

"Magic Words: How Language Augments Human Computation" in
P. Carruthers and J. Boucher (Eds) LANGUAGE AND THOUGHT:
INTERDISCIPLINARY THEMES (Cambridge University Press: Cambridge, 1998) P.162-
183

MAGIC WORDS: HOW LANGUAGE AUGMENTS HUMAN COMPUTATION

Andy Clark

*Philosophy/Neuroscience/Psychology Program
Department of Philosophy
Washington University
St. Louis, MO 63130
e-mail: andy@twinearth.wustl.edu*

0. Word Power.

Of course, words aren't magic. Neither are sextants, compasses, maps, slide rules and all the other paraphenelia which have accreted around the basic biological brains of homo sapiens. In the case of these other tools and props, however, it is transparently clear that they function so as to either carry out or to facilitate computational operations important to various human projects. The slide rule transforms complex mathematical problems (ones that would baffle or tax the unaided subject) into simple tasks of perceptual recognition. The map provides geographical information in a format well-suited to aid complex planning and strategic military operations. The compass gathers and displays a kind of information that (most) unaided human subjects do not seem to command. These various tools and props thus act to generate information, or to store it, or to transform it, or some combination of the three. In so doing, they impact our individual and collective problem-solving capacities in much the same dramatic ways as various software packages impact the performance of a simple pc.

Public language, I shall argue, is just such a tool -- it is a species of external artifact whose current adaptive value is partially constituted by its role in re-shaping the kinds of computational space that our biological brains must negotiate in order to solve certain types of problems, or to carry out certain complex projects. This computational role of language has been somewhat neglected (not un-noticed, but not rigorously pursued either) in recent cognitive science, due perhaps to a (quite proper) fascination with and concentration upon, that other obvious dimension: the role of language as an instrument of interpersonal communication. Work on sentence parsing, language use and story understanding has thus concentrated on the role of language in processes of information transfer between agents and on information retrieval from texts. But it has had little to say about the computational role of the linguistic formulations themselves, or about the special properties of the external media that support linguistic encodings.

In this treatment, I hope to display the broad shape of such an alternative interest. I begin by discussing the views of some recent (and not-so-recent) authors, who recognize in various ways, the potential role of language and text in transforming, reshaping and simplifying the computational tasks that confront the biological brain. Section 2 pursues this broad vision across a variety of cases involving planning, coordination, learning and the construction of complex thoughts and arguments. The third section extends these last considerations to encompass the rather special class of *meta-cognitive* operations and tries to implicate language as an essential part of the process of thinking about our own thoughts and cognitive profiles. The final section suggests some broader implications and raises some questions concerning the boundary between the intelligent agent and the world.

1. **Supra-Communicative Views of Language.**

The idea that language may do far more than merely serve as a vehicle for communication is not new. It is clearly present in the work of developmentalists such as Vygotsky (1962), and more recently that of Laura Berk and others (see e.g. essays in Diaz and Berk (1992)). It figures in the philosophical conjectures and arguments of e.g. Peter Carruthers (to appear) and Ray Jackendoff (to appear). And it surfaces in the more cognitive science oriented speculations of Daniel Dennett (1991). It will be helpful to begin by rehearsing some of the central ideas in this literature, before pursuing our own version viz. the idea of language as a computational transformer which allows pattern-completing brains to tackle otherwise intractable classes of cognitive problems.

Lev Vygotsky, a Soviet psychologist of the 1930's, pioneered the idea that the use of public language had profound effects on cognitive development. He posited powerful links between speech, social experience and learning. Two especially pertinent Vygotskian ideas, for present purposes, concern the role of private speech, and of scaffolded action (action within the so-called zone of proximal development -- see Vygotsky (trans., 1962)). We may call an action scaffolded to the extent that it relies on some kind of external support. Such support could come from the use of tools, or the knowledge and skills of others; that is to say, scaffolding (as I shall use the term) denotes a broad class of physical, cognitive and social augmentations -- augmentations which allow us to achieve some goal which would otherwise be beyond us. Simple examples include the use of a compass and pencil to draw a perfect circle, the role of other crew members in enabling a ship's pilot to steer a course and the infants ability to take its first steps only while suspended in the enabling grip of its parents. Vygotsky's focus on what was termed the Zone of Proximal Development was concerned with this latter type of case, in which a child is temporarily able to succeed at designated tasks only by courtesy of the guidance or help provided by another human being (usually, a parent or teacher). This idea dovetails with Vygotsky's interest in private speech in the following way. When the child, confronted by a tricky challenge, is talked through the problem by a more experienced agent, the child can often succeed at tasks which would otherwise prove impossible (think of learning to tie your shoelaces). Later on, when the adult is absent, the child can conduct a similar dialogue, but this time with herself. But even in this latter case, it is argued, the speech (be it vocal or internalized) functions so as to guide behavior, to focus attention, and to guard against common errors. In such cases, the role of language is to guide and shape our own behavior

-- it is a tool for structuring and controlling action and not merely a medium of information transfer between agents.

This Vygotskian image is supported by more recent bodies of developmental research, such as that carried out by Laura Berk and Ruth Garvin. Berk and Garvin (1984) observed and recorded the ongoing speech of a group of 5-10 year olds in Kentucky. They found that most of the children's private speech (speech not addressed to some other listener) seemed keyed to the direction and control of the child's own actions. They found that the incidence of such speech increased when the child was alone and engaged in trying to perform some difficult task. In subsequent studies (Bivens and Berk (1990), Berk (1994)) it was found that the children who made the most self-directed comments were the ones who subsequently mastered the tasks best. Berk's conclusions, from these and other studies, was that self-directed speech (be it vocal or silent inner rehearsal) is a crucial cognitive tool that allows us to highlight the most puzzling features of new situations, and to direct and control our own problem-solving actions.

