Philosophical Psychology
Vol. 19, No. 3, June 2006, 1–17

Routledge
Taylor & Francis Group

Material Symbols

Andy Clark

What is the relation between the material, conventional symbol structures that we encounter in the spoken and written word, and human thought? A common assumption, that structures a wide variety of otherwise competing views, is that the way in which these material, conventional symbol-structures do their work is by being translated into some kind of content-matching inner code. One alternative to this view is the tempting but thoroughly elusive idea that we somehow think in some natural language (such as English). In the present treatment I explore a third option, which I shall call the "complementarity" view of language. According to this third view the actual symbol structures of a given language add cognitive value by complementing (without being replicated by) the more basic modes of operation and representation endemic to the biological brain. The "cognitive bonus" that language brings is, on this model, not to be cashed out either via the ultimately mysterious notion of "thinking in a given natural language" or via some process of exhaustive translation into another inner code. Instead, we should try to think in terms of a kind of coordination dynamics in which the forms and structures of a language qua material symbol system *play a key and irreducible role. Understanding language as a complementary cognitive resource is, I argue, an important part of the much larger project (sometimes glossed in terms of the "extended mind") of understanding human cognition as essentially and multiply hybrid: as involving a complex interplay between internal biological resources and external non-biological resources.*

Keywords: ■■■

1. Translation Models of Language

Jerry Fodor famously holds that "knowing a natural language is knowing how to pair its expressions with Mentalese expressions" (Fodor, 1998, p. 67). To have a certain

Correspondence to: Andy Clark, Department of Philosophy, George Square, Edinburgh EH8 9JX, Scotland, UK.
Email: andy.clark@ed.ac.uk

ISSN 0951-5089 (print)/ISSN 1465-394X (online)/06/030001-17 © 2006 Taylor & Francis
DOI: 10.1080/09515080600689872

thought, on this view, is to token a certain mentalese sentence. Language impacts thought, on such accounts, in virtue of a process of translation that transforms the public sentence into a content-capturing inner code. This is the prime example of what might be dubbed a ''translation view of language.'' Encountered language (be it speech or the written word), if this view is correct, merely serves to activate complexes of internal states or representations that are the real cognitive workhorses. It turns up too, though with a radically different twist, in Paul Churchland's connectionist-inspired vision of human cognition. For Churchland (1989, p. 18; 1996, p. 107) public language offers only what might be dubbed ''thin translations'' of the much richer meanings made available by vector codings and high-dimensional state spaces. Public words and sentences, Churchland suggests, offer at best a shallow or ''one-dimensional'' (1989, p. 18) echo of the rich and supra-linguistic meanings encoded using the formidable resources of these high-dimensional state-space encodings. But despite disagreeing over the precise ''fit'' (excellent versus disappointingly sparse) between the public structures and the inner realm, Churchland, like Fodor, retains what is essentially a translation view of how public language works. It is just that the *contents* of the internal translations (the contents proper to the real cognitive workhorses) for Churchland typically exceed, rather than simply replicate, those of the public language structures themselves. Thus insofar as public language is a useful tool at all, it works, according to Churchland, by activating one or more suites of rich internal (connectionist) representations—one might dub them ''neuralese'' —that then encode the meaning. The actual public language items are thus once again mere scaffolding to be kicked away once content has, however imperfectly, been transmitted from person to person.

According to the translation picture, then, language works its magic by being understood, and understanding is in turn conceived as consisting wholly in something like translation into some other content-matching (or content-exceeding) format. Such a view depicts language as a kind of high-level code that needs to be compiled or interpreted (in the computer science sense) to do its work. As a result, the material forms can then be thrown away as the essence—the meanings carried, conveyed, implied—has been fully extracted and rendered in some alternative inner format.

Compare now the use of a standard tool. When I use a spade to dig the garden, the spade makes an *ongoing and complementary* contribution to that made by my biological body. There is, in such a case, no obvious sense in which I biologically replicate the essence of the spade's activity. Instead, the digging power resides in the larger coupled system.

The alternative to the translation picture, that I wish to pursue here, makes the role of public language more like that of the spade. On the view to be explored, language (and material symbols more generally) play a double role. On the one hand they do (crucially, always) activate other kinds of cognitive resource, bits of mentalese or neuralese as you prefer. But they *also* play an irreducible role as the material symbols they are. For part of the role (and the power) of such items (spoken or written words and sentences) is to *complement* the basic modes of operation and representation

endemic to the biological brain. Understanding language, if this view is correct, involves getting to grips with a special kind of coordination dynamics: one in which the actual material structures of public language (or sometimes their shallow "imagistic" internal representations) play a key and irreducible role. This view, as I shall develop it, is related to, but I think remains distinct from, Dennett's famous (1991) account of language as installing a new serial virtual machine in the head. For whereas Dennett depicts experience with language as essentially transformative, as changing the fundamental nature of (part of) the in-head machinery, I shall attempt to depict public language as a complementary resource that works *with* the more basic machinery without installing any fundamentally new styles of representation or processing *within* that machinery. This view of language, I shall finally suggest, can usefully be seen as part of the much larger project (sometimes glossed in terms of the "extended mind" —see Clark, 1997; Clark & Chalmers, 1998) of understanding human cognition as essentially and multiply hybrid: as involving a complex interplay between internal biological resources and external non-biological resources. Language, however, occupies a wonderfully ambiguous position on any hybrid cognitive stage, since it seems to straddle the internal-external borderline itself, looking one moment like any other piece of the biological equipment, and at the next like a peculiarly potent piece of external cognitive scaffolding.

