Representations reclaimed

Accounting for the co-emergence of concepts and experience

Joel Parthemore and Anthony F. Morse
University of Sussex / University of Plymouth

Understanding the relationship between concepts and experience seems necessary to specifying the content of experience, yet current theories of concepts do not seem up to the job. With Peter Gärdenfors's conceptual spaces theory as a foundation and with enactivist philosophy as inspiration, we present a proposed extension to conceptual spaces and use it to outline a model of the emergence of concepts and experience. We conclude that neither is ultimately primary but each gives rise to the other, i.e., that they co-emerge. Such a model can then serve as the anchor to a theory of concepts more generally. Concepts are most naturally understood in symbolic and representational terms, while much of experience, in contrast, is non-symbolic and non-representational; yet the conflict between the two will, herein, be shown to be more apparent than real. The main contribution of this paper is to argue for, by means of this account of co-emergence, a continuum between "low-level" mental content that is more appropriately understood in highly context-sensitive and directly sensorimotor-based terms, and "high-level" mental content that is more appropriately understood in context-free and representational or symbolic terms. In doing so we conclude that the extreme positions of representationalism and anti-representationalism are fatally flawed.

Keywords: concept, conceptual spaces, enaction, mental representation, representation, sensorimotor, sensorimotor profile, symbol

1. Introduction

Humans are paradigmatic producers and consumers of representations. Representations and symbols are ubiquitous in everyday life: when we talk, when we read, when we sit at our computers, and indeed, as we go about most of our daily routines. They are with us in our thoughts. Cognitive science must account for them

Pragmatics & Cognition 18:2 (2010), **273–312**. DOI 10.1075/p&c.18.2.03par
ISSN 0929–0907 / E-ISSN 1569–9943

if it is to lay any claim to being a science of the mind; and yet how to do so lies at the heart of a longstanding, bitter, and so far unresolved debate. The debate takes many forms and is couched in various terms, pitting representationalists against anti-representationalists, cognitivists against connectionists, symbolists against associationists, rationalists against empiricists, and so on (cf. Brooks 1991a, 1991b; Chalmers 1990; Fodor and Pylyshyn 1988; Perry 1986).

It is necessary at the outset to say something about how we intend to use terms. Although these will be defined more precisely in due course, for now suffice it to say that "symbol" will be taken to mean a sign/signified dyad *per* Wittgenstein (Wittgenstein 2001), where the relationship between sign and signified appears to be arbitrary. (*Apparent* arbitrariness need mean nothing more than that, as we shall see.) Consider this definition from Robert Rupert: "A symbol in a model is arbitrary if there is no obvious relation between the mark or sound we use to designate that symbol and the things represented by implementations of that symbol (or realizers of it, or objects onto which that symbol is mapped during modeling, etc.)" (Rupert 2009: 221).

"Representation" will be understood in a similar way, except that the relationship between sign and signified need *not* be seen to be arbitrary. This raises a standard distinction between *iconic representations*, where the relationship is meant to be non-arbitrary; and *symbolic representations*, where the relationship *is* meant to be arbitrary. We will take "symbolic representation" to mean the same thing as "symbol". With symbols or representations, the relationship between sign and signified is normally one of significant simplification of the former over the latter, where the former may be said to stand in the place of the latter.

If humans are paradigmatically representers, then concepts are paradigmatically representational. (Note that some — notably Machery 2009 and even more so Harnad 2009 — take the concept of concept to border on the uselessly vague. Needless to say, we will not ourselves take that position!). A concept may be understood here to mean a structured unit of thought that conforms to Gareth Evans's *Generality Constraint* (Evans 1982: 100–105): it can be used *systematically* across many contexts of application and, together with a finite set of other concepts, it can be used *productively* to generate an unbounded set of complex concepts and propositionally structured thoughts. Furthermore, concepts are, like other representations, discrete (they can be distinguished from one another) and generally if not always simplified from what they are representing.

Experience, by contrast, is often understood in non-representational terms. Much of it is, or seems to be at any rate, unstructured. It appears to be continuous as opposed to discrete. Intuitively, to many if not most people it does not 'signify' — at least not normally — but rather presents the world "as it is". So the debate over representations will be seen in the context of this paper in terms of establishing the

proper relationship between concepts and experience and the precedence, if any, of one over the other.

After all, representationalists do not deny that humans form associations, and associationists do not deny that there is some role for symbols. The question herein is rather what are their proper roles and, in particular, whether these things are of any use at all in explaining and modeling cognition.

The paper is structured as follows. We begin with a short summary of the representational debate (Section 2), showing that a careful definition of terms helps to reveal where the real disagreements lie. We offer our own position, which is to reject the extreme positions that tend to dominate both sides. Section 3 offers an overview of Peter Gärdenfors's conceptual spaces theory, which he offers as a "bridging" account between associationist and symbolic accounts of cognition. We believe that its role as a bridging account is even greater than Gärdenfors allows. Section 4 presents the unified conceptual space theory, which proposes an extension of conceptual spaces in the direction of something that is more conducive to implementation in, e.g., a computer model, and so more conducive to empirical testing. Section 5 puts this theory to use in accounting for the co-emergence of concepts and experience. Section 6 revisits the representational debate in the light of lessons from Sections 3–5. Section 7 concludes both that our most abstract representational interpretations of the world must logically be a consequence of cognition that is grounded in sensorimotor engagements and that they have a conceptually *ineliminable* role to play in shaping those engagements.

2. The nature of the debate

One way to frame the debate is like this: Is mental content, be it conceptual or non-conceptual, best understood in terms of symbols (where the brain is thought of as, e.g., an instantiation of a Turing machine) or associations (between objects, between actions or events, or between observed regularities)?

Of course a symbolic account does not deny that associations play a role. But the meaning is in the things being associated — the symbols — not in the associations, and context is largely if not entirely irrelevant. Likewise an association-based account does not deny that we deploy symbols, though it may well downplay their usefulness to the discussion. Critically, however, the meaning is not in the things being associated but in the associations themselves. Context is key, and local details of structure (what might lead us to talk about this symbol or that symbol) are ultimately irrelevant.

2.1 So what's wrong with symbols?

Symbols are commonly described as having certain properties. As noted earlier, they are supposed to be arbitrary, so that form need bear no relation to function; and their meaning is supposed to be universal, unaffected by context.

Following Peter Gärdenfors, "the central tenet of the symbolic paradigm is that representing and processing information essentially consists of *symbol manipulation* according to explicit *rules*. The symbols can be concatenated to form expressions in a *language of thought*" (Gärdenfors 2004: 35), "A further claim of the symbolic paradigm is that mental representations *cannot be reduced* to neurobiological or other physicalistic categories" (Gärdenfors 2004: 37). In other words, cognition is fully independent of how it happens to be implemented (Fodor and Pylyshyn 1988: 54–56, 64–66). Thus the symbolic paradigm is inherently functionalist.

In contrast, "for [philosophers like] Locke and Hume, thinking consists basically in forming associations among 'perceptions of the mind'" (Gärdenfors 2004: 40). Translated into modern context, many of the contemporary intellectual descendants of Locke and Hume believe that cognition should be understood, in the main or entirely, as some kind of dynamical system strongly coupled with its environment.

One of the most famous and persistent advocates of the symbolic paradigm is Jerry Fodor, who set out the case for a "language of thought" (LOT) in his 1975 book (Fodor 1975) and has been defending it with some modifications ever since. According to the LOT hypothesis, thought is linguistically and hence symbolically structured. The relationship of thoughts (mind) to sub-personal mechanisms (brain) is like that commonly attributed to the relation of software and hardware in a computer; mind is functionally independent from brain as software is meant to be from hardware. This is the classical characterization of the computer metaphor of mind: the mind "just is" a computer (in the relevant, abstract sense), and cognition "just is" rule-based manipulation of symbols.

For advocates of the physical symbol system hypothesis, this is a full explanation of cognition: accounting for any particular cognitive skill is a matter of deducing the symbol manipulation rules in use (e.g., Newell 1980; Newell and Simon 1979). Their critics smell a homuncular regress: Who is reading the rules? Who is doing the representing, and to whom are the rules represented?

2.2 Re-considering terms

Part of the difficulty with framing the debate has to do with knowing what people mean by their terms, as terms are frequently used without being defined, raising

Figure 1. Forms of "One"

the possibility that people may, part of the time at least, just be talking past each other.

There are well-known problems with symbols as traditionally interpreted. Their meaning is often taken to be inherent to them, but logically this is suspect: this, in essence, is the *symbol grounding problem* (Harnad 1990). Their form is meant to be arbitrary, but often it is not. They are supposed to be discrete, and yet what should one make of the various related ways of writing the number "one" (see Figure 1). Or consider the two ways of expressing the number two in German: "zwei" or "zwo"; or compare the German "zwo" with the Swedish "två", which look quite different but are pronounced and mean the same. In each case, are they the same symbol with slightly different expressions or different symbols that happen to be closely related?

But perhaps the biggest difficulty with symbols is the question of *who* is using them to represent what to *whom*: for in what sense is a symbol meant to be a symbol in the absence of an agent to interpret it (explicitly) as such? This is, at best, left unclear by many on both sides of the debate. Inman Harvey writes: "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: 7). Failure to acknowledge the role of the observer in the act of representing leads to a confusion, and yet, "in practice, the role of the observer in the act of representing something is ignored" (ibid.: 5).

With these considerations in mind, we suggest that symbols may most usefully be understood as:

1. Modally grounded (that is to say, derived from sensory experience), but in such a way that the links back to that grounding may be difficult or impossible to reconstruct. To wit: symbols need be at most only *functionally* amodal.
2. Possessing an *apparent* arbitrariness of form.
3. Discrete, but in such a way as the boundaries between symbols may shift depending on context of application and over time. (See for example Harnad 1987).

Further, we agree with Harvey that it is critical to determine who is doing the representing and whom it is being represented to. Taken together, these points suggest that symbols as classically defined are best understood as idealizations.

This is why, for example, a computer is not, on its own, doing symbol processing — a point that Winograd and Flores (1986) elaborate upon. 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, those signs never become symbols, even in the impoverished sense of number crunching. We treat computers as idealized machines, operating on their own, neither embedded in an environment nor embodied in any particular form (when in fact they are both), unable to make a mistake (which they can, and do), for the same reason we treat symbols as a modal/discrete/context-free/arbitrary — likewise idealizations — because we find it conceptually useful.