The theme of language as a tool has also been developed by the philosopher Christopher Gauker. Gauker's concern, however, is to re-think the intra-individual role of language in terms of (what he calls) a cause-effect analysis. The idea here is to depict public language not as a tool for representing the world or expressing one's thoughts but a tool for effecting changes in ones environment (Gauker (1990) p. 31). To get the flavor, consider the use of a symbol, by a chimpanzee, to request a banana. The chimp touches a specific key on a key-pad (the precise physical location of the key can be varied between trials) and learns that making that symbol light tends to promote the arrival of bananas. The chimp=s quasi-linguistic understanding is explicable Gauker suggests, in terms of the chimp=s appreciation of a cause-effect relationship between the symbol production and changes in its local environment. Gauker looks at a variety of symbol-using behaviors and concludes that they all succumb to this kind of analysis. This leads him to hypothesize that, although clearly more complex, human beings linguistic understanding likewise consists in a grasp of the causal relations into which linguistic signs may enter (Gauker, op.cit., 44).

Gauker tends to see the role of language as, if you like, directly causal: as a way of getting things done, much like reaching out your hand and grabbing a cake. However, the idea that we learn, by experience, of the peculiar causal potencies of specific signs and symbols is in principle much broader. We might even, as in the Vygotskian examples and as argued in Dennett (1991), discover that the self-directed utterance of words and phrases has certain effects on our own behavior! We might also learn to exploit language as a tool in a variety of even less direct ways, as a means of altering the shape of computational problem spaces (see Section 2 following).

One obvious question which the putative role of language as a self-directed tool raises is how does it work?. What is it about, for example, self-directed speech which fits it to play a guiding role? After all, it is not at all clear how we can tell ourselves anything we don't already know! Surely, all that public language can ever be is a medium for expressing ideas which are already formulated and understood in some other, more basic, inner code?

It is precisely this view which a supra-communicative account of language has ultimately to reject.

One way to do so is to depict public language as itself the medium of a special kind of thought. Another (not altogether distinct) way is to depict linguaform inputs as having distinctive effects on some inner computational device. Peter Carruthers¹ (to appear) champions the first of these, while Daniel Dennett (1991) offers a version of the second. Thus Carruthers argues that, in this case at least, we should take very seriously the evidence of our own introspection. It certainly often seems as if our very thoughts are composed of the words and sentences of public language. And the reason we have this impression, Carruthers argues, is because it is true: Ainner thinking is literally done in inner speech (see Carruthers (to appear) ch. 2 for an extensive discussion). By extension, Carruthers is able to view many intra-personal uses of language as less a matter of simple communication than of (what he nicely terms) public thinking. This perspective fits satisfyingly with the Vygotskian view championed by Berk, and is also applicable to the interesting case of writing down our ideas. Here Carruthers suggests one does not first entertain a private thought and then write it down: rather, the thinking is the writing (Carruthers, to appear, MS p. 56). I shall return to this point later (see section 2), since I believe that what Carruthers says is almost right, but that we can better understand the kind of case he has in mind by treating the writing as an environmental manipulation which transforms the problem space for human brains.

Carruthers, in depicting language as itself the vehicle of (certain types of) thought, is nonetheless careful to reject what he calls the Whorfian Relativism of the Standard Social Science Model (op.cit., MS p. 302). The reference here is to the idea, promoted by Benjamin Whorf (1956), that human minds are profoundly shaped and altered by the particular public languages we come to speak. Carruthers view is not that specific languages somehow deeply alter or re-program the brain, but rather² that certain kinds of human thinking are actually constituted by sequences of public language symbols (written down, spoken, or internally imagined). Such a hypothesis, Carruthers argues, can help account for a wide range of both introspective and experimental and pathological data³.

An alternative way to unpack a supra-communicative view of language, we noted, is to suppose that the linguistic inputs actually re-program or otherwise alter the high-level computational structure of the brain itself. The exegesis is delicate (and therefore tentative), but something akin to this view seems to be held by Daniel Dennett when he suggests that conscious human minds are more-or-less serial virtual machines implemented-inefficiently-on the parallel hardware that evolution has provided for us (Dennett, 1991, p. 278). In this and other passages, the idea seems to be that the bombardment of (something like) parallel processing, connectionist, pattern-completing brains by (amongst other things) public language texts and sentences (reminders, plans, exhortations, questions, etc.), results in a kind of cognitive reorganization akin to that which occurs when one computer system simulates another. In such cases, the installation of a new program allows the user to treat e.g. a serial LISP machine as if it was a massively parallel connectionist device. What Dennett is proposing is, he tells us (op.cit., p.218) the

same trick in reverse viz. the simulation of something like a serial logic engine using the altogether different resources of the massively parallel neural networks which biological evolution rightly favors for real-world, real-time survival and action.

Strikingly, Dennett suggests that it is this subtle re-programming of the brain by (primarily) linguistic bombardment which yields the phenomena of human consciousness (our sense of self) and enables us to far surpass the behavioral and cognitive achievements of most other animals (see e.g. Dennett (1995) p. 370-373). Dennett thus depicts our advanced cognitive skills as in large part a result not of our innate hardware (which may differ only in small, though important, ways from that of other animals) but of the special way that various plastic (programmable) features of the brain are modified by the effects of culture and language. As Dennett puts it, the serial machine is installed courtesy of myriad microsettings in the plasticity of the brain (Dennett (1991) p. 219). Of course, mere exposure to culture and language is not sufficient to ensure human-like cognition. You can expose a cockroach to all the language you like and get no trace of the cognitive transformations which Dennett sees in us. Dennett's claim is not that there are no initial hardware level differences. Rather it is that some relatively small hardware differences (e.g. between us and a chimpanzee) allow us to both create and benefit from public language and other cultural developments in ways which lead to a great snowball of cognitive change and augmentation, including, crucially, the literal installation of a new kind of computational device inside the brain.