2. Some Trial Cases

It will be helpful to put a range of concrete cases on the table as a kind of (somewhat hopeful) anchor for the subsequent discussion. The examples that follow may be familiar, but I ask the reader's patience. It is not the cases themselves that matter, so much as the general pattern, displaying some of the interlinked variety of ways that the actual material forms of language may impact cognition. The cases that follow are arranged in (what seems to me to be) ascending order of cognitive impact.

2.1. First Grade of Cognitive Involvement: Language as a Source of Additional Targets for Attention and Learning

There are three examples falling into this category. The first, and by far the simplest, is the well-known case of Sheba and the treats, as recounted in Boysen, Bernston, Hannan and Cacioppo (1996). Sheba (an adult female chimpanzee) has had symbol and numeral training: she knows about numerals. Sheba sits with Sarah (another chimp), and two plates of treats are shown. What Sheba points to, Sarah gets. Sheba points to the greater pile, thus getting less. She visibly hates this result, but (unless the reward matrix is greatly exaggerated) can't seem to improve. However, when the treats arrive in containers with a cover bearing numerals on top, the spell is broken and Sheba points to the lesser number, thus gaining *more* treats.

What seems to be going on here, according to Boysen et al., is that the material symbols, by being simple and stripped of most treat-signifying physical cues, allow the chimps to sidestep the capture of their own behavior by ecologically-specific

fast-and-frugal subroutines. The symbol loosens the bond between agent and world, and between perception and action, and it does so not in virtue of being the key to a rich mental representation (though it may be that too) but rather by itself, qua material symbol, providing a new target for selective attention and a new fulcrum for the control of action.

In much the same way the act of labeling creates a new realm of perceptible objects upon which to target basic capacities of statistical and associative learning. The act of labeling thus alters the computational burdens imposed by certain kinds of problem. I have written quite a bit on this elsewhere, so I'll keep this brief. My favorite example (Clark, 1998) begins with the use, by otherwise language-naïve chimpanzees, of concrete tags (simple and distinct plastic shapes) for relations such as sameness and difference. Thus, a pair such as cup–cup might be associated with a red triangle (sameness) and cup–shoe with a blue circle (difference). This is not in itself surprising. What is more interesting is that after this training, the tag-trained chimps (and only tag-trained chimps) prove able to learn about the abstract properties of higher-order sameness, i.e. they are able to learn to judge of two presented pairs (such as cup–cup and cup–shoe) that the relation between the relations is one of higher order difference (or better, lack of higher-order sameness) since the first pair exhibits the sameness relation and the second pair the difference relation (Thompson, Oden, & Boysen, 1997). The reason the tag-trained chimps can perform this surprising feat is, so the authors suggest, because by mentally recalling the tags the chimps can reduce the higher-order problem to a lower-order one: all they have to do is spot that the relation of difference describes the pairing of the two recalled tags (red triangle and blue circle). The learning made possible through the initial loop into the world of stable, perceptible plastic tokens has allowed the brain to build circuits that, perhaps by simply imaging the tokens themselves at appropriate moments, reduce the higher-order problem to a lower-order one of a kind their brains are already capable of solving. *Experience* with external tags and labels thus enables the brain itself—by *shallowly representing* those tags and labels—to solve problems whose level of complexity and abstraction would otherwise leave us baffled.[1]

A related effect may also be observed (and this is our third and final case in this category) in recent connectionist work on language learning. Thus in a recent review, Smith and Gasser (2005) ask a very nice question. Why, given that human beings are such experts at grounded, concrete, sensorimotor driven forms of learning, do the symbol systems of public language take the special and rather rarified forms that they do?

> One might expect that a multimodal, grounded, sensorimotor sort of learning would favor a more iconic, pantomime-like language in which symbols were similar to referents. But language is decidedly not like this . . . there is no intrinsic similarity between the sounds of most words and their referents: the form of the word *dog* gives us no hints about the kind of thing to which it refers. And nothing in the similarity of the forms of *dig* and *dog* conveys a similarity in meaning. (Smith & Gasser, 2005, p. 22)

The question, in short, is "Why in a so profoundly multimodal sensorimotor agent such as ourselves is language an arbitrary symbol system?" (p. 24).

One possible answer, of course, is that language is like that because (biologically basic*) thought* is like that, and the forms and structures of language reflect this fact. But another answer, and the one I want to pursue, says just the opposite. Language is like that, it might be suggested, because thought (or rather, biologically basic thought) is *not* like that. The computational value of a public system of essentially context-free, arbitrary symbols, lies, according to this opposing view, in the way such a system can push, pull, tweak, cajole and eventually cooperate with various non-arbitrary, modality-rich, context-sensitive forms of biologically basic encoding.

Consider, to take the main case presented by Gasser and Smith, the development of one-trial word learning. This powerful capacity may be multiply dependent, Smith and Gasser suggest, on the presence of a public code comprising arbitrary labels. Early word learning, they suggest, is all about building up multimodal clusters of associated properties. But later on, as is well known, children become rapid word learners, adding four to nine new words a day, and generalizing the new words in ways appropriate to their distinct categories. Such rapid-fire learning looks to require the deployment of what Smith and Gasser describe as "second-order, rule-like generalizations." Such generalizations, they argue, are driven by properties of arbitrary public symbol systems.