What then of representations? We have already said that symbolic representations and symbols can be taken as the same thing. Iconic representations can then be understood as symbolic representations with the requirement for relative arbitrariness between form and meaning relaxed: with iconic representations, the relationship between form and meaning is still, to greater or lesser extent, apparent; with symbolic representations, that relationship has, for most practical purposes, been lost. On a scale of symbolic to non-symbolic, iconic representations are located a little more toward the non-symbolic end. Turning that around, symbolic representations can be seen as an impoverished form of iconic representation. Both lie on a continuum where symbols or symbolic representations, traditionally interpreted, lie at an unreachable extreme.

What makes a symbol an effective symbol — what makes it recognizable as a symbol in the first place — is the extent to which it abstracts away from any particular context of interpretation, to be applicable across the broadest possible range of contexts. 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. What makes a(n iconic) representation an effective representation is *both* the extent to which it abstracts away from any particular context of interpretation *and* the extent to which it retains its links back to particular contexts. Symbols are unstructured relative to the domain of interpretation. Representations — at least iconic ones — are not. The form of a symbol need not relate in any obvious way to its meaning; the form of an iconic representation should.

2.3 Re-framing the debate

As this paper will attempt to establish, the real question is not, *are there symbols in the brain?*, but rather, for the committed representationalists, can a full account of cognition (or at least, of "higher-order" cognition) be given solely in terms of symbolic and representational language; and, for the committed anti-

representationalists, can a full account of cognition (or at least, of *the vast majority of* cognition) be given without any resort to such terms?

That we, as observers of a world, interpret that world symbolically seems undeniable. As one introspects, observing one's own experiences, it seems that those experiences "just are" representationally and symbolically structured because that is how we (seem to) naturally interpret them. For example: most people report an inner linguistic dialogue *sotto voce*, and likewise find their dreams full of meaningful representations.

On an anti-representational account, symbol use, while generally enduring as something to be explained, is seen as emergent, even epiphenomenal, and not necessary for an account of the mechanisms constituting our cognitive processes. In contrast, on a representational account symbols are typically treated as *the atomic components of thought*, the appropriate rule-based manipulations of which give rise to cognition (Anderson and Lebiere 1998).

The position we will take in this paper is that representations, properly understood, are both *necessary* for understanding cognition, even for the most basic sensorimotor levels, due to our inability as conceptual agents to step aside from our representational perspective, and *not sufficient* for understanding even the most abstract levels of symbolically structured thought, given the ubiquitous involvement of (non-representational) sensorimotor engagements in cognition. Note that the representations need not in any way be in the brain of the agent being observed but rather are bound to the perspective of the agent doing the observation.

We will borrow a page from Gärdenfors and suggest that concepts are best understood as an intermediate level of cognitive explanation between associationist and symbolic accounts, where neither view on its own provides a complete account. When one level of explanation is being emphasized, concepts will look more like abilities: things to be possessed and employed for the most part non-reflectively; when the other, more like representations: things that may be reflected upon. It is to Gärdenfors's account that we now turn.

3. Conceptual Spaces Theory

Gärdenfors writes, "The fundamental cognitive role of concepts is to serve as a bridge between perceptions and actions" (Gärdenfors 2004: 122). Again: "…I argue that conceptual spaces present an excellent framework for 'reifying' the invariances in our perceptions that correspond to assigning properties to the perceived objects" (ibid.: 59).

To reify is, of course, to give concrete expression to something abstract. Arguably one of the most striking aspects of concepts is the way, much of the time,

they relate the very abstract (ideas) and the very concrete (physical objects and actions). One might be tempted to go further than Gärdenfors here and suggest that conceptual spaces present an excellent framework not only for understanding (as well as we can) the nature of those invariances but also for telling a compelling story of both how they arise out of our perceptions and how at the same time they structure those perceptions, with no clear "pride of place" to which comes first.

Conceptual spaces theory is a similarity-space-type theory, owing much to prototype and exemplar theories of concepts, whilst presenting concepts within the neutral language of geometry. The central tenets of conceptual spaces theory we take to be:

- Neither associationist (or connectionist) nor symbolic accounts of cognition, and likewise neither empiricist nor rationalist approaches, can, on their own, do adequate justice to the nature of concepts. The former are too reductionist, the latter too rarefied.
- Just as accounting for cognition in terms of concepts bridges different levels of explanation of cognition, so, too, concepts bridge different levels of cognition — one more unconscious and automatic if not in fact subpersonal, the other indisputably personal if not in fact conscious and deliberate.
- "There is no unique correct way of describing cognition" (Gärdenfors 2004: 2) — and there is no unique correct perspective on concepts. Conceptual spaces theory is intended, if you will, as the best compromise. "In brief, depending on which cognitive process we are trying to explain, we must choose the appropriate explanatory level" (ibid.: 57). The same might be said of what aspects of concepts we are trying to explain.
- There is no unique correct perspective on any particular concept, not least because concepts change with the agent who is using them and the context in which they are used. Gärdenfors specifically includes the so-called natural kinds concepts.
- A metaphor for objects in physical space, concepts are (best understood) either as:
  - *Points* (or associated sets of points) within conceptual spaces, whose dimensions (e.g., hue, saturation, and brightness in the case of color) may be acquired in a bottom-up activity-driven manner or a top-down intentionally-driven one (cf.: "In a conceptual space that is used as a framework for a scientific theory or for construction of an artificial cognitive system, the geometrical or topological structures of the dimensions are *chosen* by the scientist proposing the theory or the constructor building the system…. In contrast, the dimensions of a *phenomenal* conceptual space are not

obtainable immediately from the perceptions or actions of the subjects, but have to be *infered* from their behavior" (Gärdenfors 2004: 21)); or as
  - *Shapes* (or associated sets of shapes) within those same spaces.

- Those shapes are typically (though not always) *convex* shapes: that is, for any two points that lie within the concept *x*, all points between them should also lie within that concept. (Some concepts are defined as the negation of other concepts, within a certain domain: e.g., Gentiles are anyone who is not Jewish. Fodor uses the example of NOT A DUCK. If one concept [JEWISH or DUCK] is convex, its negation [GENTILE, NOT A DUCK] within a domain cannot be convex.)
- Individual convex shapes (or individual points) denote a particular type of concepts: properties. Other types of concepts — i.e., object concepts or action concepts — are associated sets of these shapes (or points). Note that, just as individual shapes can be "collapsed to" points, so too associated sets of these shapes can be collapsed to a single shape: i.e., all concepts can be treated as properties (property concepts). This is not explicitly stated, but we take to be implicit in Gärdenfors's account. (By analogy, think of the way that, in English, in general, nouns can take the role of adjectives: e.g., "**bicycle** thief", "**happiness** patrol". Verbs do something similar but change their form: "**cycling** champion".)
- The structure of concepts, if not the concepts themselves, need not be consciously introspectible: "…For many words in natural languages that denote properties, we have only vague ideas, if any at all, about what are the underlying conceptual dimensions and their geometrical structure" (Gärdenfors 2004: 168).
- The process of "carving up" a conceptual space into various possible shapes is the process of *categorization*: "…Where (possible) objects are represented as points in conceptual spaces, a categorization will generate a partitioning of the space and a concept will correspond to a region (or set of regions from separable domains) of the space" (ibid.: 60).

3.1 Conceptual spaces as located within an enactivist framework

Although not explicitly enactivist, it is remarkably easy to locate conceptual spaces theory within an enactivist framework, where enactivism is understood in the philosophical tradition of Francisco Varela, Evan Thompson, and Eleanor Rosch, co-authors of *The Embodied Mind* (Varela et al. 1991): "I have proposed using the term *enactive* to… evoke the idea that what is known is brought forth, in contraposition to the more classical views of either cognitivism or connectionism"

(Maturana and Varela 1992: 255). "The roots of mental life lie not simply in the brain, but ramify through the body and environment. Our mental lives involve our body and the world beyond the surface membrane of our organism, and so cannot be reduced simply to brain processes inside the head" (Thompson 2007: ix).

Of course different researchers use "enactivism" in different ways.[1] Enactivism, as we wish to use the term, should not be confused or equated with the twin notions of embeddedness (or situatedness, i.e., an agent is located in a particular spatiotemporal context) and embodiment (i.e., an agent takes a particular physical form), as much as it does embrace them. Enactivism goes beyond embeddedness/embodiment by:

- Understanding cognition, at least in the first instance, as a *skillful activity*, and in any case as a lived, dynamic process and not a static entity.
- Typically perceiving continuities as underlying that which appears individuable and discrete, most notably the continuity between agent and environment.
- Taking an agent/environment, internal/external distinction to be both conceptually necessary and, at the same time, meaningful *only with respect to an observer* (and not to the organism itself independently of some observer).
- Giving a foundational role to phenomenology and emphasizing the essential contribution to be made by first-person perspective and first-person methods.

Contemporary usage has much of its roots in a book by Humberto Maturana and Francisco Varela (1992), for whom in many ways the concept is bound up with another notion, *autopoiesis*. Autopoiesis is intended as an alternative description of what qualifies as a living organism, in terms of operational closure (processes of the system are produced from within the system; anything external to the system can play only the role of catalyst), autonomy, and the observation that organisms "are continually self-producing" (Maturana and Varela 1992: 43). On the other hand, Alva Noë (2004) has called his own approach to cognition enactive but does not talk about autopoiesis, is more specifically focused on sensorimotor engagements (and less on the coupling between cognition and life), and is more recognizably (and self-avowedly) externalist (rather than seeking to avoid either internalist or externalist labels). Noë has more recently preferred to call his position "actionism" rather than "enactivism" (e.g., in Noë 2007), perhaps to avoid possible confusions.

