could go something like this: composing upward, anywhere you have a block like *this* and a block like *that*, you can join them together them *so*. Composing downward (or decomposing), anywhere you have a single block, you can replace it with a structured set of blocks like *so*.

[36] At subpersonal and at personal but unconscious levels of cognition, the question asking will be strictly metaphorical.

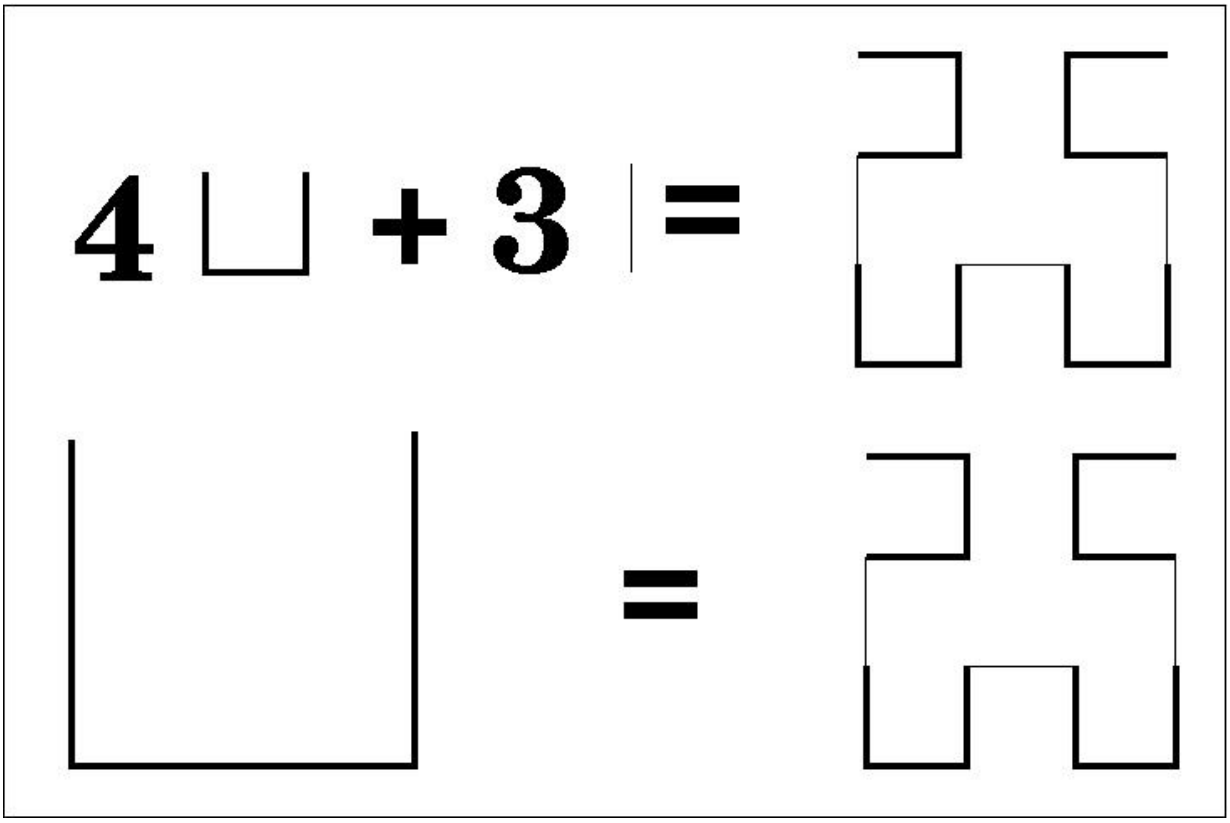

Figure 4: Consider the U shapes as concepts (the building blocks) and the | shapes as non-conceptual "glue" (the rules for assembling them). Given four U shapes and three | shapes, one can assemble a new "U" shape (upward composition). Likewise, going the other direction, given a U shape, one can replace it with four smaller U shapes and three | shapes (downward composition, or decomposition).