For example, a new word for an artifact will probably apply to *similarly shaped* things (think of tractors, frying pans, toothbrushes). Whereas a new word for a substance will apply to other things made of the same material (wooly hats, wooly jumpers, wooly mittens etc.). Rapid word learning looks to involve just such abilities of higher-order generalization. Neural network simulations by Eliana Colunga (Colunga & Smith, 2005) suggest that the formation of such second-order generalizations depends on the arbitrariness and orthogonality of the linguistic labels provided. Make the labels non-orthogonal, and the second-order knowledge is not acquired (non-arbitrary labels must tend towards non-orthogonality due to property overlaps in the objects and events labeled).

It is not fully clear why this should be so, but it seems likely that experience with concrete orthogonal labels helps the system to pull perceptually similar categories apart, and thus supports new kinds of grouping that make visible deeper commonalities and differences, yielding the kinds of implicit knowledge (e.g., concerning the typical *kinds of feature* that individuate artifacts rather than substances) that underpin rapid-fire learning and that would otherwise be buried too deep in the search space for basic sensorimotor forms of intelligence like ourselves.

2.2. Second Grade of Cognitive Involvement: Language as a Resource for Directing and Maintaining Attention on Complex Conjoined Cues

The key case in this category concerns spatial reasoning in infants and adults. In a famous study by Hermer-Vazquez, Spelke, and Katsnelson (1999), pre-linguistic

infants were shown the location of a toy or food in a room, then were spun around or otherwise disoriented and required to try to find the desired item. The location was uniquely determinable only by remembering conjoined cues concerning the color of the wall and its geometry (the toy might be hidden in the corner between the long wall and the short blue wall). The rooms were designed so that the geometric or color cues were individually insufficient, and would yield an unambiguous result only when combined together. Pre-linguistic infants, though perfectly able to detect and use both kinds of cue, were shown to exploit only the geometric information, searching randomly in each of the two geometrically indistinguishable sites. Yet adults and older children were easily capable of combining the geometric and non-geometric cues to solve the problem. Importantly, 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. Only children who were able to spontaneously conjoin spatial and (e.g.) color terms in their free speech (who would describe something as, say, to the right of the long green wall) were able to solve the problem.

Hermer-Vazquez et al. (1999) then probed the role of language in this task by asking subjects to solve problems requiring the integration of geometric and non-geometric information while performing one of two other tasks. The first 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. Yet subjects engaged in speech shadowing were unable to solve the integration-demanding problem, while those shadowing rhythm were unaffected. An agent's linguistic abilities, the researchers concluded, are indeed actively involved in their ability to solve problems requiring the integration of geometric and non-geometric information.

There are currently various competing models of just how this involvement is best unpacked (see especially Carruthers, 2002). But probably the simplest story is that here too linguistic resources provide a convenient fulcrum for the complex distribution of attention. They enable us better to control the disposition of selective attention to ever-more complex feature combinations. The shadowing result is then explained by the idea that active attention to a complex conjoined cue requires the (possibly unconscious) retrieval of at least some of the relevant lexical items. Laying the emphasis on attentional effects thus allows us to accommodate this case in a way that dovetails with the earlier ones. In each case, linguistic activity (some kind of conscious or unconscious access to representations of language-specific lexical items) helps us to target our attentional resources on complex, conjunctive, or otherwise elusive, elements of the encountered scene.

2.3. Third Grade of Cognitive Involvement: Language as Providing Some of the Proper Parts of Hybrid Thoughts

At last, then, we arrive at the highest grade of cognitive involvement I want to scout, and surely the most contentious. This is the idea (to be explained shortly) of language

as providing some of the proper parts of hybrid thoughts. The key example here concerns the role of number words in mathematical reason.

What is going on when you think the thought that "98 is one more than 97"? According to the translation-based model, to think that thought is to translate the English sentence into something else, where that something else might be a sentence of mentalese (for Fodor) or a point in some exotic high-dimensional state space (for Churchland).

But consider a recent account due to Stanislas Dehaene and colleagues (see Dehaene, 1997; Dehaene, Spelke, Pinel, Stanescu, & Tviskin, 1999). Dehaene depicts this kind of precise mathematical thought as emerging at the productive intersection of three distinct cognitive contributions. The first involves a basic biological capacity to individuate small quantities: 1-ness, 2-ness, 3-ness and more-then-that-ness, to take the standard set. The second involves another biologically basic capacity, this time for approximate reasoning concerning magnitudes (discriminating, say, arrays of 8 dots from arrays of 16, but not more closely matched arrays). The third, not biologically basic but arguably transformative, is the learnt capacity to use the specific number words of a language, and the eventual appreciation that each such number word names a distinct quantity. Notice that this is not the same as appreciating, in at least one important sense, just what that quantity is. Most of us can't form any clear image of, e.g., of 98-ness (unlike, say, 2-ness). But we appreciate nonetheless that the number word '98' names a unique quantity in between 97 and 99.