Dennett's vision is complex, and not altogether unambiguous. The view I want to develop is clearly deeply related, but differs (I think) in one crucial respect. Where Dennett sees public language as effecting a profound but subtle re-organization of the brain itself, I am inclined to see it as in essence heart an external resource which complements -- but does not profoundly alter -- the brain's own basic modes of representation and computation. That is to say, I see the changes as relatively superficial ones, geared to allowing us to use and exploit various external resource to the full. The positions are not, of course, wholly distinct. The mere fact that we often mentally rehearse sentences in our head and use them to guide and alter our behavior means that one cannot (and should not) treat language and culture as wholly external resources. Nonetheless, it remains possible that such rehearsal neither requires nor results in the installation of any fundamentally different kind of computational device in the brain, but rather involves the use of the same old (essentially pattern-completing) resources to model the special kinds of behavior observed in the public linguistic world. And as Paul Churchland (1995, p. 264-269) points out, there is indeed a class of connectionist networks (recurrent networks -- see Elman (1993), and further discussion in Clark (1993)) which do seem well-suited to modeling such behavior.

This view of inner rehearsal is nicely developed by the connectionists Rumelhart, Smolensky, McClelland, and Hinton who argue that the general strategy of mentally modeling the behavior of selected aspects of our environment is especially important insofar as it allows us to imagine external resources with which we have previously physically interacted, and to replay the dynamics of such interactions in our heads. Thus experience with drawing and using Venn diagrams allows us to train a neural network which subsequently allows us to manipulate imagined Venn diagrams in our heads. Such

imaginative manipulations require a specially trained neural resource to be sure. But there is no reason to suppose that such training results in the installation of a different kind of computational device. It is the same old process of pattern completion in high dimensional representational spaces, but applied to the special domain of a specific kind of external representation. The link to a Vygotskian image is clear and remarked upon by the authors who summarize their view saying:

We can be instructed to behave in a particular way. Responding to instructions in this way can be viewed simply as responding to some environmental event. We can also remember such an instruction and tell ourselves what to do. We have, in this way, internalized the instruction. We believe that the process of following instructions is essentially the same whether we have told ourselves or have been told what to do. Thus even here we have a kind of internalization of an external representational format.

(1986) (Rumelhart, Smolensky, McClelland, and Hinton p.47)

The larger passage (p. 44-48) from which the above is extracted is, in fact, remarkably rich and touches on several of our major themes. The authors note that such external formalisms are especially hard to invent and slow to develop, and are themselves the kinds of product which (in an innocently bootstrapping kind of way) can evolve only thanks to the linguistically-mediated processes of cultural storage and gradual refinement over many lifetimes. They also note that by using real external representations we put ourselves in a position to use our basic perceptual/motor skills to separate problems into parts and to attend to a series of sub-problems, storing intermediate results along the way.

The Rumelhart et al vision thus depicts language as a key element in a variety of environmentally extended computational processes. This notion of computational processes inhering in larger systems (ones that may incorporate the activities of many individual biological brains) is further developed and defended in Hutchins (1995). Hutchins offers a beautiful and detailed treatment that highlights the ways representation may flow and be transformed within larger, socially and technologically extended systems. Hutchins' main example involves the way maps, instruments, texts and vocalizations all contribute to the complex process of ship navigation: a process that is best analyzed as an extended sequence of computational transitions, many of whose role is to transform problems into formats better situated to the perceptual and pattern-completing capacities of biological brains. The environmental operations thus *complement* the activities of the biological brains.

The tack I am about to pursue likewise depicts language as an external artifact designed to complement, rather than recapitulate or transfigure, the basic processing profile we share with other animals. It does not depict experience with language as a source of profound inner re-programming (pace Dennett). Whether it depicts inner linguistic rehearsal as literally constitutive of specific human cognizings (as Carruthers claims) is moot. Certainly, inner rehearsals, when they occur, are quite literally models of linguistic productions. But what is most important, I believe, is not to try to answer the question, do

we actually think in words (to which the answer is in a way yes, in a way no!) but to try to see what computational benefits accrue to biological pattern-completing brains in virtue of their ability to manipulate and sometimes model external representational artifacts.

2. Language and Computation: The 6 Ways.

Here, then, are six broad ways in which linguistic artifacts can complement the activity of the pattern-completing brain.

i. Memory Augmentation.

This is, of course, the most obvious and oft-remarked case. Here we simply use the artifactual world of texts, diaries, notebooks and the like as a means of systematically storing large and often complex bodies of data. We may also use simple external manipulations (such as leaving a note on the mirror) to prompt the recall, from on-board biological memory, of appropriate information and intentions at the right time. Here, the use of linguistic artifacts is perfectly continuous with a variety of other, simpler, environmental manipulations, such as leaving the empty olive oil bottle by the door so that you cannot help but run across it (and hence recall the need for olive oil) as you set out for the shops.

ii. Environmental Simplification.

This has both an obvious and a not-so-obvious aspect. The obvious (but still important) aspect concerns the use of labels to provide perceptually simple clues to help us negotiate complex environments. Signs for the cloakrooms, for nightclubs, and for city centers all fulfill this role. They allow a little learning to go a very long way, helping you find your targets in new cities without knowing in advance what, in detail, to seek or even where exactly to seek it. McClamrock (1995 p. 88) describes this strategy as one in which we enforce on the environment certain kinds of stable properties that will lessen our computational burdens and the demands on us for inference.