Conceptual spaces theory, as well as our own position, fits in best, we think, with the enactivism of Maturana and Varela, or such contemporary philosophers as John Stewart or Evan Thompson, both of whom published with Varela. This might seem to locate us with what one might call the "strong enactivists" (as opposed to Noë, for instance). There are, however, certain important caveats, which will describe our own position as well:

- Conceptual spaces theory lacks their often strongly anti-representational bias and is even favorably disposed toward mental representations, properly understood.
- In consequence, conceptual spaces theory is favorable to their view of cognition as skillful activity but only when interpreted sufficiently broadly as to leave room for mental representations and to bridge the knowing that/knowing how divide.
- As with Noë's project, conceptual spaces theory recognizes their relationship but is not concerned to make such a tight coupling between cognition and life. (Its focus is not so much on self-organizing systems, and it makes no mention of anything resembling autopoiesis.)
- Conceptual spaces theory is therefore more sympathetic to the conceivability, at least, of artificial intelligence as distinct from artificial life.

Both enactivism and conceptual spaces theory can be seen in the context of the history of cognitive science as part of a broader movement away from largely disembodied and "purely" symbolic accounts of cognition 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 as independent from semantics, and so on. At the same time neither should be taken as final destinations (conceptual spaces theory is quite clear about this) but only as points along a path.

3.2 Empirical testing to date

One might well rue the frequent disconnection between abstract theory and empirical testing, and nowhere may this be clearer than in the intersection between philosophy of mind and cognitive science. The analytically inclined philosophers of mind decry the continental philosophers for their lack of empirical grounding, and yet their attempts at naturalization have met, at best, with mixed results. Theories translate imperfectly into implementable models, and empirical results are nearly always open to interpretation. It can be difficult to find the middle ground between armchair reflection on the one hand and cleverly designed tests or applications of dubious theoretical import on the other, and it might seem there is a tendency to slide off in one or the other direction.

One philosopher who has taken particular pains to exploit the middle ground is Ron Chrisley with his work on expectation-based architectures: roughly, this is the role expectations play in being partly constitutive of experience: i.e., expectations as the *products* of experience also *shape* experience (Chrisley and Parthemore

2007). The theory informs the (robot-based) model and the model informs revisions in the theory, which then informs revisions in the theory.

What then of conceptual spaces theory? As with this present paper, conceptual spaces theory as is strongly on the theoretical side of the divide. At the same time, Gärdenfors has attempted to create a theoretical structure that invites testing. We can but hope to do the same.

Three papers deserve mention here. The first (Gärdenfors and Williams 2001) does not present any new empirical research but rather seeks to locate conceptual spaces theory in the context of existing evidence in psychology for prototypes and their relationship to categorization, and indeed argue that conceptual spaces theory can provide a *better* account of that relationship relying on computation rather than fuzzy intuition. The goal is a more algorithmically precise description of conceptual spaces theory by relating it to something called the Region Connection Calculus (RCC). More algorithmically precise is, of course, easier to implement in a computer model (or test in a psychology experiment). Voronoi tessellations are used to determine category boundaries, and then the RCC is used to reason about them. Particular attention is paid to the "crisping" or "blurring" of boundaries and how that may be used to account for non-monotonic reasoning (i.e., "if *X* then *Y*, *ceteris paribus*").

The second (Chella et al. 2004) *does* offer new empirical research — involving two mobile robots, each using conceptual spaces to navigate their environment — but the account (less than half a page out of a six-page paper) is extremely brief, making it difficult to know what actually has been implemented and what conclusions can reasonably be drawn. The main concern of the paper is, again, more to lay the groundwork for further empirical testing: in this case, focused on what the authors call *perceptual anchoring*, the linking up of discrete concepts as symbolic entities and the continuous observable quantities provided by the sensors, which they take to be a special case of the symbol grounding problem (Harnad 1990). The pseudocode procedures for finding, tracking, and acquiring (or re-acquiring) anchors relate nicely to our account in the introduction to Section 4 of examining one's unified conceptual space in terms of the queries "What is here?", "Is this here?", and "What if this were here?".

The final paper (Chella et al. 2008) extends the ideas about perceptual anchoring. It offers the best glimpse into how conceptual spaces theory might be tested empirically and applied concretely, in this case within the emerging (and controversial!) field of machine consciousness. Here much of the emphasis is on metacognition: "We claim that one of the sources of self-consciousness are higher order perceptions of a self-reflective agent" (ibid.: 153) — i.e., the capacity of an agent to have concepts *about* its concepts. Unfortunately for present purposes, much of the (again short) paper is devoted to describing the robot's "conceptual area" (one

of three cognitive levels being modeled, the other two being the "sub-conceptual area" and the "linguistic area") in terms of low-level mathematical justifications rather than high-level algorithmic descriptions. So again, it is hard to know what precisely has been implemented or tested: notably, the question of how the robot is to deal with the sheer volume of possibilities opened up by having first-, second- and higher-order concepts all available within its conceptual area is touched upon but not resolved. However unlike in the earlier robotic system, salience (the so-called *frame problem*) *is* addressed and indeed given a treatment quite reminiscent of Chrisley's work on expectation-based architectures.

3.3 Limitations and difficulties

> Philosophers will complain that my arguments are weak; psychologists will point to a wealth of evidence about concept formation that I have not accounted for; linguistics [sic] will indict me for glossing over the intricacies of language in my analysis of semantics; and computer scientists will ridicule me for not developing algorithms for the various processes that I describe. I plead guilty to all four charges (Gärdenfors 2004: ix)

Some, including the present authors, might find Gärdenfors's use of the term "representation" overly broad, including much that one might prefer not to call representational. We have, in Section 2.2, offered our own account of what should and should not be called a representation, based on Harvey (1992). But this is a minor point.

In our opinion, the greatest strength of conceptual spaces theory is its generality — but that, at the same time, poses its greatest limitation. In contrast to much of the literature in this area, Gärdenfors is refreshingly modest and candid about how, in many ways, conceptual spaces theory provides only the scaffolding — and like all true scaffolding, it is meant to be removed once the structure (which might itself be the scaffolding for yet another structure) is in place. Gärdenfors focuses much of his attention on the concept of colour, with its relatively uncontroversial dimensions of hue, saturation, and brightness. For any other concept in a conceptual space, how should one methodically go about determining its integral dimensions?. This is among the details to be filled in, and is especially important when dealing with the non-linearity and temporal aspects of most real world data.

It is to the fourth of Gärdenfors's "charges" that the next section of this paper will largely be devoted. In particular, when it comes to talk of concepts as the building blocks of thought, we hope to put some more literal flesh onto the bones of that familiar metaphor.

4. The unified conceptual space theory

Section 3 presented our analysis and summary of Gärdenfors's conceptual spaces theory, along with its empirical testing to date. Here we will introduce our proposed extensions to it, with the goal of offering a more algorithmically inclined account and in particular introducing a new notion, the unified conceptual space, along with ideas of how the mechanism for it should best be implemented within a particular model.

As Edouard Machery has noted in his critique on concepts (Machery 2009), most theories of concepts are trying to achieve something close to a uniform account of concepts. Jesse Prinz points out that such uniformity is easier to achieve if concepts themselves are uniformly structured (Prinz 2004: 94). In the case of a conceptual-spaces-type theory, uniformly structured concepts imply uniformly structured conceptual space if not, in fact, a single unified space.

One of the primary author's concerns upon reading Gärdenfors (2004) for the first time, was how are all the different conceptual spaces that Gärdenfors talks about unified within a single (complete, in some sense, if not strictly consistent) conceptual space? — because it might seem, and might seem consistent with many if not most people's intuitions, that our structured thoughts *do* exist within a common "space": i.e., that they all fit together somehow. One ought not to think that Gärdenfors intends to imply otherwise. But though he does offer clues about how different spaces map onto one another, the geometry of any unified space is beyond the remit of that work.

We want to suggest that all concepts exist within a common space that is the mapping together of many different conceptual spaces, all with the same general structure. We are speaking here in the first instance of concepts for an individual, though an analogous space must logically exist for a society — for those conceptual agents who are social animals — mapping together the different conceptual spaces of its members into a unified space of the whole. That is to say, the unified conceptual space is a coming together of many spaces, a space of spaces.

An account of such a unified space should be able to answer such basic questions as:

- *What is here?* Given a set of coordinates in the conceptual space (which may only be relative to some other set of coordinates), say what is found at that point or in that sub-region.
- *Is this here?* Allow mappings between different parts of the conceptual space, and therefore comparisons. Given a set of coordinates and an expectation of what to find there, say whether or not the expectation is met.

– *What if this were here?* Allow simulations and possible-worlds-type scenarios. Given a set of coordinates, allow substitution of what is not found there (in terms of what might be or what might have been) for what is. This last question implies some level of meta-cognitive abilities.

Where should one begin? Wilhelm Geuder and Matthias Weisberger attempt to show (Geuder and Weisgerber 2002) how one might go about constructing a unified conceptual space in the sub-region of action concepts, clearly intending for it to apply more widely. In so doing, they contrast the conceptual spaces approach with e.g., Fodor's informational atomism, in which Fodor is quite explicit that particular concepts could, in principle, exist in utter isolation from one another (see e.g., Fodor 2008: 54). In conceptual spaces, concepts exist within a universe of concepts, and together they define the space. The value of a unified conceptual space for the individual conceptual agent is, as Geuder and Weisberger suggest, to make the incommensurable commensurable, to abstract away from differences to perceive what is in common, to blend many different conceptual spaces into one space.

4.1 Conceptual building blocks: Two contrasting perspectives

We want to suggest further that concepts can be given two contrasting structural descriptions. One is as structured shapes within the unified conceptual space, where any point that lies within the shape lies within the concept. Note that particular points can be understood both as specific instances of the concept in question and, typically (though not always), classes of more specific instances. The other is as structured logical constructs — *not* shapes — attached to particular points in that space. Something akin to this duality of perspective we take to be implicit in Gärdenfors's account.

> Criterion 1: All concepts can be viewed either as well-behaved shapes within a unified conceptual space (spatially ordered collections of contiguous points) or as single points (spatially unstructured).
>
> Criterion 2: All well-behaved shapes that are interpretable as concepts can, optionally, be assigned an "arbitrary" symbol or label. That label then accrues to all points within that subspace. A sufficient set of such labels can be used in constructing a language.
>
> Criterion 3: All specific instances (points) can in principle (but within limits of practicality) be treated as classes (collections of points: i.e., shapes). The distinction between instances and classes is a pragmatic not an absolute one.

