Being There: Putting Philosopher, Researcher and Student Together Again

Andy Clark, *Being There: Putting Brain, Body, and World Together Again.* Cambridge, Mass: MIT Press, 1997. Pp. xix + 269. US$25 HB.

By C.A. Hooker

A SLOW revolution in cognitive science is banishing this century's technological conception of mind as disembodied pure thought, namely a material symbol manipulation, and replacing it with next century's conception: mind as the organisation of bodily interaction, intelligent robotics. Here is Clark:

> Intelligence and understanding are rooted not in the presence and manipulation of explicit, language-like data structures, but in something more earthy: the tuning of basic responses to a real world that enables an embodied organism to sense, act and survive . . . it is now increasingly clear that the alternative to the "disembodied explicit data manipulation" vision of AI is not to retreat from hard science; it is to pursue some even harder science. It is to put intelligence where it belongs: in the coupling of organisms and the world that is at the root of daily, fluent action. (p. 4)

© AAHPSSS, 1998. Published by Blackwell Publishers, 108 Cowley Road, Oxford OX4 1JF, and 350 Main Street, Malden MA 02148, USA.
Copyright © 2000. All rights reserved.

And again:

> In the natural context of body and world, the ways brains solve problems is fundamentally transformed . . . Faced with the problem of how to get a computer-controlled machine to assemble tight-fitting components, one solution [by Pure Thought] is to exploit multiple feedback loops [together with a complex fitting algorithm] . . . The solution by Embodied Thought is quite different. Just mount the assembler arms on rubber joints . . . the computer can dispense with the fine-grained feedback loops [and algorithm], as the parts 'jiggle and slide into place as if millions of tiny feedback adjustments to a rigid system were being continuously computed'. This makes the crucial point that treating cognition as pure problem solving invites us to abstract away from the very body and the very world in which our brains evolved to guide us. (p. xii; embedded quote from D. Michie and R. Johnson, *The Creative Computer.* Penguin: 1984, p. 95.)

OK, there's the broad issue in Clark's focus. It is scientifically challenging and technically important—just ask the cosmonauts in Mir trying to safely dock supply ships. But is there any philosophical significance to it.? Well, yes; lots. Clark continues:

> Might it not be more fruitful to think of brains as controllers for embodied activity? That small shift in perspective has large implications for how we construct a science of the mind. It demands, in fact, a sweeping reform in our whole way of thinking about intelligent behaviour. It requires us to abandon the idea (common since Descartes) of the mental as a realm distinct from the realm of body; to abandon the idea of neat dividing lines between perception, cognition and action [AI's triune machine method: receive and transduce (perception), devise algorithm (intelligence), reverse-transduce and execute (action)]; to abandon the idea of an executive center where the brain carries out high-level reasoning; and most of all, to abandon research methods that artificially divorce thought from embodied action taking. (pp. xii–iii.)

The revolution in conceptual foundations, basic principles and research methods is profound, and understanding it, assessing it, even contributing to it is the very stuff of creative, exciting philosophy. Clark has been getting ready to tackle a theme on this scale for more

© AAHPSSS, 1998.

Copyright © 2000. All rights reserved.

than a decade now.[1] But where to begin? There is no neat paradigm to examine.

The idea is young scientifically, conceptually immature, technically and institutionally disunified. The domain covers a bewildering array of disciplines as diverse as robotics and Continental philosophy, artificial life and neurophysiology, evolutionary psychology and economics, PDP computation and linguistics.[2] The domain is overlaid with vague intuitions (Heidegger's hammer, Merleau-Ponty's mouse), sweeping anti-AI pronouncements (Brooks' MIT robotics rubric), near-religious dedication to specialist technologies;[3] research agendas range from the conservative desire to reconstruct symbolic algorithms within the new devices to the complete banishment of the old apparatus—symbols, representations, computations, the lot—in favour of just dynamics.[4] (Not that all these folk are not pioneers.) Hurrah, I thought. I have had similar leanings for years, and made my own line of approach to this thicket through control organisation.[5] My spirit rose in anticipation at what Clark might offer.