When we add the use of number words to the more basic biological nexus, Dehaene argues, we acquire an evolutionarily novel capacity to think about an unlimited set of exact quantities. We gain this capacity not because we now have an encoding of 98-ness just like our encoding of 2-ness. Rather, the new thoughts depend directly (but not exhaustively) upon our tokening the numerical expressions themselves, as symbol strings of our own public language. The actual numerical thought, on this model, is had courtesy of the *combination* of this tokening (of the symbol string of a given language) and the appropriate activation of the more biologically basic resources mentioned earlier.

Here is some evidence for this view, as presented in Dehaene et al. (1999). First, there are the results of studies of Russian-English bilinguals. In these studies, Russian-English bilinguals were trained (quite extensively) on 12 cases involving exact and approximate sums of (the same) pairs of two-digit numbers, presented as words in one or other language. For example, (in English), a subject might be trained on the question "Four + Five" and asked to select their answer from "Nine" and "Seven". This is called the exact condition, as it requires exact reasoning since the two candidate numbers are close to each other. By contrast, a question like " 'Four + Five', select answer from 'Eight' and 'Three' " belongs to the approximate condition, as it requires only rough reasoning as the candidates are now quite far apart.

After extensive training on the pairs, subjects were later tested on the very same sums in either the original or the other (non-trained) language. After training, performance in the approximation condition was shown to be unaffected by switching the language, whereas in the exact condition, language switching resulted

in asymmetric performance, with subjects responding much faster if the test-language corresponded to the training-language. Crucially, then, there were no switching costs at all for trained approximate sums. Performance was the same regardless of language switching. Training-based speedup is thus non-language switchable for the exact sums and fully switchable for the inexact ones. Such studies, Dehaene et al. concluded, provide:

> evidence that the arithmetic knowledge acquired during training with exact problems was stored in a language-specific format . . .. For approximate addition, in contrast, performance was equivalent in the two languages providing evidence that the knowledge was stored in a language-independent form. (1999, p. 973)

A second line of evidence draws on lesion studies in which (to take one example) a patient with severe left-hemisphere damage cannot determine whether $2 + 2$ is 3 or 4, but reliably chooses 3 or 4 over 9, indicating a sparing of the approximation system.

Finally, Dehaene et al. (1999) present neuroimaging data from subjects engaged in exact and approximate numerical tasks. The exact tasks show significant activity in the speed-related areas of the left frontal lobe, while the approximate tasks recruit bilateral areas of the parietal lobes implicated in visuo-spatial reasoning. These results are presented as a demonstration "that exact calculation is language dependent, whereas approximation relies on nonverbal visuo-spatial cerebral networks" (p. 970) and that "even within the small domain of elementary arithmetic, multiple mental representations are used for different tasks" (p. 973).

Dehaene (1997) also makes some nice points about the need to somehow establish links between the linguistic labels and our innate sense of simple quantities. At first, it seems, children learn language-based numerical facts *without* such appreciation. According to Dehaene, "for a whole year, children realize that the word 'three' is a number without knowing the precise value it refers to" (1997, p. 107). But once the label gets attached to the simple innate number line, the door is open to understanding that *all* numbers refer to precise quantities, even when we lack the intuitive sense of what the quantity is (e.g. my own intuitive sense of 53-ness is not distinct from my intuitive sense of 52-ness, though all such results are variable according to the level of mathematical expertise of the subject).

Typical human mathematical competence, all this suggest, is plausibly seen as a kind of hybrid, whose elements include:

(i) Images or encodings of actual words in a specific language;
(ii) an appreciation of the fact that each distinct number word names a specific and distinct quantity; and
(iii) a rough appreciation of where that quantity lies on a kind of approximate, analog number line (e.g. 98 is just less than halfway between 1 and 200).

In a certain sense then, we rely on the coordinated action of various resources. On this view, there is (at least) an internal representation of the numeral, of the word-form, and of the phonetics, along with other resources (such as the analog number line) to which these become (with learning) roughly keyed via some sense

of relative location. What matters for present purposes (for what I am calling the third grade of cognitive involvement) is that there may be no need to posit (for the average agent), in addition to this coordinated medley, any further content-matching internal representation of, say, 98-ness. Instead, the presence of actual number words in a public code (and of shallow, imagistic, internal representations *of those very public items*) is itself part of the coordinated representational medley that constitutes many kinds of arithmetical knowing.

Thus consider the thought that there are 98 toys on the table. According to the translation view, to think the thought that there are 98 toys on the table you must have succeeded in translating the English sentence into *a fully content-providing "something else."* The "something else" might be an atom or sentence of mentalese (for Fodor) or a point in some exotic state space (for Churchland). By contrast, according to this quite radical alternative, the thought that there are 98 toys on the table is (for most of us) dependent upon the presence of a hybrid representational vehicle.[2] This is a vehicle that includes, as expected, the activation of a variety of content-relevant internal representations (in neuralese or mentalese, let's assume). But it also includes as a co-opted proper part, a token (let's think of it as an image, very broadly construed) of a conventional public language encoding ("ninety-eight") appropriately linked to various other resources (such as some rough position on an analog number line).

This half-glimpsed possibility is, I suspect, actually the most important way that language (and indeed all kinds of cultural props and artifacts) may impact thought: by actually becoming parts of the thinkings themselves. This is not, as you will have noticed, the most transparent of ideas, and I doubt I have it even halfway right. But the scope is satisfyingly large. In the case at hand, the vehicle or process, though arguably genuinely hybrid, is fully internal to the biological agent. But in other cases, there seems no reason to insist that this matters. Perhaps some of our representational vehicles and processes (the actual mechanistic underpinnings of our thinkings no less) may get spread out across biological brains and all sorts of socio-cultural artifacts, including gestures, diagrams, external text, software applications, and more.