Closely related, but much less obvious, is the provision, by the use of linguistic labels, of a greatly simplified learning environment. It can be shown, for example, that the provision of linguistic labels for classes of perceptually presented objects can speed category learning in artificial neural networks. This is because the presentation of the same label accompanying a series of slightly different perceptual inputs (e.g., different views of dogs) gives the network a heavy hint. It flags the presence of some further underlying structure and thus invites the network to seek the perceptual commonality (for a detailed discussion see Schyns (1991), Clark (1993) Ch. 5). It also seems likely (though no formal demonstration exists) that for certain very abstract concepts, the *only* route to successful learning may go via the provision of linguistic glosses. Concept such as charity, extortion and black hole seem pitched too far from perceptual facts to be learnable without exposure to linguistically formulated theories. Language may thus enable us to comprehend equivalence classes that would otherwise lie forever outside our intellectual horizons.

iii. Coordination and the Reduction of On-Line Deliberation.

Human beings often make explicit plans. We say to others that we will be at such and such a place at such and such a time. We even play this game with ourselves, perhaps by writing

down a list of what we will do on what days and so on. Superficially, the role of such explicit planning is to allow the coordination of actions. Thus, if the other person knows you have said you'll be at the station at 9:00 a.m., they can time their taxi accordingly. Or, in the solo case, if you have to buy the paint before touching up the car, and if you have to go to the shops to buy other items anyway, you can minimize your efforts and enforce proper sequencing by following a plan. As the space of demands and opportunities grows, it often becomes necessary to use pencil and paper to organize and to re-organize the options, and then to preserve the result as a kind of external control structure available to guide your subsequent actions.

Closely related to such coordinative functions is the function of oiling the wheels of collaborative problem-solving. Collaborative problem solving (see e.g., Tomasello et al (1993) involves much more than the mere exchange of information and orchestration of activity. It involves actively prompting the other to work harder at certain aspects of a problem, and allowing the other to focus your own attention in places you might otherwise ignore. Here, then, the co-ordinative function of linguistic exchange phases into the further one of manipulating attention and controlling resource allocation (see (v) below).

Such broadly co-ordinative functions, thought important, do not exhaust the benefits of explicit (usually language-based) planning. As Michael Bratman has recently pointed out, the creation of explicit plans may play a special role in reducing the on-line cognitive load on resource-limited agents like ourselves. The idea here is that our plans have a kind of stability which pays dividends by reducing the amount of deliberation in which we engage as we go about much of our daily business. Of course, new information can, and often does, cause us to revise our plans. But we do not let every slight change prompt a re-assessment of our plans, intentions, even when other things being equal, we might now choose slightly differently. Human plans and intentions, Bratman suggests, play the role of blocking a wasteful process of continual re-assessment and choice, except in cases where there is some quite major pay-off for the disruption. (See Bratman (1987) for a full discussion).

Linguistic exchange and formulation thus plays a key role in coordinating activities (both at an inter- and intra-personal level) and in reducing the amount of daily on-line deliberation in which we engage.

iv. Taming Path-Dependent Learning.

Human learning, like learning in Artificial Neural Networks, looks hostage to at least some degree of path dependency. Certain ideas can be understood only once others are in place. The training received by one mind fits it to grasp and expand upon ideas which gain no foothold of comprehension in another. The processes of formal education, indeed, are geared to take young (and not so young) minds along a genuine intellectual journey, which may involve beginning with ideas now known to be incorrect, but which alone seem able to prime the system to later appreciate a finer grained truth. Such mundane facts are a reflection of cognitive path dependence -- you can't get everywhere from anywhere; where you are now strongly constrains your future intellectual trajectory. Moreover, such path dependency is nicely explained (see e.g., Elman (1993)) by treating intellectual progress as

involving something like a process of computational search in a large and complex space. Previous learning inclines the system to try out certain locations in the space and not others. When the prior learning is appropriate, the job of learning some new regularity is made tractable: the prior learning acts as a filter on the space of options to be explored. Artificial Neural Networks which employ gradient descent learning methods are highly constrained insofar as the learning routine forces the network always to explore at the edges of its current weight assignments. Since these constitute its current knowledge, it means that such networks cannot jump around in hypothesis space. The networks current location in weight space (its current knowledge) is thus a major constraint on what new ideas it can next explore (see Elman (1993) p.94).

When confronting devices which exhibit some degree of path dependency, the mundane observation that language allows ideas to be preserved and to migrate between individuals takes on a new force. For we can now appreciate how such migrations may allow the communal construction of extremely delicate and difficult intellectual trajectories and progressions. An idea which only Joe's prior experience could make available, but which can flourish only in the intellectual niche currently provided by the brain of Mary, can now realize its full potential by journeying between agents as and when required. Moreover, the sheer number of intellectual niches available within a linguistically linked community provides a stunning matrix of possible inter-agent trajectories. The observation that public language allows human cognition to be collective (e.g. Churchland (1995) p. 270) takes on new depth once we recognize the role of such collective endeavor in transcending the path-dependent nature of individual human cognition.

v. *Attention and Resource Allocation.*

Ron McClamrock reports a nice case from Marr⁴ in which we see a control loop which runs outside the head and into the local environment, In McClamrock's words:

Flies, it turns out, don't quite know that to fly they should flap their wings. They don't take off by sending some signal from the brain to the wings. Rather, there is a direct control link from the fly's feet to its wings, such that when the feet cease to be in contact with a surface, the fly's wings begin to flap. To take off, the fly simply jumps and then lets the signal from the feet trigger the wings.

McClamrock (1995) p. 85; emphasis in original.

