Beyond the Flesh: Some Lessons from a Mole Cricket

Andy Clark
Cognitive Science Program
Indiana University
Bloomington
1033 East Third Street
Sycamore 117
Bloomington, IN 47405
andy@indiana.edu

Abstract What do linguistic symbols do for minds like ours, and how (if at all) can basic embodied, dynamical, and situated approaches do justice to high-level human thought and reason? These two questions are best addressed together, since our answers to the first may inform the second. The key move in scaling up simple embodied cognitive science is, I argue, to take very seriously the potent role of human-built structures in transforming the spaces of human learning and reason. In particular, in this article I look at a range of cases involving what I dub *surrogate situations*. Here, we actively create restricted artificial environments that allow us to deploy basic perception-action-reason routines in the absence of their proper objects. Examples include the use of real-world models, diagrams, and other concrete external symbols to support dense looping interactions with a variety of stable external structures that stand in for the absent states of affairs. Language itself, I finally suggest, is the most potent and fundamental form of such surrogacy. Words are both cheap stand-ins for gross behavioral outcomes, and the concrete objects that structure new spaces for basic forms of learning and reason. A good hard look at surrogate situatedness thus turns the standard skeptical challenge on its head. But it raises important questions concerning what really matters about these new approaches, and it helps focus what I see as the major challenge for the future: how, in detail, to conceptualize the role of symbols (both internal and external) in dynamical cognitive processes.

Keywords
Embodiment, symbols, situated reason, surrogate situations, mole cricket, singing burrows, language, attention

1 Introduction

Dynamics, embodiment, and situatedness are widely advertised as having radical implications for the sciences of the mind [49, 42, 26, 47, 48, 45]. Skeptics such as Clark and Toribio [13] have pointed out that the most radical such implications seem confined to a small class of cases. These are cases involving the dense, time-pressured, coupled unfolding of a behavior, typically regulated by an ongoing perceptual link with some external goings-on—reaching for a visually presented object, returning a tennis serve, and so on. But such cases, the skeptic insists, are cognitively marginal. The proper explanatory targets for the sciences of genuinely mental activity, they suggest, involve thought and reason that target distal, absent, highly abstract, or even impossible states of affairs. Good explanatory frameworks for the highly coupled cases look less promising, it is claimed, in these more rarified arenas.

I examine this argument and find it less compelling than I once believed. In particular, it is flawed by a conflation of two distinct properties. The first is the property of disengagement (reason operating in the absence of its ultimate target, as when we think of that which is not close to hand). The second is the property of decontextualization (reason operating without the kinds of dense, perceptually saturated local couplings that most obviously reward treatment in dynamical and situated terms). I shall argue that high-level reason is local and contextualized even when (and indeed most strongly when) it is disengaged. The most obvious examples are when we use real-world models, diagrams, and other concrete external symbols to create conditions of (what I shall call) surrogate situatedness. In such cases, we reason about what is not at hand by means of dense looping interactions with a variety of stable external structures that stand in for absent states of affairs.

Language itself, I shall finally suggest, is the most potent and fundamental form of such surrogacy. Words are both cheap stand-ins for gross behavioral outcomes, and the concrete objects that structure new spaces for basic forms of learning and reason. Language is thus conceived as primarily a form of environmental structuring rather than an information stream requiring translation into and out of various inner codes.

A good hard look at surrogate situatedness thus turns the standard skeptical challenge on its head. But it raises important questions concerning what really matters about these new approaches (in particular, it undermines a certain view of the role of simple temporal constraints). And it helps focus what I see as the major challenge for the future, viz., how, in detail, to conceptualize the role of symbols (both internal and external) in cognitive processes.

2 From Coordinated Rhythmic Motion to Foreign Policy

It seems fair to say that most (though not all) of the really compelling accounts to emerge from the stables of dynamical and embodied cognitive science depict cases of densely coupled unfolding. By this I simply mean that they involve the use of a perceptuo-motor routine whose operation exploits the continuing presence of some tangible target. The simplest example might be a wall-following or phototropic robot. But more impressive demonstrations include the robot cricket that identifies and locomotes towards the call of its mate [50], the account of how baseball outfielders run to catch a fly ball [30], the work on hopping robots [36], the detailed dynamical model of the production of rhythmic finger motions [23], the recent explosion of work on animate vision, just-in-time sensing, the use of deictic pointers [1], and the growing body of robotics work reviewed in [35].

In cases such as these (and there are many others) we confront a characteristic mix of constraints and opportunities, which I'll label the *basic signature*. The basic signature involves a task that requires the agent to keep track of a situation unfolding in some constraining (absolute) time frame, so that real timing (not just sequence) is essential to success. And it involves the use, to accomplish the task, of body, motion, and world as integral aspects of the problem solution—for example, the use of head and eye motion and just-in-time sensing to retrieve information from the visual scene, thus (as Rodney Brooks put it) "using the world as its own best model."