This view of language is a perfect fit with (though not, I suppose, essential for) a very big picture according to which human cognition gains much of its distinctive force and power from its (biologically-based) ability to build and maintain new forms of external representational structure, that are then apt for non-fully-replicated use, in other words for cognitive incorporation.[3] That is to say, we make ourselves into new kinds of cognitive engine by (amongst other things) annexing and co-opting elements of external cognitive scaffolding as proper parts of hybrid computational routines. In this context, it is worth observing (though only as a kind of coda to the main story) that the (putative) ability of material symbols to participate in cognitive processes helps show the way out of a dilemma often urged upon the friends of extended cognition. For our very best cognitive artifacts, if they are sometimes to play a role as proper parts of cognitive processes, need to be assimilated within *but not totally swallowed up by* the workings of the potent basic

biological cognitive engine itself. Thus it is very tempting, when confronted with arguments that would give a strong cognitive role to artifacts (or, in this case, to public, conventional, symbolic codes) to respond with a kind of dilemma. Either the artifact/public code is not playing a truly cognitive role (it is merely input, not part of the processing) or (insofar as it seems to be playing such a role) it does so only because it has been translated into something else, some quite different inner thing, that really is suited to play such a role. Either way, it seems, the benefits that accrue can be fully explained, at least as far as here and now thinking is concerned, without continued reference to the features and properties of the artifact/public code.[4] (For some versions of this dilemma in the literature opposing the "extended mind," see Adams & Aizawa, 2001; Rupert, 2004).

The way around the dilemma should now be clear. By stressing coordination dynamics and hybrid representational forms, we leave room for genuine complementarity between the biological and artifactual cognitive contributions. We thus begin to see how artifactual resources may sometimes be co-opted without being fully recapitulated by the biological elements. This is what the notion of hybridity was always meant to suggest, and it avoids both horns of the dilemma.

One bad reason why this can seem impossible in the case of language is, of course, if we still think that understanding "obviously" always requires translation into some other content-matching (or better) inner code. But it is pretty clear that this cannot be the case all the way down, on pain (see Fodor, 1975) of an endless regress of such codes. So once the right coordination dynamics are in place, there is no reason why some hybrid whole could not *itself* be the physical vehicle, appropriately poised to control action and choice, of the relevant understanding. Indeed, the whole of artificial intelligence is surely itself testimony to the power of the idea that well-poised physically instantiated representations can sometimes constitute understanding without needing to be (in any further way) understood themselves.

3. Hybrid Thoughts?

The idea on offer then is that the symbolic environment (very broadly construed) can sometimes impact thought and learning *not* by some process of full-translation, in which the meanings of symbolic objects are exhaustively translated into an inner code, a mentalese, or even a Churchland-style neuralese, but by something closer to coordination. On the coordination model, the symbolic environment impacts thought *both* by activating such other resources (the usual suspects) *and* by using either the symbolic objects themselves (or inner image-like internal representations of the objects) as additional fulcrums of attention, memory and control. In the maximum strength version, these symbolic objects quite literally appear as elements in representationally hybrid thoughts.[5]

Now for a confession. For quite a few years, I thought this was a radical idea that fans of (to take the most extreme example) the language of thought (LOT) hypothesis would surely reject out of hand. Their idea, after all, was that words mean what they

do in virtue of being paired with expressively parallel snippets of mentalese. Imagine my surprise then, when I found this little snippet hidden away in that 1998 review of Carruthers by Jerry Fodor:

> I don't think that there are decisive arguments for the theory that all thought is in Mentalese. In fact, I don't think it's even true, in any detail . . . . I wouldn't be in the least surprised, for example, if it turned out that some arithmetic thinking is carried out by executing previously memorized algorithms that are defined over public language symbols for numbers (''now carry the '2' ''and so forth). It's quite likely that Mentalese co-opts bits of natural language in all sorts of ways; quite likely the story about how it does so will be very complicated indeed by the time that the psychologists get finished telling it. (1998, p. 72, italics in original)

Fodor here gestures, it seems to me, at an incredibly potent mechanism of cognitive expansion. Pretty clearly though, Fodor himself attaches little importance to the concession, quickly adding that ''For all our philosophical purposes (e.g. for purposes of understanding what thought content is, and what concept possession is, and so forth) *nothing essential is lost* if you assume that all thought is in Mentalese'' (1998, p. 72, italics added).