To borrow a page from the classical definitionists, the concept acquisition model just outlined can be turned on its head, verifying instead of discovering, disassembling instead of assembling, in the same spirit in which definitions are neutral as to whether they are defining new concepts or identifying and verifying old ones. Before concepts were being abstracted away from experience, away from the particulars of the moment. Here concepts are being applied *back* to experience, back to the particulars of the moment.

3.3.1 Concepts as Expectations

With one hand in the past and the future, concepts are the expectations that drive experience.[37] Non-conceptual mental content, as it has been defined here, can in comparison offer but the most impoverished notion of expectations, being limited to treating the content of experience at some other moment *as* the content of experience in this moment.[38] Even to distinguish one moment from another moment might imply some minimal conceptual abilities, for one is abstracting away from the content of any particular moment. It has been suggested earlier that, for the conceptual agent, all experience is a mix of the non-conceptual and the conceptual, so that both fully non-conceptual content of experience and fully conceptual content of experience are unrealizable idealizations.

Consider concepts as a tool that, once you have it, you literally cannot imagine doing without. Perhaps concepts are like language in this way. There was a time, for each of us, when we were a pre-linguistic infant. Likewise there was probably a time for the human race when humans, as a species, were pre-linguistic.[39] But once we have become language-using agents, we cannot, bar some catastrophe, *stop* being language-using agents, even in our private thoughts, which often seem to be structured in words, *sotto voce*.[40]

Consider conceptualised experience as a projection over top of non-conceptualised experience, all but obscuring it. Once we become aware of past and future *as* past and future, we cannot help experiencing the present moment in light of both. In Damasio's language, we begin telling the narrative that gives us our rich sense of autobiographical self.

If concepts are a tool, then perhaps the metaphor is Heidegger's hammer. Only when the hammer breaks or the nail bends – that is, only when the hammer fails, in some manner, to perform as a hammer – do we stop and see the hammer *as* a hammer.We see, hear, and feel what we expect to until the match between expectations and current experience breaks down in a manner that we cannot ignore and we are forced to take a closer look, at which time our implicit conceptual expectations are made explicit.

3.3.2 Representation as Control

We exist, by most accounts, within a wealth of perceptual information. A world that can conceptually be carved up in infinitely many ways – regardless of whether nature does, or does not, have joints to be cut along – offers an uncountable number of possible affordances. What determines salience? Proprioceptive experience, somatic markers and motivations tell part of the story, perhaps the entire story for agents without concepts. But for agents with concepts, conceptual knowledge also plays a critical role.[41] Conceptual expec-

[37] The treatment of expectations here is complementary to Ron Chrisley's Expectation-Based Architecture (EBA), described e.g. in (Chrisley and Parthemore, 2007).

[38] This is the sense in which the SEER-3 robot, at its current state of development, could be taken to have expectations.

[39] This is controversial, however. By some accounts (e.g. [ref]), hominids are not properly classifiable as *homo sapiens* until they are language users. Thanks to Blay Whitby for pointing this out.

[40] Even stroke victims, who may temporarily or permanently lose e.g. the power of speech, do not normally lose all their language abilities: ability loss can be highly selective.

[41] Note that, although there will be agents that clearly lack concepts, insofar as there are no predictive advantages to attributing them, and agents that clearly possess concepts, on the same grounds, there need be no sharp cut off between agents lacking concepts altogether and agents possessing them. Earlier it was said that, on a non-conceptualist account, there is no sharp cut off between conceptual mental content and non-conceptual mental content. This can be taken to imply that, likewise, general concept possession exists along a continuum, and the cut off between an agent lacking concepts and an agent possessing them is a pragmatic one. The archetypal agent lacking concepts will, to the extent that it has experience at all, have the kind of experience that is typically described in the literature as "living solely in the moment." Where does concept attribution become useful? ... Arguably, where a strictly stimulus-response-based model becomes unworkably awkward to maintain: roughly, when the same agent, presented with the same stimulus in the same circumstances, appears to make flexible responses – choices –

tations based on past experience guide and massively simplify the agent's interaction with its environment; they supplement and extend the possibilities offered by non-conceptual expectations. This is very close to Imogen Dickie's (Dickie, 2006) notion of "representation as control".