This signature mix of constraints and opportunities looks to be lacking in many cases of high-level human problem solving. We can plan next year's family vacation, design a new 100-story building, or sit down and think about equality, freedom, and the bad effects of the "war on terrorism." In such cases we are forced to think and reason in the absence of the target situation. The vacation is not until next year. The building does not exist (and may even be impossible). Equality and freedom are not out there for simple perceptual reencounter as and when required. Instead, we seem forced into a mode of "offline reasoning." The tools, principles, and strategies that work so well for coordinated rhythmic finger motion and its ilk may falter, it is feared, in the face of such new demands and challenges.

There are two standard replies to this worry. The first argues that any hard-and-fast online-offline distinction is itself unclear and potentially misleading. Rather, in just about all cases, we find elements of each and a constant seamless integration of the two. The need to integrate real-time action with ongoing planning and reason is itself, it is argued, a reason to prefer a unified dynamical approach that treats perception, reason, and action in essentially the same terms.

For a powerful version of this argument, see [43]. Randy Beer's [3] experiments (and more recently, those of Slocum [38]) with "minimally cognitive agents" may also be seen as partial support for this class of response.

The second response is to argue that even imagination-based, problem-situation-decoupled reason is fully continuous with the other cases, because our imaginative routines are themselves body-based, exploit egocentric coordinate spaces, (re)deploy the same perceptuo-action-oriented inner states, and so on. For various versions of this, see [43, 29, 2, 26].

Both these responses are useful and important. But there is a third move available that may, I think, prove more fundamental than either standard reply. To introduce it, let's take a brief detour into the surprising acoustic world of the singing cricket.

3 Singing Burrows

Crickets sing to attract mates of their own species, and where these songs are concerned, louder is generally better. A louder song reaches more potential mates, and is typically deemed more attractive than any weaker-sounding competitor. The onboard mechanism responsible for producing the sound is a patch of flexible wing membrane, known as the harp, that is pulled taut and struck using a kind of tooth-and-peg arrangement. The overall process is termed *stridulation*.

There is, however, a problem. Crickets are quite small creatures, and their onboard harps are tiny relative to the wavelength of their carrier tones. This is very inefficient in a free (unobstructed) sound field, and much of their muscle energy is wasted, not turned into sound but lost doing capacitive and inertial work. The typical efficiency achieved by many cricket species is a paltry 1.5 to 2% (of expended pinging energy turned into sound). The biologist J. Scott Turner [46] offers a detailed and compelling analysis of the strategies used, by a few cricket species, to make more efficient use of this muscle energy.

One strategy is to first bite a hole in a leaf, and then position the harp over the hole, which then acts as a baffle, making the sound 2 to 3 times louder and allowing it to project to a volume 40 times greater. Starting with the basic physics of sound emission, however, Turner shows that the ideal (but impossible) solution would be to enclose the sound emitter in an infinite horn. This would preclude all loss of energy to inertial and capacitive work (you can't accelerate an infinite mass of air and it can't store energy to do other work), forcing all the energy to turn into sound. Infinite horns, like true universal Turing machines, do not exist. But a good approximation is the so-called Klipsch horn. This is a horn that flares like a trumpet (an exponential horn) with a capacious end bulb (corresponding to the musician's mouth cavity) connected by a restricted opening (such as the lips or a reed) to the horn. This setup allows the air inside the horn to be turned back to help drive the emitter via a kind of turbocharging called *loading* the horn, and it allows the trapped air to resonate at a frequency related to the length of the horn. In essence, the sound field is obstructed in a good way, vastly reducing the amount of energy lost to capacitive and inertial work.

Amazingly, several species of mole cricket (especially *Scapteriscus acletus* and *Gryllotalpa vinae*) exploit physically perfect Klipsch horns as a means of efficient sound production. These crickets tweak their own burrow-building activities so that one tunnel takes the form of a large bulb, while another flares like an exponential horn. The cricket sits in a narrow constriction between these two animal-built structures and stridulates (sings). In this way the burrow acts as a tuned impedance transformer carefully fitted to the specific carrier wave of the species. Sometimes, a single cricket will sit at the intersection of two such tunnel-bulb arrangements. With the ground itself acting as a very large (near-infinite) baffle, and the boost of a double exponential horn, these crickets obtain a remarkable increase in acoustic efficiency. In the case of *Gryllotalpa vinae* a full 34% of the muscle power used for stridulation is now turned into sound. Compare this with a maximum of 5% for an unaided field cricket, though even that is admirable in that high end audio loudspeakers, as Turner points out, manage only a meager conversion efficiency of about 2% (of power into sound). The souped-up mole cricket, however, produces one of the loudest sounds made by any animal. The sound can be heard at 600 m, while the ground itself vibrates over a 20 cm radius.

Turner argues that the mole cricket's burrows, like several other animal-built structures that he considers in the same monograph, are functionally equivalent to the possession of a near-perfect physiological sound-emitting organ. Turner's larger goal, in fact, is to argue that for many biological purposes such animal-built structures should be counted as proper parts of the animal's physiology and thus to breach what he describes as "the essentially arbitrary boundary between organisms and the environment" [46].