By contrast, I am inclined to see the potential for representational hybridity as massively important to understanding the nature and power of much distinctively human cognition. One obvious reason for this difference in assessment is that Fodor has the LOT already in place. So the basic biological engine, on his account, comes factory-primed with innovations favoring structure, integration, generality and compositionality. If, however, your vision of the basic biological engine is not one that so closely echoes the properties and features of sentences and propositional attitudes (if, for example, it is closer to Churchland's vision of a complex but thoroughly connectionist device, or to Barsalou's, 1999, vision of a ''perceptual symbol system'') then the potential cognitive impact of a little hybridity and co-opting may be much greater than Fodor concedes. It may be *essential* to such a system's ability to think rather a wide variety of thoughts that the inner goings-on involve, as genuinely constitutive elements, something like images or traces of the public language symbols (words) themselves. Words and sentences, on this view, may be potent structures many of whose features and properties (arbitrary amodal nature, extreme compactness and abstraction, compositional structure, and so on) deeply complement the contributions of basic biological cognition. In such a case, it would hardly be right to treat the co-opting strategies as marginal for the understanding of thought and concepts.[6]

This vision of mind-expansion by the use of hybrid representational forms remains visibly close to that of Dennett (1991, 1996). But Dennett, as mentioned earlier, places most of his bets on the radically transformative power of our encounters with language, and thus ends up with a story that seems more developmental than genuinely hybrid. Admittedly, drawing these lines is a delicate task (Densmore & Dennett, 1999). But where Dennett depicts exposure to language as installing a new virtual serial machine via affecting ''myriad microsettings in the plasticity of the brain'' (1991, p. 219), on the hybrid model words and sentences

remain potent real-world structures encountered and used by a basically (though this is obviously too crude) pattern-completing brain. Of course, even on this account the brain sometimes represents (shallowly, imagistically) these structures. But language need not profoundly reorganize[7] the shape and texture of the neural coding routines themselves.[8]

4. Working Models?

The idea of truly hybrid, bio-artifactually distributed cognition is, I hope to have shown, at least intelligible. Moreover, the examples arrayed in §2 are meant to suggest that it is also actual. But how, in detail, might the whole thing work? Do we have even a single existence proof, in the form of an up and running simulation, that shows how such hybridity might be mechanically implemented?

The nearest I have so far found is a small but suggestive set of simulations reported in Clowes & Morse (2005). The simulations investigate ways in which the internal re-use of a public symbol system might aid cognition. Internal re-use was enabled by the provision, in some agents, of a dedicated re-entrant loop able to recycle "heard" linguistic inputs during processing. In the simulations, simple agents were evolved to find and move geometric figures in response to commands couched in a "public" code. The commands tell the agent's (simple recurrent neural nets with "visual" and "word" inputs) which of four tasks to perform on objects in an on-screen arena. The tasks are to move the objects to the top ("up"), to move the objects to the bottom ("down"), to move the objects to the right ("right") or to move the objects to the left ("left").

Groups of agents were evolved under three conditions:

1. A control condition, with no dedicated word re-entrance loop. In this condition the agent "hears" words as commands and must act on that basis alone (but the architecture is still that of a simple recurrent neural net (SRNN), so there is memory available as the output layer cycles back to the input layer alongside new inputs at the next time step).
2. Permanent Word Re-entrance: In this condition, the "heard" command words are cycled back via a dedicated part of a recurrent loop while problem solving continues.
3. Self-controlled Re-entrance: This is as (2) except the net has an additional output unit that can gate the dedicated word re-entrance loop on and off. "Heard" words can thus be recycled during processing at the agent's discretion.

Clowes and Morse found that under the control condition (no dedicated word re-entrance) the agents take longer to learn to succeed at *any* of the tasks, and seem unable to learn to succeed at all four. This is because improvements in one task seemed to always result in impairment to performance on one or more of the others. The nets with permanent word re-entrance (condition 2) fared better. Good performance was quite rapidly evolved, and typically displayed in at least three

and often all four tasks. Most impressive of all, however, were the (condition 3) nets with self-gateable word re-entrance. These agents produced the best performance, on all tasks, and with the least evolutionary costs (in terms of numbers of generations required for competence). Overall, the authors conclude, "[the] results clearly demonstrate a qualitative difference between the control group and the [word re-entrant] conditions, despite the internal re-entrance of SRNN architectures present in all three conditions" (Clowes & Morse, 2005, p. 104).

Underlying this result, I would finally conjecture, may be something quite fundamental. Perhaps (but beware: this is now *pure* speculation) the role of re-presentations (imagistic recyclings) of words here can be understood as an example of the power of loosely coupled distinct processes. This is an effect already observed in work on so-called GasNets in which the combination of (a simulation of) freely diffusing gaseous neurotransmitters and of a more standard "electrical" artificial neural network learning resource likewise improves performance and speeds evolvability. To explain this result, Philipides, Husbands, Smith, and O'Shea (2005) suggest that when an organism must accommodate conflicting pressures (just as in the four "contradictory" tasks confronting the Clowes-Morse net) the presence of distinct but loosely coupled processes "allows the possibility of tuning one process against the other without destructive interference" (p. 154). Perhaps then part of the role of rehearsed words in aiding cognition, even on the very short time-scales of ongoing episodes of thinking, might one day be seen as another instance of the more general power of loose couplings between dynamically distinct processes. Perhaps, that is to say, words are just an especially potent resource able to enter into loosely coupled forms of online activity, allowing the system to find valuable trajectories through search space that might otherwise be blocked by destructive interference between superficially conflicting current ideas, goals, or contexts. For this to occur, ongoing control over the current degree of coupling, as in the "gated" self-cueing net, may well be crucial (again, see Philippides, Husbands, Smith, O'Shea, 2005, p. 158). All this is, to repeat, pure speculation. But I do suspect that these very general kinds of consideration, concerning search, dynamics and complex systems, will eventually prove very germane to the general project of trying to understand the advantages conferred by various forms of hybrid cognition.