So for example, WEIGHT is both a specific instance of MEASUREMENT and itself a class of different weight values: e.g., 13 kilograms. THIRTEEN KILOGRAMS is both a specific instance of WEIGHT and a class of weight values between (typically) 12.5 and 13.49 kilograms.

Likewise CAT is a specific instance of MAMMAL and itself a class of various breeds of cats. MY CAT KALI is both a particular cat and a class of all my experiences of Kali, in different times and contexts. At some point, the instances-to-classes expansion has to stop for strictly practical reasons: MY CAT KALI AT 12:03 PM YESTERDAY may be as specific as my sensory experience and memory allow.

> Criterion 4: All points in conceptual space have both local and distal connections to other points in the space.

The local connections will be treated in Section 4.1.1, as will those distal connections concerned with conceptual reference. The remaining distal connections — those concerned with conceptual sense — will be treated in Section 4.1.2.

4.1.1 *Structured shapes in conceptual space*

> One important aspect of the organization of conceptual knowledge has to do with the hierarchical structure of concepts…. This hierarchical structure can be interpreted as facilitating our thinking about objects and entities. The facilitation arises out of the relations between levels of the hierarchy and information associated with the different levels. Knowing what a mammal is and that a bat is a mammal, allows us to distinguish it from birds and group it together with other animals that nurse their young (Hemeren 2008: 24).

As physical space is a space of physical entities, a space of *material substance* and of seemingly non-material energy, so conceptual space is a space of conceptual entities, a space of *properties*. Its partitioning into concepts (one could talk of partitioning into potential concepts; but we believe it more likely that, where there is no need to partition, there is no partitioning) results in sub-regions that are themselves unified spaces and, at the same time, in many instances associated sets of separable spaces, as we will describe below.

4.1.1.1 *Dimensions of the space.* We propose that the unified conceptual space is describable along three axes: an *axis of generality*, from maximally general (the concept of a concept itself) to maximally specific (a particular concept) — this is the axis that Hemeren is referring to; an *axis of similarity*, from maximally similar to maximally different, along any or more of the concept's integral dimensions; and an *axis of perspective*, from maximally broad to maximally narrow descrip-

tions, by choice of set of integral dimensions to attend to or by context. Note that the concept of a concept itself is also a particular concept; so in some way the first axis must wrap around on itself, like the circular number line in modulus arithmetic. The values along the second axis are strictly exclusive of one another, while those along the third axis are along a sliding scale of inclusiveness (more inclusive to less inclusive).

Figure 2 visualizes the first two of the three axes (reproduced by permission from (Hemeren 2008: 25).

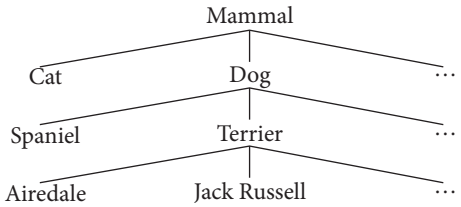

Figure 2. Two dimensions of conceptual space

Note further that all three axes are divergent in both directions. Because the axes are divergent, the space itself is divergent, fragmenting into many parallel spaces, or better, parallel slices within the same overall hyper-dimensional space. Take the first axis: any given concept can typically (though not always) be seen to belong to one of several different classes — what might elsewhere be termed *multiple inheritance*; and any general concept can typically (though not always) give rise to any of several more (mutually exclusive) specific instances. So for example, "teenagers", "infants", and "pensioners" are all more specific instances of "people", which itself is a more specific instance of "biological organism" and of "intentional agent", where the latter may or may not be a particular instance of the former, as philosophers and cognitive scientists are fond to argue over. (On the other hand, "brown" is *only* a specific instance of "colour" — more on this in Section 4.1.2.3.)

Likewise, the second axis is divergent, depending on which of the integral dimensions are attended to (presuming there is more than one). Consider a shade of brown, which may be adjusted according to hue, saturation, or brightness, or some combination of the three. Compare this to moving around within the conical colour space cited as a frequent example in Gärdenfors (2004).

Finally, the third axis is divergent, depending on what strategy is adopted in specifying the concept. Given a shade of brown, one could specify it more narrowly or more broadly in terms of any of its integral dimensions — a brown with *that* hue and *that* brightness — or one could e.g., offer examples of things that typically are that shade of brown.

> Criterion 5: A concept is an identified well-behaved sub-region of the conceptual space.
>
> Criterion 6: Most concepts — what Gärdenfors calls the natural concepts — will be convexly shaped, have prototypes, and exhibit typicality effects.

Speaking more accurately, of course, Gärdenfors talks of natural concepts as correlated sets of convex shapes. We believe we have shown an elegant way that his sets of correlated shapes can be superimposed on each other to make a single shape composed of many shapes. Note that the concepts-by-negation mentioned earlier will generally not have prototypes nor exhibit typicality effects.

4.1.1.2 *Space mapping onto space.* Concepts have referents: they map shapes in conceptual space onto other (distal) shapes in conceptual space (properties, which are themselves implicitly or explicitly higher-order concepts), onto shapes in physical space (matter- and energy-based entities that occupy that space), or onto shapes in what we will call the temporal plane (actions/events). These different sorts of concepts occupy different sub-regions of the overall conceptual space. Although physical entities, action/events and properties are conceptually quite distinct things, we want to suggest that the division between sub-regions is a pragmatic one, subject to shifts, and that, on pain of paradox, no absolute dividing line can be determined.

The temporal plane is the domain of agent-intentional actions and agent-less events, of *process*. One axis of the temporal plane is the familiar time line, extending to past and future from a particular observed moment; the other is an axis of alternative possibilities or *possible worlds* that are *not* observed (but can only be imagined), ordered from closest / most similar / likeliest; to most distance / least similar / least likely.

We now turn to the second way of specifying the structure of concepts.

4.1.2 *Shapeless logical constructs*

We have suggested earlier that, within conceptual spaces theory, all concepts can be treated as property concepts, even where they are also (and, in many contexts, more importantly) physical-object concepts or action/event concepts. In this way all concepts, sharing a common function (to describe some perceived regularity), also share a common basic form, even when in other ways they differ substantially.

> Criterion 7: All concepts, in addition to the proximal connections between contiguous points, and the distal connections to their referents, map to distal parts of the conceptual space in two additional ways:

> Criterion 7a: Integral dimensions: in this way, for example, COLOUR maps to HUE, SATURATION, and BRIGHTNESS. Integral dimensions are *necessary* (a colour *must* have a hue, saturation, and brightness) but not ordered in any way (there is no priority between hue, saturation, and brightness). We will henceforth also refer to these as *parameters*. A concept need not have more than one integral dimension, but it must have at least one; otherwise, it merges into some other, more general concept. Note that the parameters define a conceptual space of their own, whose precise nature depends on the number of parameters: e.g., for COLOUR, the parameters define a cone-shaped region of alternative possibilities.
>
> Criterion 7b: Contextual elements: things that the concept's referent is co-present with in different contexts. If the thing is associated with the concept in a majority of contexts, then one can say that it is typically associated with that concept. For example, cats are typically (though not always) associated with purring. Contextual elements are neither ordered nor necessary; however, a sufficient number of contextual elements may limit the possible scope of a particular concept to a specific instance of that concept (or to no possible instance). We will refer to these as *contextuals*. A concept cannot be given coherent interpretation except with relation to some non-empty set of contexts, and therefore contextual elements.

4.1.2.1 *Physical-object concepts.* As Hemeren (2008: 55) notes and Machery (Machery 2009) corroborates, much of the research into concepts particularly in psychology has focused on object concepts — meaning generally concrete and not abstract objects, which is how we will use the term — and their relationship to nouns in language. This may make them in some ways better understood than other sorts of concepts, but there is potentially a double distortion going on: a conflation of object concepts with nouns (and therefore with language), and a tendency to view all sorts of concepts through the lens of (physical) object concepts.

Concepts of physical objects are *first-order concepts*: i.e., they are concepts of non-concepts.

> Criterion 8: In addition to integral dimensions (parameters) and contextual elements (contextuals), object concepts can be decomposed into one or more parts, which we will call *components*, all of which are spatially ordered (that is, they bear a certain ordered relationship in physical space) and one or more of which are necessary. The components will also be object concepts. (Like decomposes into like.)

Consider the concept of a man: the (minimum) necessary components relative a particular application of the concept, and subject to revision, might be a head and

torso. A man with no arms or legs is still a man, but a man with no head or no torso is in most instances something else albeit man-related: e.g., a corpse. Which precisely the necessary part(s) is/are is not important: those can be pragmatically determined by present context and subject to change. (If one meets a living man who *is* just a head and nothing more, one might well revise one's concept to incorporate this new possibility.)

> Criterion 9: The parameters of object concepts will typically if not always lie in higher-order conceptual space, the domain of property concepts. This is not peculiar to physical-object concepts; it will be true of all concepts.
>
> Criterion 10: The contextuals of object concepts split into two groups: those that are located in a common physical space with the object in question, and those that are located in the temporal plane: both agent-centric actions and agentless events. These are two different, but complementary, senses of context: on the one hand, where you find an object *x* you may find objects *w*, *y*, and *z* as well; on the other, given an object *x* you may find action/events *a*, *b*, and *c* all involving *x*.

Components, parameters, and contextuals all inherit. A man has (must possess) a head because a human has (must possess) a head; HUMAN likewise inherits from MAMMAL, and so on. Likewise a man has a weight because all physical things (in the usual context) have a weight. On the other hand, because the contexual relation is customary rather than necessary, contextual inheritance is *ceteris paribus*: if cats typically purr, then *ceteris paribus* Siamese cats should be expected to purr (though perhaps they do not), my cat Kali should be expected to purr (though perhaps she does not), and my cat Kali should or might be expected to purr on any particular occasion (though very possibly she will not!).

4.1.2.2 *Action/event concepts.* For whatever reason, action concepts have been relatively neglected compared to physical-object concepts in empirical research. This is unfortunate, as without having an understanding of what they have in common, it is difficult to appreciate the ways in which they do, indeed, differ. With his own research, Hemeren found a similar if somewhat simpler structure, similar if somewhat shallower hierarchical organization including analogous "basic level" effects, and so on. Action concepts and object concepts do, indeed, seem to be of a common genus and not, as they would probably be on e.g., Machery's account, entirely different species of things.