Not surprisingly, conceptual expectations present several trade-offs. Simplification is necessarily, to some extent, falsification: a useful lie, as it were. The greater the simplification, the greater the remove will be from the initial truth.

Attention to certain details necessarily entails inability to attend to others. Ask the people watching a basketball game to count, and report, the number of times the ball bounces, and they will consistently fail to see the gorilla walking across the court: inattentional blindness. A control group with no such instructions will be far likelier to see the gorilla. (Simons and Chabris, 1999) Martin Langham reported on people who pull out in front of motorbikes, who "look but fail to see." What he found was that inexperienced drivers look all over the place. Experienced drivers minimize where they are looking. [ref]

But beyond all of this, the more conceptual knowledge we have, the more we come to rely on it. As a wealth of psychological evidence shows, most of the time we see, as it were, not what is in front of us but what we *expect* to see in front of us. The very expectations that simplify our interactions with our world may make us less adaptable when circumstances change, less able to learn when new information presents itself.

3.3.3 The Dynamic Dictionary

Dickie's "representation as control" is based on those aspects she wants to hold onto from Wittgenstein's picture theory. No one, including the later Wittgenstein, thinks that the picture theory as originally presented can work. What are the alternatives?

For the classical definitionists, the control mechanism was the definition, which specified the necessary and sufficient conditions for a concept's application. Again, no one thinks that the classical definitionist account can work, either.

The number-one most frequently stated reason for classical definitionism's abandonment is, as Fodor has put it, "there are practically no defensible examples of definitions. . . ." (Fodor, 1998, p. 45) One might think him generous for allowing that there are any. Once the requirement is dropped that concepts must be static entities, though, the idea of concepts as definitions may have some more mileage after all. Drop the requirement as well that concepts be strictly or primarily public entities – that is, allow for concepts to have both a significant public and a significant private aspect (a distinction that e.g. (Woodfield, 1994) thinks is important to make).

Consider concepts in the context of language, where (lexical) concepts map roughly onto the words of a language. What would a concept-as-enactive-definition look like? Again, it would belong not strictly to the agent (i.e., internal) nor to the agent's (in this case social) environment. Nonetheless, it would be convenient to see concepts-as-definitions *now* as private entities (your concept DOG, my concept DOG, both shaped by our idiosyncratic histories of encounters with dogs and stories of dogs and so on), *now* as public ones (the concept DOG signified by the English word "dog"). To borrow Frege's distinction, the reference will, most of the time at least, be the same, but the sense will often be different.

The best metaphor here might be a kind of "dynamic dictionary", where the words on the page are constantly in motion: look at a definition, look away, look again, and the definition has subtly changed. Offering a definition in the more usual sense of the word becomes an attempt to fix the concept, to take a snapshot. Something of what the concept is is captured, but something more – in particular the motion – is lost.

at different times, based on its history. Note that learning on its own is not enough, as the learning could be genetically pre-programmed and so quite rigidly bounded, in which case a strictly stimulus-response-based model can be maintained. Of course it's difficult to say what qualifies as the same stimulus, and one could give a strictly stimulus-response-based interpretation to some quite complex behaviour. But you probably wouldn't.

To return to and refine our earlier working definition, a concept, then, is or could be described as a synchronized pattern of higher-order association between some aspect of the mental world (or the experienced world) of the agent and some matching affordance(s) of her environment, that implicitly or explicitly specifies the necessary, sufficient, and customary (or contextual) conditions for its own application *relative to any particular moment.*

3.3.4 How It Could Work: Toward a Formal Model

Concept acquisition was described as a bottom-up, layer-by-layer hierarchical process of pattern recognition, building up more complex concepts from simpler concepts, more abstract concepts (less directly connected back to sensory-motor engagement) from more concrete ones (more directly connected back to sensory-motor engagement), more specific categories from more general ones, more generally applicable categories from more narrowly applicably ones. Concept application, then, is a top-down, layer by layer process of pattern matching, working the other direction from the complex toward the simple, the abstract toward the concrete, the more specific categories toward the more general, more generally applicable categories toward the more narrowly applicable. If concept acquisition is a layering process, then concept application is delayering. The basic idea is this: the concepts obtained from concept acquisition can be matched against their non-conceptual analogs in present experience, layer by layer, matching features of the one with the features of the other, until a match has been confirmed or a breakdown occurs, forcing a closer look.