And deliver he does. This is a magnificent synthesis of some of the central animating concepts and principles of the new approach; lucid, perspicuous, gloriously eclectic and illuminatingly synthetic, it is a wonderfully told tale that brings the domain into focus. I can't wait to get it into the hands of my students, philosophy, psychology, control engineering, the lot. There's an intellectual square meal for everyone in it, and so many tantalising leads running off that any student—anyone of ability and love of understanding—will find more to chase than time permits the catching.

The real strength of the book is synthesis, and its particular contribution, to my mind, is the bringing into clearer focus some important principles heretofore scattered in different guises across various disciplinary practices. A good example is what Clark calls the *scaffolding* principle: exploiting environmental order through sequential construction and intra-communal interaction in order to augment intelligence. Termites build complex social structures operating complex material mounds, complete with air conditioning, repair and defence services, departments of food supply and reproduction, and so on—a cunning, remarkably city-like contrivance. But where is the central planning department, and where the grand plan for this? There isn't one. Instead, the whole is the emergent outcome of all the interactions among individual termites, each following very much simpler rules. They are little 'TO DO' devices, responding to simple cues with the next thing to do. Among things to do is the dropping of little balls of salivered earth near others, the result being the construction of the walls and arches that

© AAHPSSS, 1998.
Copyright © 2000. All rights reserved.

eventually constitute the mound. As the balls pile up the termite response alters accordingly, no longer simply piling but also walling, filling in, finishing off the surface—in short, they react to their own alteration of their environment. Although the reactions remain simple, they result in a complex sequence of interchanges that constitutes the step-wise creation of their magnificent cities. Their spatially and socially stratified roles only emerge within, and make sense within, that construction. In this way they have scaffolded themselves from little defenceless creatures of meagre cognitive abilities into resilient, viable communities, their abilities magnified by taking organised advantage of the constructive properties of earth.

Sound familiar? Like cities, language, computers, and tech colleges for human termites? Clark explores contexts ranging from cockroaches to economic rationality, from robot 'insects' to language, to bring into focus the common principle of scaffolding: exploiting environmental structure and "our ability to actively restructure that environment so as to better support and extend our natural problem-solving abilities" (p. 32). It is fundamental to evolution which, *ceteris paribus*, prefers KISS solutions (Keep It Simple Stupid), because internal organisation can remain simpler and yet complex results be produced—our whole genetic organisation and embryogenesis is a monument to the principle (though Clark doesn't discuss the sub-cognitive). And it is fundamental to understanding the organisation of intelligence. You can construct a simple photo-tropic robot 'moth' by driving its left-side wheel motors from a right-side light sensor and vice versa.[6] Here the law of variation of light intensity is reflected in the crossed-wire design; such a subtle, cross-categorial exploitation of (electromagnetic) environmental order, yet producing an effective outcome with so simple an internal design. How could you expect to understand the principles of neuronal organisation and functional strategy without understanding the dynamics of body-in-environment in this way?

We are fundamentally very complexly organised moths or termites. Consider the research Clark cites showing that only in the right institutional environment, constructed for the purpose, do individuals display anything approaching economic rationality (cf. a termite's behaviour outside of a mound) but, conversely, more than 70 per cent of market efficiency can be explained by traders using very simple 'mindless' trading rules. (More recent experiments show that more complex traders, using trend-tracking rules, will only settle into an ordered market in sufficiently slowly changing, damped-response markets and that they otherwise track each other's tracking trends to produce the locally turbulent market fluctuations we currently

© AAHPSSS, 1998.