On our account action/event concepts, like physical-object concepts, are, in the first instance, first-order concepts. If they are more simply structured than object concepts it is, perhaps, because they map to entities in a temporal *plane* rather than entities in a physical *space*.

> Criterion 11: Like object concepts, action/event concepts can be decomposed into one or more parts or *components*, which are *temporally* ordered and one or more of which are necessary. The components will also be action/event concepts.

Consider the concept of pitching as a type of throwing. The minimum necessary components of it are, perhaps, an aiming, a shaping of the hand, a drawing back of the arm, a snapping forward of the arm, and an opening up of the hand. All of these save the aiming are inherited from throwing: PITCHING minus AIMING is no longer PITCHING but probably only THROWING. Our intuition — which would be interesting to test out empirically — is that action concepts are relatively more stable than object concepts. That is, the *necessary* components may be relatively more clear cut and any optional components correspondingly less important.

> Criterion 12: As with other concepts, action concepts take one or more parameters, which typically if not always lie in higher-order conceptual space.
>
> Criterion 13: As with physical-object concepts, the contextuals of action/event concepts split into two groups: on the one hand, other action/events happening in the same location on the temporal plane; on the other, the agents initiating the actions and the entities in physical space with which the action/events interact.

4.1.2.3 *Property Concepts.* If action concepts are somewhat more simply structured than object concepts, then property concepts are (much) more simply structured, again, by virtue of being more abstract and further removed from the more basic levels of proto-concepts and first-order concepts. They are simplified from other concepts in two significant ways:

> Criterion 14: As noted above, property concepts typically do not exhibit multiple inheritance. That is to say, they are describable *only* relative to one domain, where "domain" is defined as "*a set of integral dimensions separable from all other dimensions*" (Gärdenfors 2004: 26). BROWN is *only* a sub-category of COLOUR and so can only be described along the integral dimensions (or *parameters*) of HUE, SATURATION, and BRIGHTNESS. As Gärdenfors notes, most if not all properties we can name can be understood as convex shapes within a single domain.
>
> Criterion 15: Unlike other sorts of concepts, property concepts do not decompose into ordered parts: in our terms, they have no components. Rather than being defined in any way by their components, they are strictly defined by their parameters and contextuals.

It may be relatively clear that property concepts such as BROWN, MASS, and TRANSPARENCY have no components. But what of other candidates, such as MIND or DEMOCRACY?. We want to suggest that MIND should not be understood as having components such as e.g., SELF-CONSCIOUS, CONSCIOUS, and SUBCONSCIOUS. They need not be ordered in any kind of space, including conceptual space. Rather, these are integral dimensions (parameters) of the (human) MIND (itself an integral dimension of the human brain). Likewise, democracies do not have any identifiable parts, but they do have generally agreed upon properties. We suggest that this follows, in general, for concepts of all non-physical, non-temporal entities.

> Criterion 16: As with other concepts, property concepts take one or more parameters, which lie in higher-order conceptual space. Because property concepts are not defined in any way by their components — they have none — their parameters are perhaps heightened in importance relative to other concepts.
>
> Criterion 17: The contextuals of property concepts divide into three (rather than, as for physical-object and action/event concepts, two) groups:
> - Other (more closely or more loosely associated) properties. (Where one finds COLOUR, one often finds WEIGHT and MASS as well.)
> - *Either:*
>   - Objects in physical space that they are properties of; *or*…
>   - Action/events in the temporal plane that they are properties of.

4.2 Limitations and exclusions

The presentation of the proposed extensions and revisions to conceptual spaces theory is necessarily abbreviated. Although earlier versions of these ideas have been implemented in, e.g., a writing environment for short story design (Parthemore 1990) as well as experimented with in several smaller software projects, the unified conceptual space theory as presented here has yet to be implemented in any models or otherwise tested empirically. Putting theory into practice is, perhaps, the best way to discover what can and cannot work.

The theory presented here is static, and that is probably its greatest limitation. Although it refers to dynamic processes, it has no intrinsic dynamics of its own; all its dynamics are external. It is neither part of a biological organism nor — if the two are separable — is it part of any intentional agent, interacting with its environment to create an ever-shifting conceptual space that is neither agent nor environment. It is neither embodied nor embedded in an environment. It has neither sensory apparatus nor motor system.

The questions of concept acquisition, application and revision will be provisionally addressed in the next section. But although it will be touched upon and some preliminary suggestions offered, the critical question of salience — that is, why concepts get mapped to their referents in the first place, why e.g., a sub-region of conceptual space gets mapped to a sub-region of physical space — will remain largely unresolved.

Finally, the model of the unified conceptual space faces the twin dilemma of both not being algorithmic enough to permit direct translation into e.g., a computer model — it is a step prior even to pseudo-code; and, arguably, being already *too* algorithmic to capture some things one might well want to say about concepts. Although we agree completely with Rich Grush and Pat Churchland (1995: 190) when they write that it is unclear whether *anything* in the universe cannot be described algorithmically, nonetheless what *is* clear is that some structures and processes are easier to describe algorithmically than others, and that will inevitably prejudice our perspective.

5. The co-emergence of concepts and experience[2]

It should be said that the idea of combining bottom-up and top-down approaches to understanding cognition is far from new. The approach described here shares the spirit though not the style of Antonio Chella et al's (2000) approach to modeling dynamic scenes. Likewise the idea of co-emergence is familiar throughout the enactive literature, where the discussion is frequently on the co-emergence of agent and environment or form and meaning (see, e.g., Thompson 2007).

There is a certain unavoidable tension between concepts and experience. Experience is typically very much engaged *in* the moment and in any case is grounded in the present. Concepts, on the other hand, abstract *away* from the immediate perceptions or experience of the moment, stepping back from the present moment to take a wider view. They have, as it were — to borrow an idea from Lawrence Barsalou (see Barsalou et al. 2007) — one hand in the past and the other in the future. Not only is it unclear whether "concepts" solely of application to the present moment would be of any use, it is unclear whether they would really qualify as concepts at all.

At the same time, concepts and experience seem critically dependent on one another. Unless one wants to postulate a large body of innate or spontaneously arising concepts — and even the latter-day Fodor seems reluctant to do that (Fodor 2008) — most concepts will require experience to give rise to them. As we will argue in Section 5.1, that must ultimately be *sensorimotor-based* experience: the agent must be cognitively and physically engaged with its environment

to experience environment or self. Without such engagement, not only will there be no conceptual development, cognition in general will at the very least be drastically impaired. The standard reference here, of course, is the classic study on kittens reported in Held and Hein (1963) in which kittens, deprived of the ability to interact visually with their environment, appear subsequently unable to make sense of that environment.

Experience seems likewise dependent upon concepts. It is an open question and, perhaps, an unresolvable one, in what sense of the word an agent without any concepts could be said to have experiences. In such an agent the "experienced" world would, in every instant and 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 (as we have defined them) some at least minimal conceptual abilities.

For the conceptual agent, such experience uncoloured by concepts (if experience it is) seems no longer a possibility. Such an agent may never experience the "now" entirely on its own (though perhaps forms of meditation may get her closer than she would get otherwise); instead, the "now" is experienced in the light of past moments and in anticipation of future ones. As Gärdenfors (2004: 4) writes:

> We frequently compare the experiences we are currently having to memories of earlier episodes. Sometimes, we experience something entirely new, but most of the time what we see or hear is, more or less, the same as what we have already encountered. This cognitive capacity shows that we can judge, consciously or not, various relations among our experiences. In particular, we can tell how *similar* a new phenomenon is to an old one.

Concepts reliably shape and re-shape our experience of the world. So although 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. In the language of Noë (2004: 1–3, 78–79), once an agent has a sensorimotor-grounded profile of doorknobs, that profile is inextricably part of how that agent encounters and experiences her world.

Most if not all concepts require experience to give rise to them. Most if not all experience requires concepts to give structure to it. It's like the chicken-and-the-egg problem: which comes first? Logically something must start things off, but it need not be either concepts or experience as we understand them. Caught within our conceptual perspective, we cannot step outside of it: we cannot simply put our concepts or our conceptually structured experience aside.

Concept acquisition and application go hand in hand. Acquiring concepts is a process of applying concepts, which may themselves change in the process of acquiring the new concepts. In the language of the previous two sections of this paper, our conceptual spaces, individually and collectively, are both the product of our interaction with our environment and the basis for it. The model of causality is not linear but circular.

It is all well and good to talk about the structuring of thought and the experiencing of a world, the acquisition of concepts and their application in experience, as one single process viewed from two perspectives, like two sides of a coin. But 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 somewhat different rules. Here is where conceptual spaces theory, and our proposed extensions to it, come into their own. First, however, we must provide some more theoretical grounding.

5.1 Concepts emerge from *sensorimotor* experience

Concepts are paradigmatically abstract and cognitively high level. Sensorimotor engagements are paradigmatically concrete and cognitively low level. It should be stressed at the outset that these are *not* independent, interacting systems needing to be fitted together but rather positions toward either end of a continuum, *neither of which can be divorced from the other.*

We agree with Noë (2004), Harnad (Harnad 2007), and many others that all mental content, conceptual or otherwise, must be grounded in specific sensorimotor engagements, and so sensorimotor engagements are partly constitutive of that content. This is consistent with, and should be seen as a refinement of, the classical empiricist tradition that grounds cognition in experience.

For example: it is less accurate to say that we perceive a round plate as round because it projects a round image onto our retina, and more accurate (and useful) to understand it in terms of sensory expectations that change as we move around and interact with the plate. The roundness of the plate is in effect the sum of these expectations. Sensorimotor knowledge is not constructed from simple direct correspondences between "sensory bits" and "motor bits" as in an input-output relationship, but rather is defined in contingent terms (if I do *this*, I expect *that*) that might be called profiles of change.