Given a continuum from specific contexts and moments to generalized across contexts and over long periods of time, and given a continuum from homogeneous localized parts to heterogeneous conglomerated wholes, and to the extent (based on the earlier discussion) that those continuums can be understood in terms of discrete hierarchy, then concept application can be seen as a return down through levels of the hierarchy, toward particular encounters and toward parts as opposed to wholes. If an X violates expectations, consider previous experiences with similar Xs, or decompose the X into e.g. its functional parts.

Take a door that is in front of you. Does your present experience of that door (*as* a door) match your expectations at the most abstract conceptual levels you can apply? If you don't need to see the door as anything more than a whole with no parts (like an "unstructured" symbol), then you won't. It will register as an undifferentiated door. Of course, depending on where your attention is focused, you choose or be motivated to look more closely. Where is the handle, where are the hinges? Does the door open outward or inward? How does this particular door relate to previous doors you have encountered? The more closely you examine the door, the more directly your sensory-motor capacities with respect to that or other doors will be brought to bear, on-line or off-line.

If you need to pass through the door, you will look, minimally, for how the door opens. If it has a handle, you'll probably be inclined to pull it. If it has a flat metal plate where the handle would be, you'll be inclined to push it.

Only if the door has something perceptually un-door-like about it will you be forced to examine it more closely, e.g., if the door has a handle but is meant to be pushed instead of pulled, in which case you might look for clues such as details of the door frame. One could imagine that the "door" is only a painting on the wall of a door, or has been painted or nailed shut. Unusual doors will focus your attention and shift it from the abstract and general to the concrete and immediate, from doors as some platonic-like entities to specific door encounters.

For concept acquisition, associations and association-building were in the driver's seat. For concept application, initially at least, symbols and symbol application may be a more appropriate level of description (though only some small part of this may be consciously articulable). Of course, at some point the unusual door in front of you may confound all attempts at conceptual understanding, and you may resort to brute sensory-motor engagement with it.