Copyright © 2000. All rights reserved.

experience.) There is a pretty discussion of language from the same perspective, rooted in pioneers like Vygotsky, but more general and able to raise fresh pointed questions for the old innate/learned debate. Here is Clark on us:

> These [scaffolding] strategies are especially evident in child development. Less obvious but crucially important factors include the constraining presence of public language, culture, and institutions, the inner economy of emotional response, and the various phenomena relating to group or collective intelligence. Language and culture, in particular, emerge as advanced species of external scaffolding 'designed' to squeeze maximum coherence and utility from fundamentally short-sighted, special-purpose, internally fragmented minds . . . The Rational Deliberation turns out to be a well camouflaged Adaptive Responder. Brain, body, world and artefact are discovered locked together in the most complex of conspiracies. (pp. 32–3.)

This is terrific stuff—the synthesis across disciplines, the ease and clarity of writing. I do not want to detract from the achievement, and especially not its timeliness. But it is a tale so well told that it is also easy to miss what is not said, or said too glibly. Take scaffolding: once we understand that creatures and their environment can enter an interactive dance that re-makes both, giving both new explicit order and organisation and perhaps higher order capacities as well, how do we specify the resulting components? What are the limits to this? When does a new organism, superorganism or species result? What is the difference between a termite mound, a slime mould in aggregated sporing phase and a primate colony that has just learned a new co-operative defence trick? These are not trivial questions for a species rapidly turning its planet into a single, wired, information-processing mound. Clark does have a short discussion about one aspect of this, the 'leakage' of mind into the environment, in the last substantive sub-section of the book, but no real principle of individuation emerges.

Perhaps there isn't one, but if there were, I think it would begin with the key term from Clark's title for his first chapter, but which he never analyses: "autonomous agent". We are offered a 'general image' of autonomous agents: "a creature capable of survival, action, and motion in real time in a complex and somewhat realistic environment" (p. 6). But no more; after that there's plenty of useful description of robots designed under this rubric, and thought provoking that is, but we learn nothing more about autonomy itself or what role it might play in grounding

© AAHPSSS, 1998.
Copyright © 2000. All rights reserved.

intelligent capacities. That something more might be in order is indicated by Clark's own characterisation above, a curious one for someone who wants to re-unite body and mind. For, contrary to what is suggested, surely motion is among a creature's actions, not an extra item. Sure, effective movement is currently hard for us to implement in our simple extant robot technologies; it is a fundamental capacity, and it can be implemented pretty much independently of other capacities—but this last does not justify its separation from other action, it just reflects our limited robotics; the capacity itself would not have been biologically basic if hunting and hiding were not so selectively advantageous (as witness plants, which are still viable systems).

Again, contrary to what is suggested, action is an intrinsic component of surviving, not an additional item. An organism must so interact with its environment (and within itself) that it acquires the resources necessary for viability (for cellular repair, locomotion, heating, etc.). For this purpose it needs perception-organisation-action-feedback cycles on which it can obtain successful closure. You cannot meaningfully extract one component from these—as Clark himself elsewhere insists (see e.g. p. xiii, quoted above). But getting all the necessary interacting cyclic closures right so as to add up to a coherent creature is not simple—indeed, it is one of the most profound constraints on biological organisation generally, let alone on that of intelligent systems.

Clark misses this issue, but it returns to haunt him. He complains that purely dynamical models of mind often leave us with "an impoverished understanding of the adaptive role of components" (p. 101), but in fact just this is the result of ignoring autonomy. Independently viable systems have to be autonomous in the above sense to at least some minimal degree. These organisational requirements provide a set of constraints, further to those imposed by environmental selection, on biological dynamics, constraints which are suppressed in the standard selectionist models. But autonomy constraints are crucial to understanding the structure of adaptive strategies available to an organism type because the required modifications must be so organisationally coherent that autonomy is preserved. These factors are key, for example, to explaining the difference between genetic variety coupled with shallow organisation and behavioural adaptability coupled with deep organisation but genetic uniformity, as divergent adaptive strategies, the latter alone leading to intelligence. The robots Clark mentions are not autonomous, they do not have closure over any essential function and no self-control, though they share some of the same general functional features as autonomous systems and (most importantly) their construction is moving in that direction. (A clearer, if isolated, functional autonomy was actually possessed by

© AAHPSSS, 1998.