This already implies a certain abstraction away from particular contexts even at very low levels of cognition; so for example, detection of a visual flow field (surely a more directly sensorimotor relationship than identifying round things) is to discover a profile of change in the visual stream contingent upon certain movements, or activities in the motor stream. As we turn our head the visual image

slides across our retina in a consistent manner, but this relationship is independent of what that image contains; while it can be used to predict the changing activity of individual rods and cones, it remains quite independent from their actual activity. Identifying a profile of change contingent on some action or on some other sensory data (Morse and Ziemke 2009), is not a matter of finding a relationship between specific sensory data and specific motor engagements but already implies an abstraction over a number of specific such engagements.

5.1.1 *Sensorimotor ++*

For all that we enthusiastically support Noë's work, we believe there are problems with his or any account that focuses too narrowly on sensorimotor explanations. The difficulty is how one gets beyond specific sensorimotor engagements: how one generalizes to the sensory-motor profiles needed to explain specific affordances, never mind abstract conceptual thought (i.e., "the problem of extracting reliable categories from experience" (Harnad 1987: 538)). Vittorio Gallese and George Lakoff (Gallese and Lakoff 2005) are quite clear that no further mechanism is needed; the most abstract of concepts is, on any occasion one can name, no more than a specific sensory-motor engagement, with parts of it (e.g., the activation of the motor cortex, with consequent movement) suppressed. We disagree, and propose the unified conceptual space as the additional mechanism. What "pure" sensorimotor accounts cannot do, and (we believe) our account can, is explicate the relationship between instances and classes. By trading on an essential ambiguity between the two — most instances can be treated as classes; all classes can be treated as instances — we believe we can show how one can get from the one to the other, on whatever cognitive level.

As the unified conceptual space theory was presented as an extension of Gärdenfors's work, so the position we would like to take on sensorimotor grounding of cognition can best be taken as an extension of Noë's work. One might be tempted to call it "sensorimotor ++": sensorimotor *plus* somatic and other bodily information (*per* Morse and Ziemke, forthcoming) *plus* (with appropriate qualifications) symbolic/representational language (*per* Chrisley and Parthemore 2007), as located within a conceptual spaces framework.

As on standard associationist accounts, the story begins with pattern recognition, albeit with the caveat that there may well be a minimal pre-existing notion in the system of what patterns are: i.e., proto-concepts, mental structures that are concept-like yet fail to meet one or more of the conditions standardly placed on concepts. Such structures would be (for practical purposes at least) governed by nomic relations rather than experientially derived.

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 as an initial structuring of that agent's unified conceptual space: an initial partitioning (unless that initial partitioning is provided by pre-existing proto-concepts). A minimal perceptual regularity is a mapping of one point in the perceptual stream to another.

The intuition here is that regularities in the regularities, and regularities in *those* regularities, yield increasingly complex, increasingly abstract mental content that, as abstraction progresses, eventually becomes recognizable as first- and higher-order conceptual content and, in the limit, as symbolic content. The result is an associational hierarchy that at its base is strongly associational and at best weakly symbolic, and at its summit is strongly symbolic and at best weakly associational.

As one ascends through the hierarchy, 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 sub-regions within the unified conceptual space. In this manner the unified conceptual space is gradually transformed from a largely unstructured entity to a richly structured one.

If one inverts the associational hierarchy, then what at first looked very much like symbols will shift into iconic representations and 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, and we are left with "bare" interactions.

To return to the earlier definition of concepts and refine it: a concept then is a synchronized pattern of relatively higher-order association between some aspect of the mental world of the agent and some matching aspect of her experienced environment. It is recognizably a concept to the extent that it is abstracted away from the particulars of context, even as it is then applied back to particular contexts. It is both abstracted away from and structurally isomorphic to its referent in perception (which need not necessarily correspond in any way to whatever caused that perception). We will return to and revise this definition once more before we close the paper.

Note that we have talked about salience without giving any account of it. Salience is a whole separate large topic outside the scope of this paper. Nonetheless we find intuitively appealing the attempt by many in the enactivist camp — recently e.g., by Thompson and Stapleton (2009) — to ground minimal salience in the survival of the organism. What is salient is what enables the organism to survive, and other saliences are meant to follow directly or indirectly from there.

5.1.2 *Partitioning the conceptual space*

We can set aside, at least for now, the question of whether an agent's initial conceptual space is a *tabula rasa*, or is minimally pre-configured with a partitioning into proto-conceptual sub-regions that are e.g., physical-object-type things, action-type things, and property-type things. If not preconfigured, that partitioning will shortly follow, for each of the three types of "thing" depends upon the others, indeed is defined as *per* Section 4.1.2 in terms of the others. It is important to bear in mind that the structuring of the conceptual space begins at a point far below what we are inclined to think of as conceptual: not *fully* non-conceptual (otherwise it would not be a conceptual space) but not recognizably conceptual, either. This initial situation is described in Figure 3. Object- and action-type things dominate the space, all of which is compressed below the level of first-order concepts (but above zeroth-order).

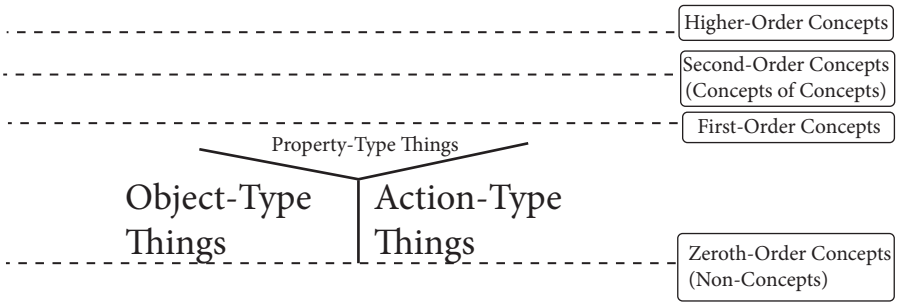

Figure 3. The initial partitioning

As the partitioning proceeds, objects and actions continue to dominate (see Figure 4). Particularly at the proto-conceptual level, the space becomes recognizable as a Voronoi tessellation. (A Voronoi tessellation tiles a plane that is initially populated by a set of points, which in conceptual spaces theory is taken to represent the most prototypical member of a conceptual category. The plane is then divided up

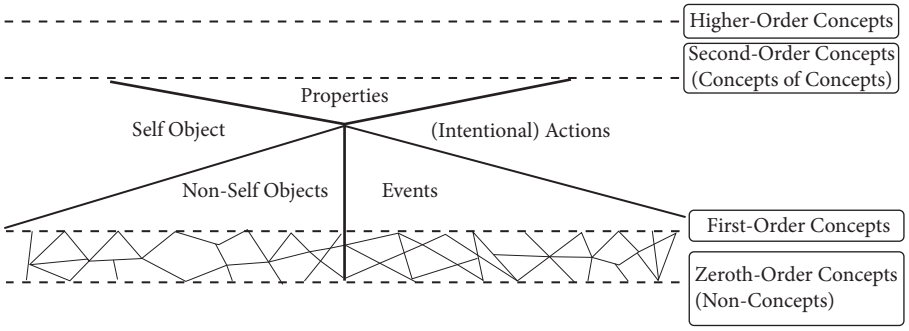

Figure 4. Initial first-order concepts

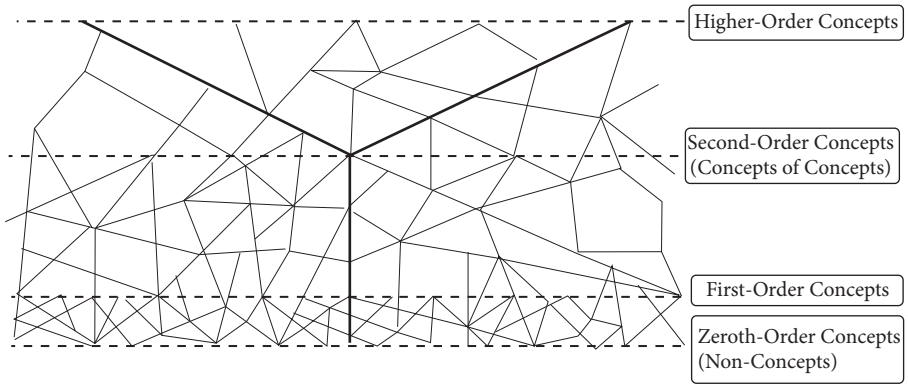

Figure 5. Conceptual space extends into second-order (and beyond)

according to which of those points the remaining points in the plane are closest to. Boundaries arise wherever there is equidistance to two of the existing points, junctions wherever there is equidistance to three or more).

As meta-cognitive abilities develop (i.e., the capacity for thoughts about thoughts), the space expands into the area between second- and higher-order concepts. See Figure 5. Properties take on increasing importance even while sub-partitioning of object and action sub-spaces continues. The space is still largely unstructured, however, and there will be a consequent tendency to over-generalize.

5.1.3 *Mapping space onto space*

The process we have just described, where the conceptual space begins as largely if not entirely un-partitioned and ends up very intricately partitioned, is a movement from the maximally general (*something* is salient) to the maximally specific (the salient thing is e.g., my weight on the scales at 7:03 this morning). But there is a competing perspective, just as valid, whereby concept formation is a movement from the maximally specific (applicable to the narrowest possible range of contexts) to the maximally general (applicable to the widest possible range of contexts). This is because at the same time that the conceptual space is being partitioned, it is becoming more and more structured in another way, as parts of it map onto each other, both through conceptual reference, in those cases where the referent also lies within the conceptual space, and through the mechanism of components (sub-parts), parameters (integral dimensions), and contextuals (Section 4.1.2). These mappings are initially very limited, rendering the proto-concepts and initial concepts applicable only within quite narrow contexts.

So on the one end of the continuum, one has maximally general proto-concepts that are applicable only within quite specific contexts (everything is just a *thing*); at the other end of the continuum, one has maximally specific concepts

(relating only to one particular thing) able to locate their referents in the broadest possible range of contexts.

The description of space mapping onto space given in Section 4.1 was quite high level. How can we conceptualize the process as looking initially?

Given some set of "raw" perceptions — 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 locations, or sudden changes in pixels. Those regularities describe a space of their own, but it's no longer a perceptual space, strictly; rather it is one step removed: a space of regularities, albeit very, very basic ones. Call it a regularity space. The proto-conceptual entities at this level are to be used and very quickly discarded (or recycled).