5. Conclusions: Leaps and Boundaries

So what is the "cognitive bonus" that language brings? In this treatment, I have begun to explore one of the less-visited regions of this surprisingly mysterious landscape: the region in which the material structures of language play a cognitive role that in some way actually depends on, and exploits, that very materiality. To even glimpse this region we need to look beyond a seemingly inescapable model of how language must do its work, the model according to which encountered linguistic tokens act solely in virtue of a process of exhaustive translation into some other content-matching (or exceeding) internal representational format. This was what

we dubbed the "translation view" of language. On a pure translation view, it is hard to see how our linguistic encounters can do anything more than inculcate a kind of useful shorthand for ideas whose very thinkability requires only on the more fundamental tokenings (in mentalese or neuralese) with which they have come to be associated. The alternative on offer is a "hybrid model" according to which some of the cognitive benefits that language brings depend on the *complementary* action of actual material symbols (and image-like inner encodings of such symbols) and more biologically basic modes of internal representation.

Effects tentatively explored under this umbrella included the ideas that:

1. Otherwise inacessible contents can be learnt and grasped by agents skilled in the use of perceptually simple tokens that reify complex ideas.
2. The presence of material symbols (or images thereof) can productively alter the fulcrums of attention, perception and action.

And most contentiously of all:

3. Material symbols (or their shallow imagistic encodings) can coordinate with more basic representational resources to yield new forms of hybrid thought.

If this kind of story is even halfway correct, then minds like ours are indeed transformed by the web of material symbols and epistemic artifacts. But that transformation may neither require nor result in the installation of brand new internal representational forms. Instead, there may be much under-explored merit in the canny use of the external forms (and internal images of those very forms) themselves. Such forms may help sculpt and modify processes of selective attention, and act as elements within hybrid representational wholes.

One immediate merit of such a view is a more nuanced attitude to the vexed question of evolutionary cognitive continuity. Jesse Prinz (2004) makes the point well:

> Researchers who presume that we think in amodal symbols face a dilemma. If they argue that nonhuman animals lack such amodal symbols, they must postulate a radical leap in evolution. If they suppose that animals have amodal thoughts, they must explain why human thought is so much more powerful. Empiricism [Prinz's favorite, though not obligatory in the present context!] when coupled with the assumption that we can think in public language, explains the discrepancy in cognitive capacities without postulating a major discontinuity in evolution. (p. 427)

Needless to say, much remains to be done. It would be good to have a clear account of just what attention, that crucial variable that linguistic scaffolding seems so potently to adjust, actually *is*. It would be good to have much more in the way of genuine, implementable, fully mechanistic models of the various ways that internalized language might enhance thought. And it would be good to know just what it is about human brains and/or human history that has enabled structured public language to get such a comprehensive grip on minds like ours. But shortfalls aside, I hope to have at least brought the artifact model into clearer view, and to have

shown why it might be attractive to anyone who thinks that language makes a truly *deep* contribution to human thought and reason.

Acknowledgements

This paper grew out of material produced for the workshops on Memory, Mind and Media organized by John Sutton at Macquarie University, Sydney, Australia in December 2004. Thanks to John Sutton, Rob Wilson, Mark Rowlands, and all the speakers and participants at those meetings for their invaluable input and criticism. Thanks also to two anonymous referees for important and thought-provoking comments. This project was completed thanks to teaching relief provided by Edinburgh University and by matching leave provided under the AHRC Research Leave Scheme.

Notes

[1] Note that the suggestion here is not that processes of abstraction always or even typically require the loop through public tokens or symbols. Rather it is that such loops, when present, can play a distinctive cognition enhancing role. For some important explorations of the nature, scope and possible limits of such roles, see Schwartz and Black (1996), and Schyns, Goldstone, and Thibaut (1998).

[2] A possible worry (thanks to an anonymous referee for raising this issue) is that the kinds of rich interaction between different resources posited by hybrid accounts may first require the translation of the various different elements into a ''common code,'' thus undermining any claim of genuine hybridity. A possible analogy here is with cases of intermodal interaction, also sometimes said to require the existence of a common code. But in both cases a possible response, it seems to me, is simply to deny the requirement. Potent coordinated interaction need not require a common code. Consider the case of coding in the dorsal and ventral visual streams. The two streams (see Milner & Goodale, 1995) look to trade in highly distinct representational forms, yet in daily life (in uncompromised subjects) they work together seamlessly in the service of goal directed behaviour.

[3] See work on ''tools for thought,'' the ''extended mind,'' ''wide computation,'' ''vehicle externalism'': Clark (1997, 2003); Clark & Chalmers (1998); Dennett (1991, 1996); Hurley (1998); Rowlands (1999); Wilson (1994, 2004).

[4] Perhaps there are effects on learning trajectories (see the grade one examples) that resist the dilemma but for here and now thinking (so the argument goes) the options are as stated.

[5] From this point on, whenever I speak of 'hybrid representational forms' I shall mean forms that include both standard kinds of internal representation (mentalese, neuralese, perceptual symbol systems, . . .) and, as proper parts of a kind of distributed encoding, either the material symbols of some public language, or shallow imagistic encodings of those very forms.