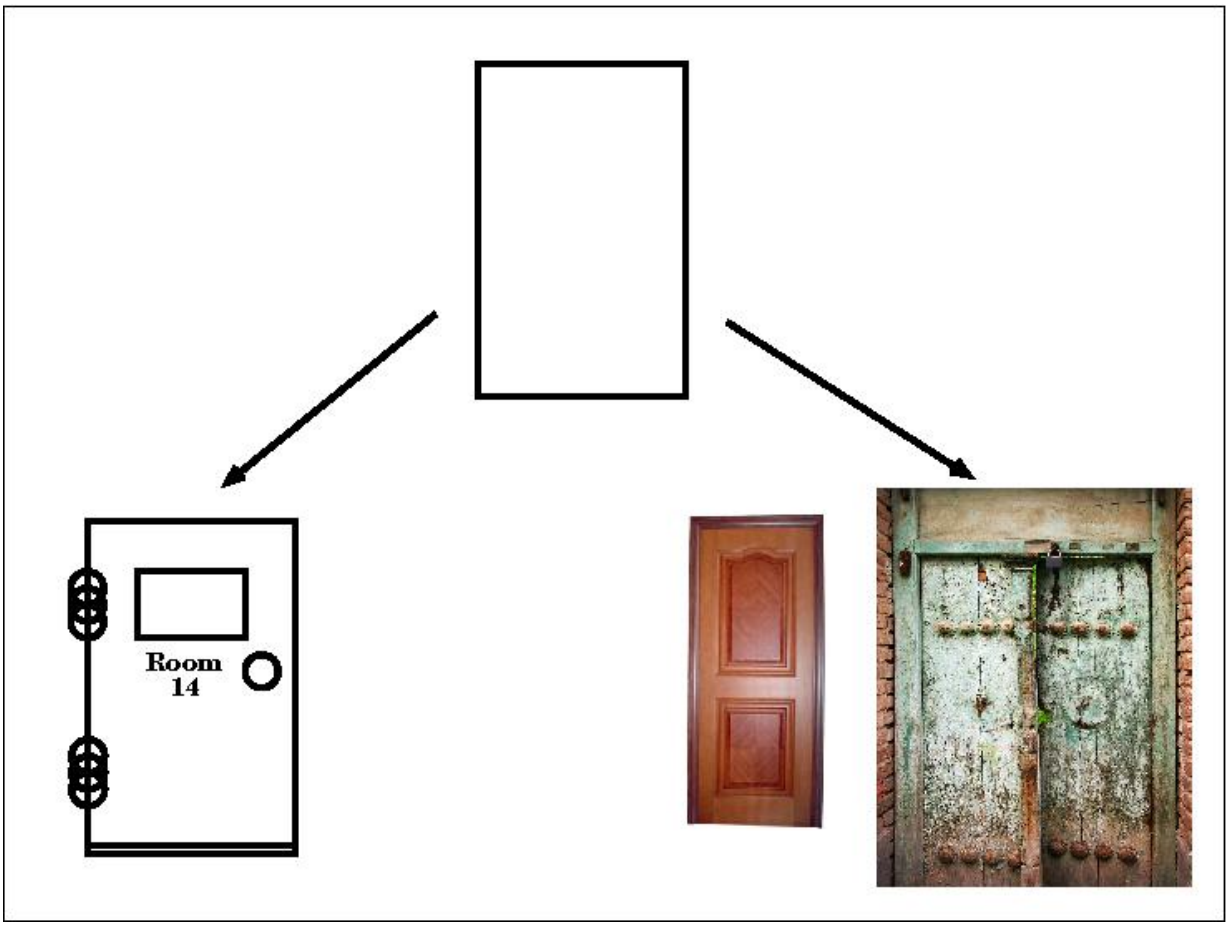

Figure 5: Given a door, one can see it as an undifferentiated whole. Or, one can focus attention on its (e.g., functional) parts, or on previous door encounters.

In practice the pattern matching can only go so far. As previously stated, one's concepts will be more *unlike* their referents structurally than they are like them. At some point sooner than later, conceptual understandings break down. Also one cannot get more specific than calling up specific relevant experiences and individually matching against each of them. There will furthermore be practical considerations like a balancing of cost (e.g. time and effort required) with return (presumed degree of certainty). The goal is sufficient correspondence, as determined by the agent's current needs in its present environment.[42] The moral: look as close as you need to, and no closer![43]

When breakdown occurs – when the thing you took to be a door is not a door after all – some amount of re-conceptualization takes place. The simplest account would be that re-conceptualization begins from the first layer in the conceptual hierarchy at which the breakdown can be accommodated without further de-layering. That is to say, when re-conceptualisation is necessary, the minimum re-conceptualization needed to bring conceptual expectations and experience back into alignment should be undertaken. If you previously understood all swans to be white but now see a black one, you should only adjust your most general understanding of swan-ness and not, e.g., revise your understanding of previous encounters with swans. (You thought they were white but you were mistaken, the victim of some illusion.) You should likewise

[42] What counts as sufficient? Organism survival is a good starting point. Even were it possible for an agent to bring *all* of its conceptual capacities to bear on understanding *every* aspect of its environment, such an approach would probably not be conducive to its survival! Then there are considerations like maximizing safety and comfort, minimizing risk, and stress.

[43] This is related to Andy Clark's 007 principle: "know only as much as you need to know to get the job done." (Clark, 1989, p. 64)

revise your understanding of swans and not that of birds more generally.