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.

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: what had been a continuous period of moments *t* is the new minimal, individual unit *u* sampled over some period of duration *v*, as the agent steps further back from the moment and the "moment" becomes larger and larger scale. Likewise one could compare "second-order" spaces of regularities to derive "third-order" spaces, and so on through to *n*th order spaces, limited only by practical boundaries (e.g., available time, energy).

At each stage the vast majority of content is being discarded. At each level, some patterns are being pulled out but infinitely many other possible patterns are being ignored. At one end of the continuum, one is, as it were, drowning in a sea of detail; go too far in the other direction, however — toward the very rarefied, the very abstract — and no useful detail remains.

5.2 Turning things around: Experience emerges from concepts

Of course, if the account of concept acquisition given so far was *all* there was to be said, then concepts really would be static entities, with no means for update or obsolescence. But concepts are dynamic. They have no value, no meaning, unless at the same time they are being acquired they are also being applied. Concepts are at least as much skilful abilities as they are expressible knowledge (cf. Morse and Ziemke 2007).

To borrow a page from the classical definitionists, the concept acquisition account 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.

Again, we begin with some more theoretical grounding before relating this application process back to conceptual spaces and to the unified conceptual space theory in particular.

Unfortunately we do not have, as we had with Noë and the bottom-up-driven process of concept acquisition, a similar guide to the top-down-driven process of concept application. As do many in the enactive camp, we reject as not very useful (at least as a general model for cognition, even viewed top down) the cognitivist input-output based model of cognition exemplified by SMPA: sense-model-plan-act. Although there are, of course, many differences, what cognitivist approaches generally have in common is a splitting apart of sensory input from motor output and a treatment of higher-level cognition as independently explainable from lower-level details of "mere" implementation. That said, they remain surprisingly popular: witness the recent publication of Adrian Torey's book (2009), which ironically is, in many ways, much more traditional than it is revolutionary.

So for this part of the story, we will help ourselves to several diverse guides. The resulting picture is of high-level cognition, conceptually removed from but logically continuous with its sensorimotor roots.

5.2.1 *Concepts as expectations*

With one hand in the past and the other in the future, concepts are the expectations that drive experience. Consider concepts as a tool that, once you have it, you literally cannot imagine doing without. Perhaps concepts are like language in this way. For many people, language is so much a part of their thinking that it seems their thought is *structured as* language. Torey makes language the basis of his account of cognition: no language, no thoughts, and no mind.

Consider conceptualized experience as an emergent projection over top of non-conceptualized 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 (Damasio 2000: 195–233), 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 (Heidegger 1962). Only when the hammer breaks or the nail bends — i.e., only when the hammer fails somehow 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.

This account of concepts as expectations is very reminiscent of Chrisley's Expectation-Based Architecture (EBA) (Chrisley and Parthemore 2007), even though Chrisley grounds those expectations (correctly, we think) in non-conceptual content. It is reminiscent as well of Imogen Dickie's notion of "representation as control" (Dickie 2006), whereby representational (conceptual) expectations based on past experience guide and massively simplify the agent's interaction with its environment. Compare Gärdenfors: "The prime problem is that the information received by the receptors is too rich and too unstructured. What is needed is some way of transforming and organizing the input…" (Gärdenfors 2004: 21).

Not surprisingly, conceptual expectations present several trade-offs. One simplifies in order to understand. But if one over-simplifies, one no longer understands.

The set of patterns potentially discernible in any perceptual context may be infinite (cf. Dennett 1991: 33–35). Any context can be perceived from a bewildering variety of perspectives. Attention appears to be limited by finite resources; there is ample psychological evidence that working memory can attend to only a small number of items at any time. 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, the phenomenon known as 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 (Langham 1999: PAGES will be supplied by author at first proofs).

The lesson more broadly is this: experience teaches us to become more and more selective of what we attend to. A simplified experience of the driving scene, *in most instances*, leads to improved performance. A simplified experience of the world in general may, in many instances, lead to improved performance, even better survival and reproductive opportunities. A simplified world is easier to understand and respond to. But sometimes the simplified model will make mistakes, because the simplified model is *not* the original. Information is lost. The driver pulls out right in front of the motorbike, even though he swears, truthfully enough, that he looked and saw no one.

But beyond all of this, the more conceptual knowledge we have, the more we come to rely on it. As a wealth of further 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; or we see some of what is in front of us and infer the rest. Perception is not independent of reality, nor is it solely constituted by it! Instead of simply revealing the world, concepts-as-representations help to construct it even as they are constructed by it.

Concepts have this dual nature: if their nature is increasing abstract, their application is quite context specific. 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*.

With this revised definition, then, we can begin to see how conceptual spaces theory comes into its own. Just as representations are neither *internal* nor *external*, being relational entities, standing between an agent and her perceived environment; just as concepts are, likewise, *between* the agent and her environment, created out of the dynamic interaction of the agent with her environment; so a theory of concepts properly belongs between associational and representational accounts of mental content, showing how conceptual mental content arises from the dynamic interaction of symbolically interpretable structures with associations, the dynamic interaction of the cognitively abstract with the sensorimotorly concrete.

5.2.2 *The conceptual space in use: Conceptual space mapping onto perceptual space*

Structuring the conceptual space was described as a bottom-up, layer-by-layer hierarchical process of pattern recognition, finding patterns in patterns, patterns in patterns of patterns, and so on. Using the conceptual space then is a top-down-driven, layer-by-layer process of pattern matching, a return down through levels of the hierarchy, toward particular encounters and toward parts as opposed to wholes. Call it "de-layering". If an *x* violates expectations, consider previous experiences with similar *x*s or *x*-like things, or decompose the *x* into e.g., its functional parts.

For concept acquisition, concepts looked more like abilities. Associations and association building were in the driver's seat. For concept application, concepts look more like representations. Initially, at least, symbols and symbol application may be a more appropriate level of description (though only some part of this need be consciously articulable).

The basic idea is this: conceptual space can be compared to perceptual space, concepts in the conceptual space matched against their non-conceptual analogs in present experience, attempting the most specific match possible: a *particular* concept instance to a *particular* sub-region of present experience. If a match cannot be confirmed at the most specific levels of the conceptual space, one can, as it were, relax the resolution, attempting to match against larger and larger portions of the conceptual space, going more and more general until a match finally does occur. (At some point a match *must* occur, since everything perceivable is, minimally, a *something*).

Breakdown of varying degrees occurs anytime there is a need to reduce the resolution, because the highest resolution of the conceptual space fails to match. There are three strategies available for dealing with breakdowns, which we can order from least to most radical.

1. Adjust the *logical structure* of the closest matching concept (Section 4.1.2) in terms of its parameters, contextuals and, if appropriate, components, so that a match now does occur. Formerly you conceptualized all swans as white; now you conceptualize swans as either white or black.
2. Partition an as-yet-unpartitioned sector of the conceptual space. Formerly you recognized only a general category of swans, all of which were white. Now you recognize two sub-categories of swans: white swans and black swans.
3. Most radically, *remove partitioning* from some sector of the conceptual space and re-partition it: i.e., divide it into a different set of shapes (Section 4.1.1). As opposed to conceptual change via 1 or 2, this is conceptual obsolescence and replacement. This would be if your encounter with black swans forced you to e.g.. re-structure your whole "bird" space.

5.2.3 *An example: On encountering a door*

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 (which is to say, the maximal resolution of the conceptual space) 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 may 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 sensorimotor 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 yet 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.

In the process, you may come to change your understanding of doors a little; or you may derive a concept of a new kind of door. You might re-structure your entire "door" space. 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 sensorimotor engagement with it!

6. The debate revisited

If one may borrow a line from Mark Twain, the death of representations has been greatly exaggerated. Concepts are more than representations, but representations are *unavoidably, for us* part of what it is to be a concept — for when an agent uses a concept reflectively she is using it to represent *something* to *someone*, be it herself or another agent. As David Kirsh writes (in response to Brooks 1991b), "There is a limit… to how far a creature without concepts can go… Concepts are either necessary for certain types of perception, learning, and control, or they make those processes computationally simpler. Once a creature has concepts its capacities are vastly multiplied" (Kirsh 1991: 191). Again, he writes, "This capacity to predicate is absolutely central to concept-using creatures. It means that the creature is able to identify the common property which two or more objects share and to entertain the possibility that other objects also possess that property" (ibid.: 163).

What representations gain us, argues Richard Shusterman in response to anti-representationalist Merleau-Ponty, is the capacity for explicit reflection and the consequent ability to recognize and to modify bad habits. "…In order to effect… improvement, the unreflective action or habit must be brought into conscious critical reflection (if only for a limited time) so that it can be grasped and worked on more precisely" (Shusterman 2008: 63). He could as well be responding to Brooks when he says, "The claim that we can do something effectively *without* explicit or representational consciousness does not imply that we cannot also do it *with* such consciousness and that such consciousness cannot improve our performance" (Shusterman 2008: 68).

We are under no illusions that the debate over representations will end with our paper. But we can and do hope that this paper can be part of an emerging re-conceptualizing of representations the better to understand both their power and their limitations.

7. Discussion and conclusions

This paper has sought to provide, from a perspective within enactive philosophy, an account of a continuum between basic sensorimotor engagements and abstract mental content. On the one hand, we can never stop being symbol users viewing the world representationally. On the other hand, we require means to talk about contexts that we cannot directly observe but that we logically conclude must exist: contexts in which, for lack of a homunculus, there *is* no agent representing anything to anyone. This includes not only what must be happening at our lowest levels of cognition but also what is happening at quite abstract levels of cognition when we employ concepts non-reflectively.

An enactive approach sees symbolic (representational) and associationist (non-representational) accounts, suitably framed, 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 as a way of bridging the apparent gap between associationist and symbolic accounts. We have taken his conceptual spaces theory (Section 3) and extended it with our presentation of the unified conceptual space theory (Section 4), which attempts to show how all our many different conceptual spaces come together within a single unified space.

These ideas then are pressed into use in the discussion of co-emergence (Section 5). Any embodied theory of concept acquisition probably needs to start from a discussion of sensorimotor engagement (5.1). Any comprehensive theory of concept application needs to take representations seriously (5.2). Section 5 shows the compatibility between these very different looking perspectives and so connects bottom-up with top-down understanding of cognition. Either perspective, on its own, will be incomplete.