Copyright © 2000. All rights reserved.

Grey Walter's 1950's 'turtle', which could search for and find its own power source, though it moved more crudely than the newer models, and plug itself in to recharge its batteries; it at least had one complete, essential process closure.)

I think it clear that the book cries out for an analysis of the concept of autonomy to undergird it. And while this is not the occasion to expound my own views, as just hinted, I believe that autonomy is also the right organisational constraint on which to ground a truly embodied account of intelligence. Very briefly, the internal organisation and control autonomy requires can be elaborated into a platform for intelligent capacities. Autonomous systems beyond some minimum level of complexity have an intrinsic tendency to adaptability, that is, to the capacity to adapt adaptations, because they must be capable of coherent sequencing and modification of actions. And such autonomous, adaptable systems are intrinsically anticipative. Their functionalities imply that their environmental (and internal) actions anticipate responses that will support those properties. Hunting is feed forward action anticipating subsequent eating and satiation signals. Anticipative feedforward is fundamental to all self-controlling systems, it combines with error-corrective feedback to deliver powerful learning and response capabilities. And thus we arrive at Autonomous, Adaptable, Anticipative systems (AAA systems), which already show all the hallmarks of intelligence. They display complex internal control of anticipative response, conditionalising it on many subtle signals and, to the degree their control architecture is coherently adaptable, they are able to modify it and thus learn. Thus cognitive capacities are grounded in the organisational and control capacities of AAA systems.[7] This is not an account of intelligence that involves a return, with Clark, to "computation and representation" as explanatorily fundamental (p. 101). Autonomy is a dynamically grounded organisational property, but it does involve, as Clark argues, going beyond mere dynamical pattern formation.

This is just part of the story that underlies Clark's elegant synthesised surface, and needs developing. There needs to be (and are) related analyses of action, semiotics for control, epistemics and error correction, semantics and off-line emulation, and so on. Some of these analyses are partially embryonic in the book; for example, an account of semiotic signal information as defined by the modifications it produces (rather than by its sender, the traditional account) is implicit in note 42 to Chapter 8—though the reference to control, the part I would espouse, is there clouded by appeal to some independently characterised 'representation system'. Other analyses remain unnoted; for example, emulation is briefly mentioned, but its significance left undiscussed.[8]

© AAHPSSS, 1998.
Copyright © 2000. All rights reserved.

None of this is criticism in the sense of attributing blunder. One could dispute the whole approach, but that would be to ask Clark to write a different book (and anyway, I am far too much in agreement with him). One could dispute mere details, but Clark's account is too balanced and clear for that to be profitable here. There is the occasional editing blemish—redundant notes (e.g. 13 and 38 to Chapter 8) and several mangled bibliographical entries. But overall a splendid and timely work, whatever your theoretical proclivities. Enjoy.

Department of Philosophy, University of Newcastle,
Newcastle, New South Wales, Australia.

1. See Clark's earlier MIT Press books *Microcognition: Philosophy, Cognitive Science, and parallel Distributed Processing* (1989) and *Associative Engines: Connectionism, Concepts, and Representational Change* (1993).
2. A typical recent effort is Beer *et al.*, *Biological Neural Networks in Invertebrate Neuroethology and Robotics.* New York: Academic Press, 1993.
3. Such as Paul Churchland's splendid neural netist, *The Engine of Reason, the Seat of the Soul.* Cambridge, Mass.: MIT Press, 1995.
4. Try *Mind as Motion* (1995), edited by Port and van Gelder, or Smith and Thelen's *A Dynamics Systems Approach to the Development of Cognition and Action* (1993: yes, both published by MIT Press).
5. See Hooker, *Reason, Regulation and Realism*, Albany: SUNY Press, 1995, at the level of science itself as an intelligent organisation; and also papers in *Topoi 11*, 71 (1992) and chapter 14 of W. O'Donoghue and R. Kitchener (eds) *The Philosophy of Psychology.* London: Routledge, 1996.
6. See the tantalising *Vehicles* by Braitenberg, Cambridge, Mass.: MIT Press, 1984, not mentioned by Clark.
7. For some more details see Hooker, *Reason, Regulation and Realism*, including analysis of Piaget who made endogenous control explicitly central in the 1950s, and W. Christensen and C.A. Hooker in *Evolution and Cognition, 3* (1997), 44. Wayne Christensen, a PhD student and member of the dynamical systems research team at Philosophy, Newcastle University, has contributed as substantially as I to the story summarised here.
8. Compare his colleague Grush's nice linkage of it to representation in *Philosophical Psychology, 10* (1997), 1. Could off-line emulation be the intended source of Clark's representation?