4 Applying the Model

Given some artificial agent – say, the off-the-shelf robotic dog used for the SEER-3 project – one could imagine exploring the acquisition/application process in a number of ways, to focus on the acquisition process, or the application process, or low-level cognition (close to the purely sensory-motor level), or high-level cognition (abstracted furthest from the sensory-motor level). Looking at any one of these in isolation can be very misleading, as this paper has attempted to suggest, e.g. inclining one to believe that low-level cognition is only about sensory-motor engagement (at least, as narrowly interpreted) or that high-level symbol-manipulating cognition is functionally independent of the implementation level (a mistake that, arguably, both symbolic AI and *nouveau* connectionist AI researchers have made). At the same time, trying to capture the entire process at once is unworkable; one cannot go directly to one's destination but must rather make a journey there in stages. So one might want to look at some aspects of the process, but in a way that acknowledges and does justice to the rest of the process.

Concept acquisition and application have been discussed so far as if they were mostly if not entirely unconscious and automatic: outside intentional control. But, at least for the agent capable of self-reflection, there is the possibly that the acquisition/application process can be partly (self-)conscious and manual: explicit, intentional learning as problem solving. Given *these* concepts, what new concepts can you build with them (induction), and what properties can you derive from them (deduction)? Given *those* concepts, how could you have arrived at them (abduction)?

The SEER-3 project might usefully be extended in two ways:

1. To show how non-conceptual expectations, as modeled in the current SEER-3, might be supplemented and extended by conceptual expectations, and how Ron Chrisley's Expectation-Based Architecture could consequently be extended.

2. To pull out the conscious and manual aspects of the acquisition/application process, with the goal of showing how a theory informed by the largely unconscious and automatic process as described in this paper would lead to those conscious and manual aspects being modeled in a different way, resulting in different predictions. So for example, one would predict that the conscious and manual process should never be entirely under the agent's conscious control. The agent should be surprised sometimes at the consequences of her apparently straightforward choices.

Were one to locate the conscious and manual aspects within the artificial agent, that probably *would* require modeling the entire cognitive system, including the entire concept acquisition/application process. But one could also imagine offloading them onto an external agent, already known to be a self-conscious conceptual agent: the human researchers, or some human test subjects, or both. The entire conceptual system then would be the coupled system of self-conscious conceptual agent and non-self-conscious non-conceptually-experiencing SEER-3 (itself a coupled system of robot and wireless laptop control).

The application could be a significant extension of traditional mind-mapping software. Mind-mapping software, as currently conceived, allows users to externalize portions of their conceptual domain by drawing a map of their ideas on a computer screen and creating links between them. Like hypertext, it provides a way to structure information non-linearly, and its use is favoured for assisting with certain forms of learning disability. Some users find it a great benefit toward brainstorming and thus toward assisting with e.g. certain early stages of the writing process. No particular theory of concepts drives the mind-mapping software; no restrictions are placed on the links that can be drawn between nodes, or on the nodes themselves, or on how they may be interpreted.

The application would be an improvement over mind-mapping software in two important ways:

1. It would have a rich theory of concepts to draw upon.
2. By making use of a robot and the robot's interactions with its environment, the application would be embedded and embodied in a way that traditional mind-mapping software is not.

By being embodied in a particular way (a robotic dog) and by being embedded in a particular real-world environment, the conceptual mental content of the coupled human-SEER-3 system could be tested in the most demanding yet at the same time natural way possible. Of course one could imagine using a simulated robot in place of the real robot; but the real robot has a number of advantages over the simulated one, not least the rich ways in which people can interact with it. Using a robot that can break down (or whose motors can overheat, making it temporarily unresponsive) seems more true to the spirit of the enactive approach.

How might it work? Given some test subjects, their goal could be to define explicitly, for the SEER-3, some network of related concepts relating to its immediate environment, such that possessing and applying those concepts might be expected to encourage certain behaviors (chasing a pink ball) or inhibit others. The subjects would, at the same time, be externalising some portion of their own conceptual domains and defining a limited conceptual domain for the SEER-3. Of course there would need to be a fairly rich body of effectively hard-wired given primitives for the subjects to work with. But that need not count against the approach: after all, it's an open question how many hard-wired primitives humans might have, merely perhaps for sake of cognitive efficiency.