The account remains incomplete in many areas: both the unified conceptual space theory and the account of co-emergence. Notably missing from the discussion of concept acquisition is the question of salience: of all the possible patterns that could be derived from the perceptual stream, why do only a few get derived and most do not? Missing from the discussion of concept application is the question of just how far down the cognitive scale do our conceptual expectations reach. There is evidence that, in visual processing at least, this is surprisingly far. As Paul Hemeren notes in his recent PhD thesis, and offers results to support, "Humans appear to have early access to semantic level information in the form of basic level object categorization" (Hemeren 2008: 150).

Sitting between an association-based perspective most suitable to low-level cognition and a symbol-based perspective most suitable to high-level cognition, a concept-based perspective shows how concepts can be seen one moment as as-

sociations — pattern-recognition abilities — and another as representations or symbols. Like symbols or representations, concepts apply to specific contexts at the same time they abstract away from the particulars of context. This generalization away from the particular is essential to a certain class of autonomous agents in meeting changing circumstances and environments, giving them the ability to step back and take a wider view. To the extent that agents can do this "stepping back", they can be acknowledged as concept users; if they are aware of and able to reflect upon their use of concepts — to observe that they are observing — then they are probably symbol users as well.

Notes

1. Some, of course, would object to our use of the term at all, either as being vacuous or overly vague. We believe we have defined it well enough here for it clearly to be neither, on our usage. At the same time, we use the term guardedly, for it is both a term of fashion and a "red flag" word that, for many people, sets off immediate emotional reactions.

2. An earlier version of this paper referred to the mutual scaffolding of concepts and experience. The problem with scaffolding, of course, is that it is meant to be removed (or at least be removable) once the structure is in place, as noted in the closing paragraphs of Section 3. "Co-emergence" captures our intent much better.

References

Anderson, J. and Lebiere, C. 1998. *The Atomic Components of Thought*. Mahwah, NJ: Lawrence Erlbaum Associates.

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

Brooks, R.A. 1991a. "Intelligence without reason". In *Proceedings, IJCAI-91*. San Francisco, CA: Morgan Kaufmann, 1–21.

Brooks, R.A. 1991b. "Intelligence without representation". *Artificial Intelligence* 47: 139–159.

Chalmers, D.J. 1990. "Why Fodor and Pylyshyn were wrong: The simplest refutation". In *Program of the Twelfth Annual Conference of the Cognitive Science Society*. Hillsdale, NJ: Lawrence Erlbaum Associates, 340–347.

Chella, A., Coradeschi, S., Frixione, M., and Saffioti, A. 2004. "Perceptual anchoring via conceptual spaces". *Prooceedings of the AAAI-04 Workshop on Anchoring Symbols to Sensor Data*, Menlo Park, CA: AAAI Press, 40–45.

Chella, A., Frixione, M., and Gaglio, S. 2000. "Understanding dynamic scenes". *Artificial Intelligence* 123: 89–132.

Chella, A., Frixione, M., and Gaglio, S. 2008. "A cognitive architecture for robot self-consciousness". *Artificial Intelligence in Medicine* 44(2): 147–154.

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.

Damasio, A. 2000. *The Feeling of What Happens: Body, Emotion and the Making of Consciousness*. London: Vintage.

Dennett, D.C. 1991. "Real patterns". *Journal of Philosophy* 88(1): 27–51.

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

Evans, G. 1982. *Varieties of Reference*. Oxford: Clarendon Press.

Fodor, J.A. 1975. *The Language of Thought*. New York, NY: Thomas Y. Crowell.

Fodor, J.A. 2008. *LOT 2: The Language of Thought Revisited*. Oxford: Clarendon Press.

Fodor, J.A. and Pylyshyn, Z.W. 1988. "Connectionism and cognitive architecture: A critical analysis". In Pinker, S. and Mehler, J. (eds|), *Connections and Symbols*. Cambridge, MA: The MIT Press, 3–72.

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.

Geuder, W. and Weisgerber, M. 2002. "Verbs in conceptual space". In G. Katz, S. Reinhard, and P. Reuter (eds), *Sinn und Bedeutung 6, Proceedings of the Sixth Meeting of the Gesellschaft für Semantik, Osnabrück, October 2001*. Osnabrück: Universität Osnabrück, 69–84.

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

Gärdenfors, P. and Williams, M.-A. 2001. "Reasoning about categories in conceptual spaces". In *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence*. San Francisco, CA: Morgan Kaufmann, 385–392.

Grush, R. and Churchland, P. 1995. "Gaps in penrose's toilings". In T. Metzinger (ed), *Conscious Experience*. Exeter: Imprint Academic, 185–214.

Harnad, S. 1987. *Category Perception: The Groundwork of Cognition*. Cambridge: Cambridge University Press [Chapter "Category Induction and Representation", 535–565].

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

Harnad, S. 2007. "From knowing how to knowing that: Acquiring categories by word of mouth". Presented at Kaziemierz Naturalized Epistemology Workshop (KNEW), Kaziemierz, Poland, 2 September 2007.

Harnad, S. 2009. "Concepts: The very idea". Presented at the Canadian Philosophical Association Symposium on Machery's *Doing without Concepts*. Carleton University, Canada, 27 May 2009.

Harvey, I. 1992. "Untimed and misrepresented: Connectionism and the computer metaphor". Cognitive Science Research Paper 245, Brighton, UK: University of Sussex.

Heidegger, M. 1962. *Being and Time*. J. Macquarrie and E. Robinson (trans). Oxford: Blackwell.

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

Hemeren, P. 2008. *Mind in Action: Action Representation and the Perception of Biological Motion*. PhD thesis, University of Lund.

Kirsh, D. 1991. "Today the earwig, tomorrow man?". *Artificial Intelligence* 47: 161–184.

Langham, M. 1999. *An investigation of the role of vehicle conspicuity in the 'looked but failed to see' error in driving*. DPhil thesis, University of Sussex.

Machery, E. 2009. *Doing Without Concepts*. New York: Oxford University Press.

Maturana, H.R. and Varela, F.J. 1992. *The Tree of Knowledge*. Boston, MA: Shambhala.

Morse, A. and Ziemke, T. 2007. "Cognitive robotics, enactive perception, and learning in the real world". In *CogSci 2007 — The 29th Annual Conference of the Cognitive Science Society*. New York: Erlbaum, 485–490.
Morse, A. and Ziemke, T. Forthcoming. "The somatic sensory hypothesis".
Newell, A. 1980. "Physical symbol systems". *Cognitive Science* 4(2): 135–183.
Newell, A. and Simon, H.A. 1979. *Human Problem Solving*. Upper Saddle River, NJ: Prentice-Hall.
Noë, A. 2004. *Action in Perception*. Cambridge, MA: The MIT Press.
Noë, A. 2007. "Magic realism and the limits of intelligibility: What makes us conscious?". *Philosophical Perspectives* 21: 457–474.
Parthemore, J. 1990. *Charley: A creative, collaborative short-story writing environment*. MSc thesis, University of Sussex.
Perry, J. 1986. "Thought without representation". *Proceedings of the Aristotelian Society* 60: 137–151.
Prinz, J. 2004. *Furnishing the Mind: Concepts and Their Perceptual Basis*. Cambridge, MA: The MIT Press.
Rupert, R.D. 2009. *Cognitive Systems and the Extended Mind*. Oxford: Oxford University Press.
Shusterman, R. 2008. *Body Consciousness: A Philosophy of Mindfulness and Somaesthetics*. Cambridge: Cambridge University 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*. Cambridge, MA: Harvard University Press.
Thompson, E. and Stapleton, M. 2009. "Making sense of sense-making: Reflections on enactive and extended mind theories". *Topoi* 28(1): 23–30.
Torey, Z. 2009. *The Crucible of Consciousness: An Integrated Theory of Mind and Brain*. Cambridge, MA: The MIT Press.
Varela, F.J., Thompson, E., and Rosch, E. 1991. *The Embodied Mind: Cognitive Science and Human Experience*. Cambridge, MA: The MIT Press.
Winograd, T. and Flores, C. 1986. *Understanding Computers and Cognition: A New Foundation for Design*. Norwood, NJ: Ablex.
Wittgenstein, L. 2001. *Philosophical Investigations*. G.E.M. Anscombe (trans). Oxford: Blackwell.

Authors' addresses

Joel Parthemore
Centre for Research in Cognitive Science
School of Informatics
University of Sussex
Falmer, Brighton BN1 9QH
UK

J.E.Parthemore@sussex.ac.uk

Tony F. Morse
Centre for Robotics and Neural Systems
University of Plymouth
Portland Square B110
Plymouth, Devon PL4 8AA
UK

mora@iki.his.se

About the authors

Joel Parthemore is a DPhil student at the University of Sussex, UK, and a visiting research student at the University of Lund, Sweden. His main research interests are in using enactive theories of concepts to bridge some of the major divides in contemporary theories of concepts and to capture the continuity of the agent with the agent's environment. He is interested as well in the limitations of conceptual knowledge and the relationship between conceptual and non-conceptual mental content. He received his MSc in Knowledge-Based Systems from the University of Sussex and his BSc in Journalism from Northwestern University in the US. His recent publications include "Synthetic phenomenology: Exploiting embodiment to specify the non-conceptual content of visual experience" (2007, with R. Chrisley).

Anthony F. Morse is a Senior Research Scientist at the University of Plymouth, UK. His current research focus is on sensorimotor based developmental robotics and biases in categorization. He worked as a post-doc on the ICEA project (www.iceaproject.eu) in at the University of Skovde, in Sweden following receipt of his DPhil in Cognitive Science and his MSc in Evolutionary and Adaptive Systems from the University of Sussex. He is currently working on ITALK, a European project on cognitive robotics (www.italk.eu). His recent publications include "On the Role(s) of Modelling in Cognitive Science" (2008, with T. Ziemke, *Pragmatics & Cognition*, *16*(1), 37–57) and "Manipulating Space: Transient Dynamics and Inattentional Blindness" (In Press, with R. Lowe, and T. Ziemke, *Connection Science*).