Notice, then, how written and spoken language at times serves a similar goal. We write down a note to do such and such, thus creating an externalized control loop for our own future behavior. We follow someones vocal instructions as we learn to windsurf. Or we mentally rehearse such instruction as we practice on our own. Such phenomena reveal linguistic formulations as somehow helping to focus, monitor and control behavior. I do not think we yet understand exactly how language (especially mental rehearsal of instructions) interacts with more basic on-line resources so as to yield these benefits. But that it does indeed play some such role seems clear.

vi. *Data Manipulation and Representation.*

This final benefit accrues most directly to the use of actual text. As I construct this chapter, for example, I am continually creating, putting aside, and re-organizing chunks of text. I have a file which contains all kinds of hints and fragments, stored up over a long period of time, which may be germane to the discussion. I have source texts and papers full of notes and annotations. As I (literally, physically) move these things about, interacting first with one, then another, making new notes, annotations and plans, so the intellectual shape of the chapter grows and solidifies. It is a shape which does not spring fully developed from inner cogitations. Instead, it is the product of a sustained and iterated sequence of interactions between my brain and a variety of external props. In these cases, I am willing to say, a good deal of actual thinking involves loops and circuits which run outside the head and through the local environment. Extended intellectual arguments and theses are almost always the product of brains acting in concert with multiple external resources. These resources enable us to pursue manipulations and juxtapositions of ideas and data which would quickly baffle the un-augmented brain. (The simple case of physically manipulating Scrabble tiles to present new potential word-fragments to a pattern-completing brain (see Kirsh (to appear)) is a micro-version of the same strategy). In all such cases, the real environment of printed words and symbols allows us to search, store, sequence and reorganize data in ways alien to the on-board repertoire of the biological brain.

The moral of the 6 ways is thus clear. The role of public language and text in human cognition is not limited to the preservation and communication of ideas. Instead, these external resources make available concepts, strategies and learning trajectories which are simply not available to individual, un-augmented brains. Much of the true power of language lies in its underappreciated capacity to re-shape the computational spaces which confront intelligent agents.

3. Mangroves and Meta-Cognition.

If a tree is seen growing on an island, which do you suppose came first? It is natural (and usually correct) to assume that the island provided the fertile soil in which a lucky seed came to rest. Mangrove forests,⁵ however, constitute a revealing exception to this general rule. The Mangrove grows from a floating seed which establishes itself in the water, rooting in shallow mud flats. The seedling sends complex vertical roots through the surface of the water, culminating in what looks to all intents and purposes like a small tree posing on stilts. The complex system of aerial roots, however, soon traps floating soil, weed and debris. After a time, the accumulation of trapped matter forms a small island. As more time passes, the island grows larger and larger. A growing mass of such islands can eventually merge, effectively extending the shoreline out to the trees! Throughout this process, and despite our prior intuitions, it is the land which is progressively built by the trees.

Something like the Mangrove effect, I suspect, is operative in some species of human thought. It is natural to suppose that words are always rooted in the fertile soil of pre-existing thoughts. But sometimes, at least, the influence seems to run in the other direction. A simple example is poetry. In constructing a poem, we do not simply use words to

express thoughts. Rather, it is often the properties which of the words (their structure and cadence) which determine the thoughts that the poem comes to express. A similar partial reversal can occur during the construction of complex texts and arguments. By writing down our ideas we generate a trace in a format which opens up a range of new possibilities. We can then inspect and re-inspect the same ideas, coming at them from many different angles and in many different frames of mind. We can hold the original ideas steady so that we may judge them, and safely experiment with subtle alterations. We can store them in ways which allow us to compare and combine them with other complexes of ideas in ways which would quickly defeat the un-augmented imagination. In these ways, and as remarked in the previous section, the real properties of physical text transform the space of possible thoughts.

Such observations lead me to the following conjecture. Perhaps it is public language which is responsible for a complex of rather distinctive features of human thought viz, our ability to display *second order cognitive dynamics*. By second order cognitive dynamics I mean a cluster of powerful capacities involving self-evaluation, self-criticism and finely honed remedial responses.⁶ Examples would include: recognizing a flaw in our own plan or argument, and dedicating further cognitive efforts to fixing it; reflecting on the unreliability of our own initial judgements in certain types of situations and proceeding with special caution as a result; coming to see why we reached a particular conclusion by appreciating the logical transitions in our own thought; thinking about the conditions under which we think best and trying to bring them about. The list could be continued, but the pattern should be clear. In all these cases, we are effectively thinking about our own cognitive profiles or about specific thoughts. This thinking about thinking, is a good candidate for a distinctively human capacity -- one not evidently shared by the other, non-language-using animals who share our planet. As such, it is natural to wonder whether this might be an entire species of thought in which language plays the generative role -- a species of thought which is not just reflected in, or extended by, our use of words but is directly dependent upon language for its very existence. Public language and the inner rehearsal of sentences would, on this model, act like the aerial roots of the Mangrove tree -- the words would serve as fixed points capable of attracting and positioning additional intellectual matter, creating the islands of second-order thought so characteristic of the cognitive landscape of *homo sapiens*.

It is easy to see, in broad outline, how this might come about. For as soon as we formulate a thought in words (or on paper), it becomes an object for both ourselves and for others. As an object, it is the kind of thing we can have thoughts about. In creating the object, we need have no thoughts about thoughts -- but once it is there, the opportunity immediately exists to attend to it as an object in its own right. The process of linguistic formulation thus creates the stable structure to which subsequent thinkings attach.

Just such a twist on potential role of the inner rehearsal of sentences has been suggested by the linguist Ray Jackendoff. Jackendoff (to appear) suggests that the mental rehearsal of sentences may be the primary means by which our own thoughts are able to become objects of further attention and reflection. The key claim is that linguistic formulation makes complex thoughts available to processes of mental attention, and that this, in turn open

them up to a range of further mental operations. It enables us, for example, to pick out different elements of complex thoughts and to scrutinize each in turn. It enables us to stabilize very abstract ideas in working memory. And it enables us to inspect and criticize our own reasoning in ways that no other representational modality allows.