Because all of the concepts should be interconnected (locally and distally) and hence inter-defining, the subjects might find that, in the process of defining "chase", they have to define the object of the chase (say, a pink ball); in the process of defining the object of the chase, they have to define the colour of the object; in defining the colour of the object, they have to define other objects in the environment that share that colour; in defining those objects, they have to define actions that can be performed on those objects; and so on. This being a dynamic system, all those definitions should be subject to change, as the SEER-3 interacts with its environment and as the subjects see how their definitions are and aren't working out in the ways that they intended. When breakdown occurs, subjects would need to revise their definitions to correct the robot's behaviour.

5 Conclusions

This paper attempts to set out a simple model for the mutual scaffolding of concepts and experience, within the context of developing an enactive theory of concepts, one of whose basic premises should be that concepts are to be found not internal to the agent nor external to the agent in that agent's environment, but rather between the two: concepts are *enacted* out of the agent's dynamic engagement with her environment.

In order to set the stage for the discussion of concept acquisition (experience scaffolding concepts) and concept application (concepts scaffolding experience), the paper set out to explain roughly what the authors take concepts, experience, symbols, and representations to be.

Concepts are, to borrow a phrase, "persistent mental particulars"; at the same time, they are skillful abilities. The one thing everyone (or nearly everyone) can agree on is that they must meet something like Evans' Generality Constraint. To do that, the authors contend, they must abstract away from the perceptions of any particular moment to show what diverse moments and contexts have in common. Paradigmatic conceptual mental content has, as it were, one hand in the past and the other in the future.

Experience is about how the world is presented to the agent. Conceptualists and non-conceptualists argue over whether experience is or is not fully conceptualised; but what both sides seem to concede is that understanding concepts is an important piece of the puzzle to understanding experience. The authors suggest that, using Evans' Generality Constraint as a metric, for the conceptual agent *all* experiences have aspects of the conceptual and the non-conceptual.

Symbols as traditionally understood are problematic to any account of cognition. But symbols as traditionally understood are, it is argued, unrealizable idealizations. Following Wittgenstein's account, symbols are sign-meaning dyads, where the form of the sign is, to most practical purposes, arbitrary relative to the meaning. This allows a contrast with representations, where something can be seen to be a good representation to the extent that the representation can be seen to resemble the represented. This need not mean that resemblance yields representation; rather, representation yields resemblance.

An enactive approach sees symbolic and associationist accounts not as opposed but as complementary, both required within an overall account of cognition in general or conceptual mental content in particular. Gärdenfors has offered his conceptual spaces theory of concepts as a way of bridging the apparent gap between associationist and symbolic accounts – say, something like Gallese and Lakoff's account on the one hand, and something like Fodor's on the other – and showing how both are necessary to an overall account.

That experience and concepts mutually scaffold each other is easy to say but difficult to understand. Though ultimately the two scaffolding processes of acquisition and application should probably be seen as part of one unified process, nonetheless conceptually it may be useful, or even necessary, to treat them as two interacting but independent processes instead.

Any embodied theory of concept acquisition is likely to need to start from a discussion of sensory-motor engagement. But any account that talks about sensory-motor engagement without, at the same time, talking about bodily states risks being overly externalist, as well as failing to give a proper account for where our non-conceptual and conceptual expectations come from, as well as failing to give any account of salience. All these issues are related. It is also not *necessary,* when starting from a sensory-motor-based account, to eschew symbolic and representational language, although many enactive theorists do, perhaps out of justifiable fear for how those terms have been abused. Considering how Noë's sensory-motor-grounded account might be extended (see 3.1 leads naturally to a discussion about how enactive theories of concepts relate both to traditional concept empiricist accounts and definitionist accounts (which most often were grounded in concept empiricism).

As experience scaffolds concepts in concept acquisition, so concepts scaffold experience in concept application. From Imogen Dickie the authors borrow the notion of "representation as control". Conceptual expectations offer a number of trade-offs: giving with one hand and taking with the other.