Notice, however, a clear difference in evolutionary costs. To create the singing burrows, all that needs to be canalized by the evolutionary process is the reliable emergence of a strategy for feedback-tuned excavation and tunneling. The male mole cricket digs, settles into the constriction, and emits a special kind of chirp (a test chirp). He then repeatedly alters the size and shape of the burrow until the right resonant frequency is heard. Finding this feedback-driven behavioral strategy and loading it into the genome, Turner argues, was plausibly much faster and cheaper than making the long sequence of small changes needed to install a Klipsch horn as a biological-physiological organ.

The mole cricket scenario is simply a comparative springboard: a dramatic device to bring home the staggering transformative potential of our own *cognitive singing burrows*. The dramatic device is helpful in that, despite our familiarity with the notion of human technologies as cognition amplifiers [19, 15], the ubiquitous presence of these amplifiers (pen, paper, models, words, numbers, blueprints, compasses) often blinds us to the depth and importance of their role in distinctively human thought. By the close of the present treatment we will also see many fundamental respects in which our own cognitive singing burrows far exceed (in the cognitive realm) the transformative impact of the simple *acoustic* singing burrow. We thus better appreciate the deeply hybrid (biological and artifactual) nature of the human mind.

To anticipate one such case, notice immediately that certain cognitive singing burrows (such as the use of external media to store new ideas) take even more pressure off the genome than they did in the acoustic case. Now, even the *strategies needed* to build specific external structures (such as paper mills and factories) can be preserved and transmitted by nonbiological means. Moreover (and this is especially important in the present context), many of our cognitive singing burrows function precisely so as to relax specific constraints on real-time, engaged and embodied response (see Section 4 below). The question of whether human cognition is best understood as "embodied all the way up" is thus revealed as importantly ambiguous, as are claims of "seamless continuity" between basic animal and infant minds and our own.

4 Surrogate Situations

The most basic form of cognitive singing burrow, I want to argue, is the use of *surrogate situations*. By a surrogate situation I mean any kind of real-world structure that is used to stand in for, or take the place of, some aspect of some target situation. By a target situation I mean an actual, possible, or at least superficially possible real-world event or structure that is the ultimate object of my cognitive endeavor. For example, suppose I use a dotted line, or a small stick, to indicate, on a rough drawing, the proposed location of a strut on a barn door that I am about to build. The target situation is the (as yet nonexistent) door, and the surrogate situation is the context provided by the drawing (and the stick, if I am using one).

Real-world processes of design, as Henrik Gedenryd [18] has argued in great detail, are marked by multiple complementary uses of surrogate situations. Examining cases as diverse as designing a building or laying out a magazine cover, Gedenryd details the different uses of sketches, prototypes, thumbnails, storyboards, and scenarios, to name but a few. What these all have in common, of course, is that they allow human reason to be disengaged (to reach out to that which is absent or distant or otherwise unavailable) while at the same time providing a concrete arena in which to deploy perceptuo-motor routines of a fundamentally world-engaging kind. In such cases, human reason is, we may say, disengaged but not disembodied.

While it is no doubt obvious enough that we often rely on such a strategy, its pervasiveness, variety, and importance are easily overlooked. Considered in the light of the typical worries about scaling up the embodied approach to higher human cognition (Section 2 above), such tactics are revealing indeed. The mockups (etc.) serve no primary purpose other than that of allowing human reason to get a grip on what might otherwise prove elusive or impossible to hold in mind. Any given project will often rely on the use of multiple kinds of surrogate situation, each of which highlights or makes available some specific dimension of what Gedenryd calls "the future situation of use." In this way, surrogate situations are not simply miniature versions of the real thing. Rather, they are selected so as to allow us to engage specific, and often quite abstract, aspects of the future situation of use. For example, a 4 ft eye-level simulation of a walk through a new living and teaching space may be selected to address the need to develop a "safe and inspiring environment for 4–6 year olds" [18]. In a similar fashion, page layout designers use very rough thumbnails to work out potential relations between graphic and textual objects, and are explicitly counseled to omit distracting detail.

In fact, despite certain recent educational trends that overemphasize the creation of rich and realistic teaching contexts, it may well be that our fluency with surrogate situations depends, to a certain extent, on actively keeping the level of nonessential detail quite low. Judy DeLoache, in a 1991 [16] study of symbolic functioning in very young children, found that the more realistic the surrogate situation, the harder it became for children to use it as a tool to understand a target problem. In a series of experiments, DeLoache and her coworkers had children (2–3 years old) watch as a model toy was hidden under a model piece of furniture. They were then told that the real toy was hidden in the same place in the real room. Despite ensuring that the children understood the instructions, the experimenters found that more realistic models actually degraded performance. For example, children performed better when shown a 2D picture rather than a 3D scale model. And with the scale model, performance could be improved by placing it behind glass, thus impeding full physical interaction. DeLoache suggests this is because as the mock-ups' physical properties become more salient and afford a wider range of interactions, the child's ability to use it as a symbol decreases. Related findings in adult performance include Markman and Genter's [28] demonstration that grasp of abstract relations improves when the richness of the representation of the related objects is decreased, and Goldstone and Sakamoto's [20] demonstration that transfer of learning of abstract principles is often better when concrete details (and hence many superficial similarities) are suppressed in the original learning situation. One important lesson is thus that surrogate situations should be purpose-built to serve specific cognitive needs, and that any pretheoretical commitment to maximal detail and realism is premature and may prove counterproductive.