© AAHPSSS, 1998.

Copyright © 2000. All rights reserved.

By Gerard O'Brien

PERHAPS it's a mark of the sheer vitality of the relatively young field of cognitive science that it is grappling with its third major paradigm in the space of just thirty years. While the roots of the discipline can be traced back to the 1960s, its real beginnings in the early 1970s involved the application of ideas derived from conventional digital computers to human cognition, spawning the now appropriately named *classical* computational theory of mind: the doctrine that cognition is a species of symbol manipulation. Then, in the mid-1980s, the field witnessed its first major shake-up with the advent of neurally inspired, parallel distributed processing (PDP) computational models, which substituted operations over activation patterns for symbol manipulations, and many theorists in the field started talking passionately about *connectionism*. Now, scarcely ten years later, the field is once again in tumult, this time with the arrival of *dynamical systems theory*, which, because it eschews the concept of representation, threatens to create an even greater rift in the field than that which occurred between connectionism and classicism.

It is in this revolutionary milieu that Andy Clark's latest book *Being There* is situated. Clark rose to prominence through his advocacy of connectionism, with his two previous books (*Microcognition* and *Associative Engines*) containing some of the most penetrating philosophical work to be found on this alternative approach to the mind. But Clark, who might have expected to spend a few more years developing connectionism in a relatively stable intellectual environment, now finds himself defending it against the even newer dynamical systems vision of cognition.

Clark's response to this predicament is to preach ecumenism. Just as *Microcognition* argued that we shouldn't throw out all the classical insights as we stampede towards connectionism, *Being There* puts the case for combining the embodied, embedded aspects of cognition highlighted by dynamical models, with the commitment to representation, and hence computation, that we find in connectionism (and classicism, for that matter). This is a sensible position, in my view. And there is much to admire in Clark's latest book. He is a gifted expositor, and *Being There* is brimming with detailed and entertaining discussions of the new light that dynamical systems theory is throwing on the role played by both the body and the environment in shaping cognitive processes. At the same time he doesn't shy away from providing incisive critiques of the excesses of this programme, especially when these bubble over into what Clark terms the "Thesis of Radical Embodied Cognition", the claim that

© AAHPSSS, 1998.
Copyright © 2000. All rights reserved.

"embodied cognition is best studied by means of noncomputational and nonrepresentational ideas and explanatory schemes" (p.148). His point here is that such radicalism is unjustified and counter-productive, inviting competition between dynamical and computational conceptions of cognition where progress is more likely to be achieved through cooperation (see especially Chapter 8).

While there is much in *Being There* that I like, the nature of review symposia forces the commentator to look for points of discord rather than concurrence, simply because disagreements are bound to more interesting and response-provoking. In what follows, therefore, I will focus on the one major ecumenical theme propounded in *Being There* that I find difficult to accept. This is Clark's advocacy, especially in the third and final part of the book, of the *extended* nature of the embedded, embodied mind.