[6] A second reason (for Fodor's downplaying the power of hybridity) flows from his (in)famous views concerning concept learning. For given those views, the meaning of hybrid representational forms could not be learnt unless the learner *already* had the resources to represent *that very meaning* using more biologically basic (indeed, innate) resources. This, however, is not the time or place to engage in this important discussion (for some countervailing thoughts, see Prinz & Clark, 2004).

[7] It is a moot point exactly what constitutes ''profound'' reorganization. But in essence, the most radical version of the view I am defending holds that although the brain must learn to deal with the special class of linguistic structures, in so it need not reorganize its neural coding routines in any way that is deeper or more profound than might occur, say, when we first learn to swim, or to play volleyball.

[8] A further question is exactly how the hybrid view defended in this paper relates to that of Carruthers (2002). The relation here is hard to determine, as the starting points of the two accounts are very different. Carruthers buys into large-scale mental modularity and sees natural language as cognition enhancing in virtue of being the sole medium of all module-integrating thoughts. The notion of hybrid cognitive vehicles defended here seems to me to be attractively weaker than this. It is indifferent to the truth or falsity of modularity.

References

Adams, F., & Aizawa, K. (2001). The bounds of cognition. *Philosophical Psychology*, *14*, 43–64.

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, *22*, 577–609.

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.

Carruthers, P. (2002). The cognitive functions of language. *Behavioral and Brain Sciences*, *25*, 657–726.

Churchland, P. M. (1989). *A neurocomputational perspective*. Cambridge, MA: MIT Press.

Churchland, P. M. (1995). *The engine of reason, the seat of the soul*. Cambridge, MA: MIT Press.

Churchland, P. M. (1996). The neural representation of the social world. In L. May, M. Friedman & A. Clark (Eds.), *Minds and morals* (pp. 91–108). Cambridge, MA: MIT Press.

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

Clark, A. (1998). Magic words: How language augments human computation. In P. Carruthers & J. Boucher (Eds.), *Language and thought: Interdisciplinary themes* (pp. 162–183). Cambridge, England: Cambridge University Press.

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

Clark, A., & Chalmers, D. (1998). The extended mind. *Analysis*, *58*, 7–19.

Clowes, R. W., & Morse, A. F. (2005). Scaffolding cognition with words. In L. Berthouze, F. Kaplan, H. Kozima, Y. Yano, J. Konczak, G. Metta, J. Nadel, G. Sandini, G. Stojanov & C. Balkenius (Eds.), *Proceedings of 5th International Workshop on Epigenetic Robotics: Modeling Cognitive Development in Robotic Systems* (Lund University Cognitive Studies, Vol. 123, pp. 101–105). Lund, Sweden: Lund University Cognitive Studies.

Colunga, E., & Smith, L. B. (2005). From the lexicon to expectations about kinds: A role for associative learning. *Psychological Review*, *112*, 347–382.

Dehaene, S. (1997). *The number sense*. Oxford, England: Oxford University Press.

Dehaene, S., Spelke, E., Pinel, P., Stanescu, R., & Tviskin, S. (1999). Sources of mathematical thinking: Behavioral and brain imaging evidence. *Science*, *284*, 970–974.

Dennett, D. C. (1991). *Consciousness explained*. New York: Little Brown.

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

Densmore, S., & Dennett, D. C. (1999). The virtues of virtual machines. *Philosophy and Phenomenological Research*, *59*, 747–767.

Fodor, J. A. (1975). *The language of thought*. New York: Crowell.

Fodor, J. A. (1998). Do we think in mentalese: Remarks on some arguments of Peter Carruthers. *In critical condition: Polemical essays on cognitive science and the philosophy of mind* (pp. 63–74). Cambridge, MA: MIT Press.

2

Fodor, J. A. (2004). Having concepts: A brief refutation of the 20th century. *Mind & Language*, *19*, 29–47.

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.

Hurley, S. (1998). *Consciousness in action*. Cambridge, MA: Harvard University Press.

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

Philippides, A., Husbands, P., Smith, T., & O'Shea, M. (2005). Flexible couplings: Diffusing neuromodulators and adaptive robotics. *Artificial Life*, *11*, 139–160.

Prinz, J. (2004). Sensible ideas: A reply to Sarnecki and Markman and Stilwell. *Philosophical Psychology*, *17*, 419–430.

Prinz, J., & Clark, A. (2004). Putting concepts to work: Some thoughts for the 21st Century. *Mind & Language*, *19*, 57–69.

Rowlands, M. (1999). *The body in mind: Understanding cognitive processes*. Cambridge, England: Cambridge University Press.

Rupert, R. (2004). Challenges to the hypothesis of extended cognition. *Journal of Philosophy*, *101*, 389–428.

Schwartz, D. L., & Black, J. B. (1996). Shuttling between depictive models and abstract rules: Induction and fallback. *Cognitive Science*, *20*, 457–498.

Schyns, P. G., Goldstone, R. L., & Thibaut, J.-P. (1998). The development of features in object concepts. *Behavioral and Brain Sciences*, *21*, 1–54.

Smith, L., & Gasser, M. (2005). The development of embodied cognition: Six lessons from babies. *Artificial Life*, *11*, 13–30.

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 Processes*, *23*, 31–43.

Wilson, R. A. (1994). Wide computationalism. *Mind*, *103*, 351–372.

Wilson, R. A. (2004). *Boundaries of the mind: The individual in the fragile sciences—Cognition*. Cambridge, England: Cambridge University Press.