Nonetheless, the fact that the surrogate situation provides *some* leverage for real-world action and intervention is crucial. In cases such as we have considered, the one thing the agent *cannot* do is to "use the world as its own best model." A nonexistent building cannot act as its own best model, nor can a (merely) proposed route for a new road. Instead, in such cases, the agent must *let a model serve as its own best microworld*. In so doing, she creates an arena in which many basic perceptuo-motor strategies (such as just-in-time sensing and binding) can be redeployed, while at the same time constraining the kinds of thing that can and cannot be done and relaxing some of the most potent temporal constraints on normal world-engaged thought and action.

Much attention has been paid in the literature to the way good models, aids, and mockups constrain action in productive ways (see, e.g., work on graphical representations, such as [37, 14]). But equally important, though less noticed and much harder to study, is the way such stand-ins alter and relax the temporal constraints on ordinary performance. Many of the classic cases of environmentally coupled perceptuo-motor action cited by dynamicists involve nonnegotiable temporal restrictions: The tennis player must deal with a serve whose speed is determined by the opponent, the diving plover must penetrate the water while catching the fish, and the coordinated finger movers must attain a certain speed of rhythmic motion. By contrast, routines applied to a model or mock-up seem relatively unconstrained by fine demands of timing. It really doesn't matter precisely how fast I move the model car into position, or how slowly I draw the lines in the sand describing a new football strategy. In one whole, and very important, class of cases, I actively create, as I go along, the very structures to which I then respond. I am here thinking of cases where I create my own model or mock-up, by writing symbols on a page or laying items out in front of me. In these cases, the agent is in charge of the temporal generation of the items that then structure and guide her own problem solving. Evidently, there is much scope in such cases for an agent to time the production of new symbols or stand-ins in ways that are maximally conducive to her own success.

It is possible (though this is sheer speculation) that this temporal relaxation allowed us to begin to deploy a phylogenetically more recent set of neural cognitive strategies: ones that make richer contact with episodic memory systems and explicit stored knowledge, and are known to be major players in time-delayed as well as imagination-based responses. We have learned, for example, that time-delayed and imagination-based responses, at least in visual processing, depend on distinctive activity in the so-called *ventral visual processing stream* [31, 9]. Surrogate situations may thus provide a kind of evolutionary and developmental halfway house between fully offline imaginative modes of thought and reason and the more time-constrained domains of normal real-world response.

It will help to be as explicit as possible about the time-scale conjectures here. The basic, older circuitry would (in this model) be involved in slow learning by repeated experience, but would (once trained) be capable of supporting extremely rapid, fluent responses in the normal range of situations. The newer circuitry, though slower in actual execution, would be capable of supporting effortful responses to novel incoming information, perhaps by the construction of temporary action-guiding representations. This circuitry would also be in contact with episodic and short-term memory. The kind of picture I have in mind is thus very similar to that developed in some detail by connectionist modelers such as [32, 33].

In sum, there exist clear qualitative differences between surrogate situations and normal situations, both in terms of content (the surrogates profit from idealization, abstraction, and the omission of much concrete detail) and in terms of timing (surrogate situations help relax some of the temporal constraints on normal real-world response). These differences may have favored the gradual coevolution, both through adaptation and through learning, of novel neural and cognitive problem-solving strategies.

5 Words as Anchors in the Sea of Thought

Words and symbols, whether spoken or written, are human-built structures that (like the mole cricket's burrows) alter and amplify our powers of thought and reason. By treating words and symbols as cognition-transforming elements of external scaffolding, we open up new and productive ways of accommodating linguaform reason in a broadly embodied, dynamical framework.

In a fascinating recent study, Hermer-Vazquez, Spelke, and Katsnelson [23] speculate that knowledge of words may enable us to combine otherwise encapsulated mental resources into unified bodies of knowledge. More specifically, they claim that:

> Humans' flexible spatial memory depends on the ability to combine diverse information sources rapidly into unitary representations and that . . . in turn, depends on natural language [23].