What fits internal sentence-based rehearsal to play such an unusual role? The answer, I suggest, must lie in the more mundane (and temporally antecedent) role of language as an instrument of communication. For in order to function as an efficient instrument of communication, public language will have been molded into a code well-suited to the kinds of interpersonal exchange in which ideas are presented, inspected and subsequently critiqued. And this, in turn involves the development of a type of code which minimizes contextuality (most words retain more-or-less the same meaning in the different sentences in which they occur), is effectively modality-neutral (an idea may be prompted by visual, auditory or tactile input and yet be preserved using the same verbal formula), and allows easy rote memorization of simple strings.⁷ By freezing our own thoughts in the memorable, context-resistant and modality-transcending format of a sentence we thus create a special kind of mental object -- an object which is apt for scrutiny from multiple different cognitive angles, which is not doomed to alter or change every time we are exposed to new inputs or information, and which fixes the ideas at a fairly high level of abstraction from the idiosyncratic details of their proximal origins in sensory input. Such a mental object is, I suggest, ideally suited to figure in the evaluative, critical and tightly focused operations distinction of second order cognition. It is an object fit for the close and repeated inspections highlighted by Jackendoff under the rubric of *attending* to our own thoughts. The coding system of public language is thus especially apt to be co-opted for more private purposes of inner display, self-inspection and self-criticism, exactly as predicted by the Vygotskian treatments mentioned in Section 1 above. Language stands revealed as a key resource by which we effectively redescribe⁸ our own thoughts in a format which makes them available for a variety of new operations and manipulations.

The emergence of such second order cognitive dynamics is plausibly seen as one root of the veritable explosion of types and varieties of external scaffolding structures in human cultural evolution. It is because we can think about our own thinking that we can actively structure our world in ways designed to promote, support and extend our own cognitive achievements. This process also feeds itself, as when the arrival of written text and notation allowed us to begin to fix ever more complex and extended sequences of thought and reason as objects for further scrutiny and attention.

To complete this picture, we should reflect that once the apparatus (internal and external) of sentential and text-based reflection is in place, we may expect the development of new types of non-linguistic thought and encoding – one's dedicated to the task of managing and interacting with the sentences and texts in more powerful and efficient ways.⁹ The linguistic constructions, thus viewed, are a new class of objects which invite us to develop new (non-linguistically based) skills of use, recognition and manipulation. Sentential and non-sentential modes of thought this co-evolve so as to complement, but not replicate, each others special cognitive virtues.

It is a failure to appreciate this deep complementarity that, I suspect, leads Paul Churchland (one of the best and most imaginative neurophilosophers around) to dismiss linguaform expression as just as shallow reflection of our real knowledge. Churchland fears that without such marginalization we might mistakenly depict all thought and cognition as involving the unconscious rehearsal of sentence-like symbol strings, and thus be blinded to the powerful, pattern-and-prototype-based encodings which look to be biologically and evolutionarily fundamental. But we have now scouted much fertile intermediate territory.¹⁰ In combining an array of biologically basic pattern-recognition skills with the special cognitive fixatives of word and text, we (like the Mangroves) create new landscapes, new fixed points in the sea of thought. Viewed as a complementary cognitive artifact, language can genuinely extend our cognitive horizons -- and without the impossible burden of recapitulating the detailed contents of non-linguistic thought.

4. Studying the Extended Mind.

Speech and text, we have seen greatly extend the problem-solving capacities of humankind. More profoundly, the practice of putting thoughts into words alters the nature of human experience. Our thoughts become determinate and public objects, apt for rational assessment and for all kinds of meta-cognitive scrutiny. In thus recognizing public language as a powerful transformer of individual computational and experiential space, we invite reflection of a number of further topics. I will end by mentioning just two.

The first concerns the nature of the internal representations that guide human action. A popular image, often associated with Jerry Fodor's reflections on the need for a language of thought (e.g., Fodor (1975)(1986)), depicts the internal representational arena as itself a locus of propositionally structured items -- sentences in Mentalese. This image has lately been the subject of a damaging series of criticisms stemming from the successes of non-linguaform computational approaches -- especially those of connectionist (or parallel distributed processing) models. The perspective developed above might, I suspect, encourage us to approach some of these issues in a slightly different way. For the Fodorian view is at least intuitively linked to views of language as essentially a communicative tool. This is because the Fodorian sees linguistic formulations as reasonably faithful reflections of both the contents and the structural forms of internal representations. The view I have been developing is quite different insofar as it depicts the linguistic formulations as importing genuine novelties onto our cognitive horizons. The linguistic formulations are seen as novel both in content and in structure. There is content-novelty insofar as linguistic expression makes new thoughts available by effectively freezing other thoughts as types of static object¹² (images can do this too, but they are not so easily traded in public exchange). And there is structural novelty insofar as the value of the linguistic formulations (especially in written text) partly consists, we saw, in their amenability to a variety of operations and transformations that do not come naturally to the biological brain working in non-linguistic mode. Such novelties, I contend, are not at all predicted by the image of a pre-existing inner code whose basic features and properties are merely recapitulated in our public language formulations. By contrast, they are exactly what would be expected if the public

code is not a handy recapitulation of our non-linguistic resources so much as a powerful complement to them.¹³

Such a view suggests a certain gloss on the history and origin of the Fodorian image itself. For perhaps one mistake of classical Artificial Intelligence (upon which the image purports to be based) lay in its mistaking the properties of the linguistically augmented and environmentally extended cognitive agent (the person plus a variety of external representations, especially texts) for the cognitive profile of the basic biological brain. Thus the neat classical separation of data and process and of static symbol structures and CPU may have reflected nothing so much as the gross separation between the biological agent and an external scaffolding of ideas persisting on paper, in filing cabinets and in electronic media.