Talk of the mind leaking out of the brain and into the world is in the air these days. In philosophy it's primarily driven by externalist theories of mental content, which hold that the meaning of some mental states is determined by the causal relations that internal brain states bear to extrinsic, environmental factors. But Clark is quite explicit that his motivation is quite different (see especially note 23, p. 246). For him the seepage of the mind into the environment is licensed by the subtle couplings between the brain and aspects of the environment, so emphasised by dynamical systems theory, that make it reasonable to suppose that certain extra-bodily resources play a *constitutive* role in some cognitive operations. This kind of extension is most plausible, he thinks, "in cases involving the external props of written text and spoken words, for interactions with these external media are ubiquitous . . . reliable, and developmentally basic" (p. 214). And his conclusion is that in such cases, "what we commonly identify as our mental capacities may . . . turn out to be properties of the wider, environmentally extended systems of which human brains are just one (important) part" (p. 214).

Clark is well aware that, without qualification, this thesis is in danger of foundering on the reef of common sense—the distinction between my mind and yours should not be allowed to collapse "just because we are found chattering on the bus" (p. 217). So there must be principled ways of isolating those external props that become part of the mind from the absolutely vast number that don't. Some of the constraints Clark suggests here are that the requisite information must be "easy to access and use", "automatically endorsed", and "originally gathered . . . by the current user" (p. 217). His favourite example is that of a notebook, which is our constant companion, and in which we make all manner of scribblings. The crucial point in such a case, he argues, is that "the entries

© AAHPSSS, 1998.

Copyright © 2000. All rights reserved.

in the notebook play the same explanatory role, with respect to the agent's behaviour, as would a piece of information encoded in long-term memory" (p. 218). It is principally this "functional isomorphism" that licenses his contention that our "beliefs, knowledge, and perhaps other mental states now depend on physical vehicles that can (at times) spread out to include select aspects of the local environment" (p. 218).

In making these claims about the mind's extension beyond the skin and skull, Clark is opting for one of the two traditional ways of distinguishing between the mental and the merely physical. One way is to suppose that *consciousness* is the mark of the mental, and hence determines the extent of the mind. But Clark thinks that conscious experience is fully explained by the current state of the brain, so there is no basis here for any mind expansion (see pp. 216–17). The other way, on which Clark relies, is to focus on the property of *intentionality*, whereby mental states possess the property of aboutness or, in the language of cognitive science, representational content. The mind's boundaries, according to this second approach, are drawn around the representational vehicles it manipulates in the course of cognition. And so intimate is the causal commerce between human brains and certain written and spoken words, according to Clark, that these external artefacts themselves constitute part of the mind's representational substrate.

But I'm not convinced. It's not that I object to the general criterion by which Clark seeks to include these representational vehicles in the mind (namely, that they are functionally isomorphic with those that standardly encode information in long-term memory). The problem, as I see it, is that, at least in the context of a broadly connectionist understanding of cognition, even his best examples fail to satisfy this condition. No matter how vigorous the causal commerce between parts of my mind and information I record in a personal notebook, these external symbols do not have the same causal properties as the representational vehicles responsible for my memories. To see this, it's necessary to step back somewhat and rehearse some of the now fairly familiar details of the mind's information coding and processing capacities, as these are understood from a connectionist perspective.

It's commonplace for theorists to distinguish between *explicit* and *nonexplicit* forms of information coding in a computational device. Representation is typically said to be explicit if each distinct item of information in the device is encoded by a physically discrete object. Information that is either stored dispositionally or embodied in a device's primitive computational operations, on the other hand, is said to be nonexplicitly represented. It is reasonable to conjecture that the brain employs these different styles of representation. Connectionists make

© AAHPSSS, 1998.
Copyright © 2000. All rights reserved.

much of this distinction by pointing to the two different ways in which information is coded in PDP networks and hence, by extension, in the brain's neural networks.