Evidence for this claim comes from studies in which rats and prelinguistic infants were shown the location of a toy or food in a room, then disoriented and left to try to find the desired item. The location was determinable by remembering cues concerning color, geometry, scent, and the like. But the rooms were designed so that the geometric cues (e.g., length of wall) were insufficient, and would yield an unambiguous result only when combined with other information (e.g., scent or color of the wall). Rats, as Cheng [7] has shown, stick with the geometric information, searching randomly in each of the two geometrically indistinguishable sites. Yet the rats are capable of using other cues, such as color and scent, in other tasks. Cheng concluded that the rat uses an insulated geometric-information-based module for the navigational task. Hermer and Spelke [21, 22] reproduced this pattern of results with prelinguistic infants, who likewise relied solely on the geometric information, despite the apparent availability of other cues. Yet adults and older children are easily capable of combining the various (geometric and nongeometric) cues to solve the problem. Success at combining the cues was not predicted by any measure of the children's intelligence or developmental stage except for the child's use of language. Children who were able to spontaneously conjoin spatial and (e.g.) color terms in their speech (who would describe something as, say, to the right of the long green wall) were able to solve the problem. Those not displaying this pattern of word use were unable to outperform the rats.

In an elegant 1999 study, Hermer-Vazquez et al. [23] probed the possible role of language in this task by asking subjects to solve problems requiring the integration of geometric and nongeometric information while performing one of two other tasks. One task involved shadowing (repeating back) speech played over headphones. The other involved shadowing, with their hands, a rhythm played over the headphones. The working memory demands of the latter task were at least as heavy as those of the former. Subjects engaged in speech shadowing were unable to solve the integration-demanding problem, while those shadowing rhythm were unaffected. This result (alongside other experiments designed to control for other possible explanations) suggests to the researchers that language is actively involved in our ability to solve the problem requiring the integration of geometric and nongeometric information. More generally, it might even seem to suggest (see especially [6]) that language is the unique medium (at least in humans) for the cross-modular integration of information. Perceptually encountered or recalled symbols and sentences act, according to Carruthers, like inner data structures, replete with slots and apt for genuine inner combinatoric action. This combinatoric action allows information from otherwise encapsulated modules to enter into a unified inner representation. This latter, rather sweeping claim finds some support in a number of other studies discussed by Carruthers, including work on human mathematical abilities [40]. But I would like to offer an alternative interpretation, one that makes contact with my story about the use of surrogate situations.

In previous work [8, 10] I have discussed the role of tags and labels in making available forms of higher-order problem solving. Chimps trained to use concrete tags (e.g., a red plastic square) for the relations of sameness and difference are able (unlike non-tag-trained chimps) to learn to solve second-order problems requiring them to judge of two pairs of objects whether the relations exhibited are themselves the same or different (for example, shown two identical cups and two identical shoes, the correct response is "same," as the same relation (sameness) is exhibited within each pair). In attempting to explain these results, the authors of the original study [44] suggest that the early experience with the tags allows the chimps to imaginatively reduce the second-order problem to a first-order one. If the chimps, on encountering (e.g.) a pair of identical shoes, now call to mind an image of the tag (a red square), and if they do the same for the pair of identical cups, then all they need do to solve the harder problem is to judge that the two tag images (two red squares) are the same. The higher-order problem is thus reduced to a lower-order one they are already able to solve.

Here, the skills acquired by practice using concrete symbolic objects (the plastic tags) seem to allow the chimps to perceive the scene in a new way. The effect is perhaps rather like that of an augmented reality display, in which a suitably equipped user (courtesy of an eyeglass display, or, one day, retinal or even cortical implants) sees signs and information superimposed upon the actual visual scene in front of his eyes. Thus augmented, the user is able to distribute attention around the visual scene in new ways (for example, he might now follow a moving arrow traced in the sky and picking out a specific constellation).

Experience with public language symbols, I want to suggest, augments human cognition in the same kind of way. It allows us to direct and distribute attention in new ways. And it does so by in effect creating a special kind of surrogate situation: one in which what is otherwise unavailable is not the visual scene itself, but a particular way of parsing the scene into salient components and events.

Further evidence in favor of such a view comes from the well-known studies by Boysen et al. [5] in which two chimpanzees are presented with two bowls containing different numbers of pieces of fruit, and then one subject is asked to select the bowl the other chimp will receive. To get the most reward, the chimp must thus point at the bowl containing less fruit. Chimps are unable to do this. But when the same task is presented using numerals instead of fruit (numerals the chimps have, of course, been previously taught to understand), they can succeed. Here, the use of the surrogate (numeral-choosing) situation apparently enables the chimp to inhibit an otherwise overwhelming response. This effect is not quite the same as the one in the tagging experiment, but it is similar in that here too, the use of labels frees the chimps from the gravitational pull of their ordinary perception-action routines. Contra Carruthers, then, I think we may conceive perceptually encountered or recalled symbols and sentences as acting less like inner data structures, replete with slots and apt for genuine inner combinatoric action, and more like cheap ways of adding task-simplifying and attention-reconfiguring structure to the perceptual scene.