This notion of the biological agent leads nicely to the second issue I wish to mention. It concerns the question of where the mind ends and the rest of the world begins. Otherwise put, the question concerns how to conceive and locate the boundary between an intelligent system and its world. For certain external (to the biological unit) props and aids may play such a deep role in determining the shape and flow of our thoughts as to invite depiction as part and parcel of the very mechanism of human reason. This depiction is most plausible in the case of the external props of written text and spoken words. For interactions with these external media are ubiquitous (in educated modern cultures), reliable and developmentally basic. Our biologic brains, after learning, expect the presence of text and speech as much as they expect to encounter weight, force, friction and gravity. Language for us is a constant, and as such can be safely relied upon as the backdrop against which on-line processes of neural computation take shape and develop. Just as a neural network controller for moving the arm to a target in space must define its commands to factor in the spring of muscles and the effects of gravity, so the processes of on-board reason may come to factor in the potential contributions of textual off-loading and reorganization, and vocal exchange. The overall cognitive competencies which we identify as mind and intellect may thus be more like ship navigation than capacities of the bare biological brain. Ship navigation (see Hutchins (1995)) is a global emergent from the well-orchestrated adaptation of an extended complex system (comprising individuals, instruments, and practices). Much of what we uncritically identify as our mental capacities may likewise, I suspect, turn out to be properties of the wider, extended systems of which human brains are just one (important) part. In constructing an academic paper, for example, it is common practice to deploy multiple annotated texts, sets of notes and files, plans, lists and more. The writing process often depends heavily on manipulations of these props -- new notes are created, old ones juxtaposed, source materials are wheeled on and off of work surfaces, etc. In giving credit for the final product, however, we often marginalize the special contributions of these external manipulations, and speak as if the biological brain did all the work. No parallel temptation afflicts the person who uses a crane to lift large weights, or a motorized digger to plough trenches! In these cases, it is clear that the person uses additional tools whose capacities extend and complement those of the unaided laborer. The relative invisibility of the special cognitive roles of text and words are a reflection, I think, of their ubiquity and ease of use: a reflection, indeed, of our tendency to think of these operations as proper to the biological agent rather than (like the crane) as technological

additions. Perhaps the truth lies midway -- the use of spoken words may be as biologically proper, to the human agent, as the use of webs is to the spider.¹⁵ And the use of written text may thus straddle the intuitive divide between the web (biologically proper) and the crane (a true artifact).

The point, in any case, is that use of words and texts may usefully be seen as computationally complementary to the more primitive and biologically basic kinds of pattern-completing abilities that characterize natural cognition. These complementary operations essentially involve the creation of self-standing structures (short-term ones, like spoken sentences, or long-term ones, like text) that can perform a variety of useful functions such as the sequencing and control of behavior and the freezing of thoughts and ideas into objects for further attention and analysis. The availability of these functions extends the bound of human cognition as surely as the provision of a new board extends the bounds of a personal computer. In particular, it is our capacity to create and operate upon external representations that allows us to use manipulations of the physical environment as integral parts of so many of our problem-solving routines. In thus reaching out to the world we blunt the dividing line between the intelligent system and the world. We create wider computational webs whose understanding and analysis may require us to apply the tools and concepts of cognitive science to larger, hybrid entities comprising brains, bodies and a variety of external structures, traces and processes.

To endorse a notion of computational processes as criss-crossing brain, body and world is not yet to endorse a parallel notion of cognitive or mental processes. Perhaps cognition is all in the head, but computation spreads liberally out into the world. My own inclinations are less conservative. I suspect that our intuitive notions of mind and cognition actually do pick out these larger extended systems and that as a result the biological brain is only one component of the intelligent system we call the mind.¹⁶ But I will settle for a weaker conclusion -- one that merely implicates our linguistic capacities in some highly productive transformations of our overall computational powers. This power of computational transformation constitutes a neglected virtue of linguistic practice. It reveals language as the ultimate upgrade: so ubiquitous it is almost invisible; so intimate, it is not clear whether it is a kind of tool or an added dimension of the user. But whatever the boundaries, we confront a complex coalition in which the basic biological brain is fantastically empowered by some of its strangest and most recent creations: words in the air, symbols on the printed page.

Notes

1. A major focus of both Carruthers' and Dennett's treatments is the relation between language and consciousness. I will not discuss these issues here, save to say that my sympathies lie more with Churchland (1995, Chapter 10), who depicts basic consciousness as the common property of humans and many non-linguistic animals. Language fantastically augments the power of human cognition. But it does not, I believe, bring into being the basic apprehensions of pleasure, pain and the sensory world in which the true mystery of consciousness inheres.
2. Carruthers position, unlike Whorf's, is thus compatible with both a realist conception of the mental and a fair degree of linguistic nativism.
3. A quick sampling of this data includes: the developmental lock-step of cognitive and linguistic abilities, the difficulties which language-deficient humans have with certain kinds of temporal discourse and the deficits of abstract thought found in global aphasics. See Carruthers (to appear) esp. MS p. 291.
4. Marr (1982) p. 32-33.
5. A particularly stunning example is the large Mangrove forest extending north from Key West, Florida to the Everglades region known as Ten Thousand Islands. The black Mangroves of this region can reach heights of 80 feet -- see Landi (1982) p. 361-363.
6. Two very recent treatments which emphasize these themes have been brought to my attention. Jean-Pierre Changeux (a neuroscientist and molecular biologist) and Alain Connes (a mathematician) suggest that self-evaluation is the mark of true intelligence -- see Changeux and Connes (1995). Derek Bickerton (a linguist) celebrates off-line thinking and notes that no other species seems to isolate problems in their own performance and take pointed action to rectify them -- see Bickerton (1995).
7. The modality neutral dimensions of public language are stressed by Karmiloff-Smith in her closely related work on representational re-description -- see Note 8 below. The relative context -- independence of the signs and symbols of public language is discussed in Kirsh (1991) and Clark (1993) Ch. 6.
8. The idea that advanced cognition involves repeated processes in which achieved knowledge and representation is redescribed in new formats (which support new kinds of cognitive operation and access) is pursued in much more detail in Karmiloff-Smith 1992, Clark 1993, Clark and Karmiloff-Smith 1994, and Dennett 1994. The original hypothesis of representational re-description was developed by Karmiloff-Smith (1979, 1986).
9. See e.g. Bechtel (1996) p. 125-131, Clark (1996) pp. 120-25.