In concluding, the paper briefly considers how limited aspects of the concept acquisition/application process could usefully be explored in a simple embodied mind-mapping-type application using an off-the-shelf robotic dog. By offloading the conscious and manual aspects of the process onto human agents, the problem area can be simplified while keeping the application richly embodied. The overall conceptual system becomes the coupled system of human agent and robot.

References

Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology*, 59:617–645.

Barsalou, L. W., Breazeal, C., and Smith, L. B. (2007). Cognition as coordinated non-cognition. *Cognitive Processing*, 8(2):79–91.

Barsalou, L. W., Simmons, W. K., Barbey, A. K., and Wilson, C. D. (2003). Grounding conceptual knowledge in modality-specific systems. *Trends in Cognitive Science*, 7(2):84–91.

Bermudez, J. L. (2007). What is at stake in the debate on nonconceptual content? *Philosophical Perspectives*, 21:55–72.

Chrisley, R. and Parthemore, J. (2007). Synthetic phenomenology: Exploiting embodiment to specify the non-conceptual content of visual experience. *Journal of Consciousness Studies*, 14(7):44–58.

Clark, A. (1989). *Microcognition*. MIT Press.

Cussins, A. (1990). The connectionist construction of concepts. In Boden, M., editor, *The Philosophy of Artificial Intelligence*, chapter 15, pages 368–440. Oxford University Press.

Dickie, I. (2006). Knowing-which without knowing-that. Unpublished paper presented to the University of Sussex Philosophy Society.

Evans, G. (1982). *Varieties of Reference*. Clarendon Press.

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

Fodor, J. (2006). Revenge of the given: Mental representation without conceptualization. Royal Institute of Philosophy Annual Lecture.

Gallese, V. and Lakoff, G. (2005). The brain's concepts: The role of the sensory-motor system in conceptual knowledge. *Cognitive Neuropsychology*, 22(3-4):455–479.

Gärdenfors, P. (2004). *Conceptual Spaces: The Geometry of Thought*. Bradford Books.

Goodman, N. (1976). *Languages of Art: An Approach to a Theory of Symbols*. Hackett Publishing Company, Cambridge.

Harnad, S. (1990). The symbol grounding problem. *Physica D: Nonlinear Phenomena*, 42(3):335–346.

Harvey, I. (1992). Untimed and misrepresented: Connectionism and the computer metaphor. CSRP 245.

Held, R. and Hein, A. (1963). Movement-produced stimulation in the development of visually guided behavior. *Journal of Comparative and Physiological Psychology*, 56(5):872–876.

Hurley, S. (2003). Animal action in the space of reasons. *Mind and Language*, 18(3):231–257.

Jolley, K. D. (2007). *The Concept 'Horse' Paradox and Wittgensteinian Conceptual Investigations: A Prolegomenon to Philosophical Investigations*. Ashgate.

Margolis, E. and Laurence, S., editors (1999). *Concepts: Core Readings*. MIT Press, Cambridge, Massachusetts.

McDowell, J. (1996). *Mind and World*. Harvard University Press, Cambridge, Massachusetts.

Morse, A. and Ziemke, T. (2008). The somatic sensory hypothesis.

Newen, A. and Bartels, A. (2007). Animal minds and the possession of concepts. *Philosophical Psychology*, 20(3):283–308.

Noë, A. (2004). *Action in Perception*. MIT Press.

Prinz, J. (2004). *Furnishing the Mind: Concepts and Their Perceptual Basis*. MIT Press.

Simons, D. J. and Chabris, C. F. (1999). Gorillas in our midst: Sustained inattentional blindness for dynamic events. *Perception*, 28:1059–1074.

Thompson, E. (2007). *Mind in Life Biology, Phenomenology and the Sciences of Mind*. Harvard University Press.

Winograd, T. and Flores, C. (1986). *Understanding Computers and Cognition: A New Foundation for Design*. Intellect.

Wittgenstein, L. (2001). *Philosophical Investigations*. Blackwell.

Woodfield, A. (1994). Do your concepts develop? In Hookway, C. and Peterson, D., editors, *Philosophy and Cognitive Science*, pages 41–67. Cambridge University Press.