Carruthers' assumption of widespread modularity and encapsulation is thus not needed. We can allow that human learning is sculpted by some innate biases, but we need not suppose that the thinking thus supported is architecturally isolated. Into this nexus, learning words (such as "blue") and phrases (such as "to the left of") may be seen as the developmental source of new forms of selective attention. Learning the words (which act as cheap behavioral targets for reward and reinforcement routines) shapes the child's attentional biases in language-specific ways that then promote new forms of problem solving, such as the use of conjoint geometric and color cues. And certainly, there is ample evidence that children show attentional biases that are sensitive to the language they are learning (or have learned)—for example, [4, 27, 39] explicitly suggest that learned linguistic contexts come to "serve as cues that automatically control attention" [39: p. 113].

Carruthers might reply that this kind of story is undermined by the shadowing results in which linguistic activity in a distractor task impairs performance on (only) those tasks requiring integration of information across domains. But if the very process of selective attention to a complex conjoined cue required (in humans) the retrieval of at least some of the relevant lexical items, the shadowing result would equally be predicted. According to this alternative account, then, we need to first retrieve simple quasi-perceptual placeholders (such as words) when (and only when) a task requires us to target attentional resources on complex, "unnatural," or otherwise elusive elements of the encountered scene. Such a picture is radically different from one such as Carruthers', in which the logical form of the natural-language sentence provides the skeleton for a whole new compositional internal representation unifying the outputs of multiple modules.

Words and sentences, in the view I am advocating, act as stable anchor points around which complex neural dynamics can then swirl and coalesce. Instead of thinking of linguistic encodings as enabling informational integration by acting as a common format for the outputs of multiple modules, we can then think of the whole process as one not of translation into a single unifying representation, but of *attention-based coordination*. Words and sentences here serve as kinds of simple, cheap quasi-perceptual marker posts, enabling the agent to attend to specific dimensions of a scene, including specific combinations of aspects of the scene, that would otherwise remain unnoticed. Language emerges as the source of a potent form of surrogate situatedness that makes available new ways of parsing a scene into salient, attendable components and events.

6 Conclusions: Burrows within Burrows

This article started by rehearsing a familiar challenge to embodied and dynamical approaches to cognition. The challenge was to show how such approaches will deal with more traditionally "cognitive" phenomena, such as slow, reflective "offline" forms of thought and reason. The basic form of an answer has been suggested. Approach the traditional phenomena via an indirect route, paying special attention to the role of human-built structures (especially stable public symbols and surrogate situations) in pressing additional benefit from basic strategies of embodied action and response.

Along the way, I also speculated on a possible coevolutionary process,[1] in which public symbol structures and surrogate situations form the selective niche (either developmentally or evolutionarily) for the emergence of new forms of cognitive circuitry. These would be circuits whose typical contributions differ, in a variety of important ways, from those of the circuits involved in more basic forms of learning and adaptive response. In particular, such circuits will operate on a different time scale from the others, and will support effortful response to newly received, or temporarily maintained, information.

It is important to notice, however, that I leave completely open the question of how best to conceive of these (putatively) new resources. One possibility (which I do not favor) would be to imagine a kind of all-or-nothing divide, with a wholly representational and computational understanding somehow appropriate to (and sufficient for the understanding of) the new resources, and a wholly dynamical (and nonrepresentational) understanding somehow appropriate to the older ones. But, as J. Scott Kelso (personal communication) usefully points out, this seems unnecessarily divisive and theoretically unmotivated. The picture I am defending is rather one of complex, multi-time-scale interaction. On the one hand, it is a picture in which there genuinely are multiple, qualitatively different strands involved in the production of behavior (I further defend this view in [12]). On the other hand, there can be no doubt but that these various strands must usually work together in the production of behaviors. Reasoning using surrogate situations provides a good demonstration of this cooperation. An interesting direction for future work will be understanding the detailed nature of the connections between these differently time-scaled dynamics and the way they work together to create human understanding [34, 25].

Surrogate situations, including those constructed by means of words and language, are, I have suggested, the "singing burrows" of the human mind. These "cognitive singing burrows" range from the use of simple tags and tokens, to the creation of mock-ups and models, on to the use of complex sentences and all the varied props and tools of the modern age. It is by means of this whole cascade of props and surrogacy that we humans routinely exceed the apparent limits of our basic modes of animal reason. Yet while we may be amazed and impressed to learn that the humble mole cricket can use its earthbound burrows to form a double exponential horn set in a near-infinite baffle, and can thus amplify its humble stridulations into one of the loudest sounds in nature, we are routinely blind to the depth of our own cognitive transformations.

Consider: We do not just use our cognitive singing burrows to think better. We use the burrows to help build better burrows. We build better tools to think with, and use these very tools to discover still better tools to think with. And we tune the way we use our burrows, by building some burrows (educational practices) to train ourselves to use our best burrows better. Furthermore, we tune the way we tune the way we use our best burrows, by devising burrows to help build better burrows for educating ourselves in the use of our burrows (i.e., we devise tools for teacher education and training). Finally, unlike the mole cricket, we do not encode all the recipes for our burrow-building in our genes, not even as simple feedback-driven routines. Instead, we make the burrows themselves (books, oral traditions, software) do double duty as their own encodings for production by future generations. We store in books and manuals the procedures needed to build paper factories and bookbinderies, as well as the procedures to build planes, trains, and automobiles.