10. Dennett (1991) explores just such a intermediate territory. I discuss Churchland's downplaying of language in detail in Clark (1996). For examples of such downplaying see P.M. Churchland (1989) p. 18, P.S. and P.M. Churchland (1996) p. 265-270.
11. For a perfect introduction to the debate, see the various essays gathered in MacDonald and MacDonald (eds)(1995).
12. See also Dennett's discussion of belief versus opinion, in Dennett (1987).
13. There is of course, a midway option: that public language later becomes a language of thought and that Fodor's image is misguided only in its claim that all thoughts occur in Mentalese. For discussion, see Carruthers (to appear).
14. See Clark (1989) p. 135, Hutchins (1995) Ch.9.
15. See Dawkins lovely (1982) book, *The Extended Phenotype* (Freeman) for an especially biologically astute treatment of this kind of case.
16. See Clark (to appear, Ch. 10), Clark and Chalmers (1995).

Bibliography

- Berk, L. (1994). Why Children Talk to Themselves. *Scientific American November*: 78-83.
- Berk, L. and R. Garvin (1984). Development of Private Speech among Low-Income Appalachian Children. *Developmental Psychology* 20(2): 271-286.
- Bickerton, D. (1995). *Language and Human Behavior*. Seattle: University of Washington Press.
- Bratman, M. (1987). *Intentions, Plans and Practical Reason*. Cambridge, MA, Harvard University Press.
- Carruthers, P. (to appear). *Language, Thought and Consciousness*.
- Changeux, J-P. & Connes, A. (1995). *Conversations on Mind, Matter and Mathematics*. Princeton, NJ: Princeton University Press.
- Churchland, P. M. (1989). *The Neurocomputational Perspective*. Cambridge, MIT/Bradford Books.
- Churchland, P. M. (1995). *The Engine of Reason, the Seat of the Soul*. Cambridge, MA, MIT Press.
- Clark, A. (1989). *Microcognition: Philosophy, Cognitive Science and Parallel Distributed Processing*. Cambridge, MIT Press.
- Clark, A. (1993). *Associative Engines: Connectionism, Concepts and Representational Change*. Cambridge, MIT Press.
- Clark, A. (1996). *Being There: Putting Brain, Body and World Together Again*. Cambridge, MA, MIT Press.
- Clark, A. (1996). Connectionism, Moral cognition and collaborative problem solving. In L. May, M. Friedman, and A. Clark (Eds), *Mind & Morals* (p. 109-128). Cambridge, MA: MIT Press.
- Clark, A. and A. Karmiloff-Smith (1993). The Cognizer's Innards: A Psychological & Philosophical Perspective on the Development of Thought. *Mind & Language*.
- Clark, A. and D. Chalmers (1995). *The Extended Mind*. St. Louis, Washington University.
- Dawkins, R. (1982). *The Extended Phenotype*. Oxford, Oxford University Press.
- Dennett, D. (1987). *The Intentional Stance*. Cambridge, MA, MIT Press.
- Dennett, D. (1991). *Consciousness Explained*. New York, Little Brown & Co.
- Dennett, D. (1994). Labeling and learning (Commentary on Clark & Karmiloff-Smith 1994). *Mind and Language*, 8, 54-48.
- Dennett, D. (1995). *Darwin's Dangerous Idea*. New York, Simon & Schuster.
- Diaz, R. and L. Berk, Eds. (1992). *Private Speech: From Social Interaction to Self-Regulation*. New Jersey, Erlbaum.
- Elman, J. (1993). Learning and Development in Neural Networks: The Importance of Starting Small. *Cognition* 48: 71-99.
- Fodor, J. (1975). *The Language of Thought*. New York, Crowell.
- Fodor, J. (1987). *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*. Cambridge, MIT Press.
- Gauker, C. (1990). How to Learn a Language like a Chimpanzee. *Philosophical Psychology* 3(1): 31-53.
- Hutchins, E. (1995). *Cognition in the Wild*. Cambridge, MA, MIT Press.
- Jackendoff, R. (this issue). *How Language Helps us Think. Pragmatics and Cognition*.

- Karmiloff-Smith, A. (1979). *A functional approach to child language*. Cambridge University Press.
- Karmiloff-Smith, A. (1986). From meta-process to conscious access. *Cognition*, 23, 95-147.
- Karmiloff-Smith, A. (1992). *Beyond Modularity: A Developmental Perspective on Cognitive Science*. Cambridge, MIT Press/Bradford Books.
- Kirsh, D. (1991). When is information explicitly represented? In P. Hanson (ed), *Information thought and Content*. Canada: UBC Press.
- Kirsh, D. and P. Maglio (1992). *Reaction and Reflection in Tetris*. Artificial Intelligence Planning Systems: Proceedings of the First Annual International Conference AIPS 92, San Mateo, CA, Morgan Kaufman.
- Kirsh, D. (to appear). The Intelligent Use of Space. *Artificial Intelligence*.
- Landi, V. (1982). *The Great American Countryside*. London: Collier MacMillan.
- MacDonald, C. and MacDonald, G. (Eds) (1995). *Connectionism: Debates on Psychological Explanation*. Oxford: Blackwell.
- Marr, D. (1982). *Vision*. San Francisco, CA, W.H. Freeman.
- McClamrock, R. (1995). *Existential cognition*. University Chicago Press.
- Rumelhart, D. and J. McClelland (1986). On learning the past tenses of English verbs. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*. D. Rumelhart and et al. Cambridge, MIT Press. 2: 216-271.
- Tomasello, Kruger, et al. (1993). Cultural Learning. *Behavioral and Brain Sciences* 16: 495-552.
- Vygotsky, L.S. (1986). *Thought and Language (1962, trans.)*. Cambridge, MA: MIT Press.
- Whorf, B. (1956). *Language, Thought and Reality*. New Jersey, Wiley.