Who we are, as human engines of thought and reason, is thus a matter densely determined by a cascade of extended cognitive physiologies achieved by progressively fitting an open-ended sequence of technologies to somewhat plastic human brains. Just as in the case of the mole cricket, the special genetic program for all this may be surprisingly minimal. But once it is up and running, we are barely,

1 Since writing this article, I have discovered a closely related coevolutionary proposal in [41].

if at all, constrained by the limits of the onboard apparatus that once fitted us to the good old savannah (for lots more on this, see [11]).

Scaling up new embodied approaches to confront the full gamut of human cognition thus requires a modestly brave step. It requires us to take our technological and cultural props and scaffolds seriously, and to treat these larger systems as the source of extended, and perhaps radically innovative, distributed cognitive architectures. That means developing a science of the mind that, although embodied and embedded, reaches far beyond the flesh.

Acknowledgments

Thanks to Ed Hutchins for suggesting J. Scott Turner's book, and to Rob Goldstone for useful chats about learning, abstraction, and surrogate situations. I am also grateful to an anonymous referee, and to J. Scott Kelso, for a number of useful and penetrating comments on a previous draft.

References

1. Ballard, D., Hayhoe, M., Pook, P., & Rao, R. (1997). Deictic codes for the embodiment of cognition. *Behavioral and Brain Sciences*, *20*(4), 723–767.

2. Barsalou, L. (1998). Perceptual symbol systems. *Behavioral and Brain Sciences*, *22*(4), 577–660.

3. Beer, R. (1996). Towards the evolution of dynamical neural networks for minimally cognitive behavior. In P. Maes, M. Mataric, J.-A. Meyer, J. Pollack, & S. Wilson (Eds.), *From Animals to Animats 4 (pp. 421–429). Cambridge, MA: MIT Press.*

4. Bowerman, M., & Choi, S. (2001). Shaping meanings for language: Universal and language-specific in the acquisition and shaping of semantic categories. In M. Bowerman, & S. Levinson (Eds.), *Language acquisition and conceptual development* (pp. 475–511). Cambridge, UK: Cambridge University Press.

5. Boysen, S. T., Bernston, G., Hannan, M., & Cacioppo, J. (1996). Quantity-based inference and symbolic representation in chimpanzees (*Pan troglodytes*). *Journal of Experimental Psychology: Animal Behavior Processes*, *22*, 76–86.

6. Carruthers, P. (2002). The cognitive functions of language. *Behavioral and Brain Sciences*, *25*(6), 657–674.

7. Cheng, K. (1986). A purely geometric module in the rat's spatial representation. *Cognition*, *23*, 149–178.

8. Clark, A. (1998). Magic words: How language augments human computation. In S. Boucher, & P. Carruthers (Eds.), *Thought and language* (pp. 162–183). Cambridge, UK: Cambridge University Press.

9. Clark, A. (1999a). Visual awareness and visuomotor action. *Journal of Consciousness Studies*, *6*(11–1), 1–18.

10. Clark, A. (1999b). "Leadership and influence: The manager as coach, nanny and artificial DNA." In J. Clippinger (Ed.), *The biology of business: De-coding the natural laws of enterprise* (pp. 47–66). San Francisco: Jossey-Bass.

11. Clark, A. (2003). *Natural-born cyborgs: Minds, technologies, and the future of human intelligence*. New York: Oxford University Press.

12. Clark, A. (2003). Forces, fields and the role of knowledge in action. *Adaptive Behavior*, *11*, 270–272.

13. Clark, A., & Toribio, J. (1994). "Doing without representing?" *Synthese*, *101*, 401–431.

14. Cox, R. (1999). Representation construction, externalized cognition and individual differences. *Learning and Instruction*, *9*, 343–363.

15. Dennett, D. (1996). *Kinds of minds*. New York: Basic Books.

16. DeLoache, J. S. (1991). Symbolic functioning in very young children: Understanding of pictures and models. *Child Development*, *62*, 736–752.

17. DeLoache, J. S., & Burns, N. M. (1994). Symbolic functioning in preschool children. *Applied Developmental Psychology*, *15*, 513–527.

18. Gedenryd, H. (1998). *How designers work: Making sense of authentic cognitive activities*. PhD Dissertation, Lund University Cognitive Studies, Lund, Sweden.

19. Gregory, R. (1981). *Mind in science: A history of explanations in psychology*. Cambridge, UK: Cambridge University Press.

20. Goldstone, R., & Sakamoto, Y. (2003). The transfer of abstract principles governing complex adaptive systems. *Cognitive Psychology*, *46*, 414–466.

21. Hermer, L., & Spelke, E. (1994). A geometric process for spatial reorientation in young children. *Nature*, *370*, 57–59.

22. Hermer, L., & Spelke, E. (1996). Modularity and development: The case of spatial reorientation. *Cognition*, *61*, 195–232.

23. Hermer-Vazquez, L., Spelke, E., & Katsnelson, A. (1999). Sources of flexibility in human cognition: Dual-task studies of space and language. *Cognitive Psychology*, *39*, 3–36.

24. Kelso, J. A. S. (1995). *Dynamic patterns*. Cambridge, MA: MIT Press.

25. Kelso, J. A. S. (unpublished). Computational versus dynamical accounts in cognitive science: Coexisting alternatives? Center for Complex Systems & Brain Sciences, Florida Atlantic University.

26. Lakoff, G., & Johnson, M. (1999). *Philosophy in the flesh: The embodied mind and its challenge to Western thought*. New York: Basic Books.

27. Lucy, J., & Gaskins, S. (2001). Grammatical categories and the development of classification preferences: A comparative approach. In M. Bowerman, & S. Levinson (Eds.), *Language acquisition and conceptual development* (pp. 257–283). Cambridge, UK: Cambridge University Press.

28. Markman, A., & Genter, D. (1993). Structural alignment during similarity comparisons. *Cognitive Psychology*, *25*, 431–467.

29. Mataric, M. (1992). Integration of representation into goal-driven behavior based robots. *IEEE Journal of Robotics and Automation*, *8*, 304–312.

30. McBeath, M. K., Shaffer, D. M., & Kaiser, M. K. (1995). How baseball outfielders determine where to run to catch fly balls. *Science*, *268*, 569–573.

31. Milner, A., & Goodale, M. (1995). *The visual brain in action*. Oxford, UK: Oxford University Press.

32. O'Reilly, R. (in press). The division of labor between the neocortex and hippocampus. In George Houghton (Ed.), *Connectionist modeling in cognitive (neuro-)science*. Psychology Press.

33. O'Reilly, R., & Munakata, Y. (2000). *Computational explorations in cognitive neuroscience*. Cambridge, MA: MIT Press.

34. Pattee, H. H. (1987). Instabilitites and information in biological self-organization. In F. E. Yates (Ed.), *Self-organizing systems: The emergence of order*. New York: Plenum Press.

35. Pfeifer, R., & Scheier, C. (2001). *Understanding intelligence* (2nd printing, paperback ed.). Cambridge, MA: MIT Press.

36. Raibert, M. H., & Hodgins, J. K. (1993). Legged robots. In R. D. Beer, R. E. Ritzmann, & T. McKenna (Eds.), *Biological neural networks in invertebrate neuroethology and robotics* (pp. 319–354). San Diego, CA: Academic Press.

37. Scaife, M., & Rogers, Y. (1996). External cognition: How do graphical representations work? *International Journal of Human-Computer Studies*, *45*, 185–213.

38. Slocum, A. C., Downey, D. C., & Beer, R. D. (2000). Further experiments in the evolution of minimally cognitive behavior: From perceiving affordances to selective attention. In J. Meyer, A. Berthoz, D. Floreano, H. Roitblat,, & S. Wilson (Eds.), *From Animals to Animats 6: Proceedings of the Sixth International Conference on Simulation of Adaptive Behavior* (pp. 430–439). Cambridge, MA: MIT Press.

39. Smith, L. (2001). How domain-general processes may create domain-specific biases. In M. Bowerman, & S. Levinson (Eds.), *Language acquisition and conceptual development* (pp. 101–131). Cambridge, UK: Cambridge University Press.

40. Spelke, E., & Tsivkin, S. (2001). Language and number: A bilingual training study. *Cognition*, *57*, 45–88.

41. Sterelny, K. (2003). *Thought in a hostile world*. Oxford, UK: Blackwell.

42. Thelen, E., & Smith, L. (1994). *A dynamic systems approach to the development of cognition and action*. Cambridge, MA: MIT Press.

43. Thelen, E., Schoner, G., Scheier, C., & Smith, L. (2001). The dynamics of embodiment: A field theory of infant perseverative reaching. *Behavioral and Brain Sciences*, *24*(1), 1–55.

44. Thompson, R. K. R., Oden, D. L., & Boysen, S. T. (1997). Language-naive chimpanzees (*Pan troglodytes*) judge relations between relations in a conceptual matching-to-sample task. *Journal of Experimental Psychology: Animal Behavior Processses*, *23*, 31–43.

45. Thompson, E., & Varela, F. (2001). Radical embodiment: Neural dynamics and consciousness. *Trends in Cognitive Sciences*, *5*(10), 418–425.

46. Turner, S. J. (2000). The extended organism: The physiology of animal-built structures. Cambridge, MA: Harvard University Press.

47. Van Gelder, T. (1995). What might cognition be, if not computation? *Journal of Philosophy*, *XCII*(7), 345–381.

48. Van Gelder, T. J. (1998). The dynamical hypothesis in cognitive science. *Behavioral and Brain Sciences*, *21*, 1–14.

49. Varela, F., Thompson, E., & Rosch, E. (1991). *The embodied mind*. Cambridge, MA: MIT Press.

50. Webb, B. (1996). A cricket robot. *Scientific American*, *275*, 62–67.