The representational capacities of PDP systems rely on the plasticity of the connection weights between their constituent processing units. By altering these connection weights, one alters the activation patterns the network produces in response to its inputs. As a consequence, an individual network can be taught to generate a range of stable target patterns in response to a range of inputs. These stable patterns of activation are semantically evaluable, and hence constitute a form of information coding. What is more, because these patterns are physically discrete, structurally complex objects, which each possess a single semantic value, it is reasonable to regard the information they encode as *explicitly* represented.

While activation patterns are a transient feature of PDP systems, a 'trained' network has the capacity to generate a whole range of activation patterns, in response to cueing inputs. So a network, in virtue of its connection weights and pattern of connectivity, can be said to *store* appropriate responses to input. This form of information coding constitutes long-term memory in PDP systems. Such long-term storage of information is *superpositional* in nature, since *each* connection weight contributes to the storage of *every* stable activation pattern (every explicit representation) that the network is capable of generating. Consequently, the information that is stored in a PDP network is not encoded in a physically discrete manner. The one appropriately configured network encodes a *set* of contents corresponding to the range of explicit tokens it is disposed to generate. For all these reasons, a PDP network is best understood as storing information in a *non-explicit* fashion.

These facts about information coding in PDP systems have major consequences for the manner in which connectionists conceptualise cognitive processes. Most importantly, information that is non-explicitly represented in PDP networks need not be rendered explicit in order to be causally efficacious. This is because it is a network's connection weights and connectivity structure that is responsible for the manner in which it responds to input (by relaxing into a stable pattern of activation), and hence the manner in which it processes information. There is a strong sense, therefore, in which it is the non-explicit information in a network (i.e., the network's 'memory') that governs its computational operations: *all* the information that is encoded in this fashion is causally active *whenever* that network responds to an input. The causally holistic nature of information processing in PDP systems is the reason that many theorists think that connectionism provides us with a hint as to how

© AAHPSSS, 1998.

Copyright © 2000. All rights reserved.

Nature might have solved the infamous frame problem. From a connectionist perspective it's possible to envisage how, whenever we act in the world, a very large amount of information could be automatically and unconsciously guiding our behaviour.

But these same facts about connectionism would appear to be destructive of Clark's attempts to extend the mind beyond the skull. It is quite clear that the information encoded in the form of symbols in a personal notebook doesn't have these causal properties, and hence isn't functionally isomorphic with the information contained in our long-term memory. There are two important points of difference here. The first is that the external symbols are *causally passive*: the information they encode doesn't do any work unless we bring them under the gaze of our perceptual equipment. At this point the recorded information does become causally active, but only because it is now *re-coded* elsewhere—namely, inside our skulls. The second difference is that such externally recorded information, when it does become causally engaged with parts of the mind, does so only in a *causally discrete* fashion: each separate piece of information, coded by a distinct symbol structure, must be individually accessed and processed. These differences would thus seem to mark an important natural boundary; one that makes it hard to justify, even on Clark's own terms, any extension of the mind's representational substrate to include our written and spoken words.

Incidentally, this talk of the causal passivity and discreteness of external symbols should call to mind one of the of-cited differences between connectionism and classicism. One of the reasons that classicism presents a very different picture of cognition from connectionism is because it holds that information in long-term memory, unlike that stored in a PDP network, is just like the information recorded on a piece of paper, as it must be discretely accessed by some processing mechanism before it can causally influence ongoing cognitive operations. (This is not to say that classicists are committed to the view that all long-term memories are stored explicitly. In fact, given the sheer bulk of information that is stored in the brain, classicists are committed to the existence of highly efficient, generative systems of information storage and retrieval, whereby most of our knowledge can be readily derived, when required. But such information, while stored in a non-explicit form, must first be rendered explicit before it can be causally effective.) So Clark's case for extending the mind across those symbols inscribed in various external media would be much stronger in the classical context. But this just serves to highlight why it is a mistake to enlarge the mind's boundaries in this way. As many theorists have argued, it is precisely because classicism is committed to this account of memory and information processing (that is,

© AAHPSSS, 1998.
Copyright © 2000. All rights reserved.

because it is committed to the code/process divide—see for example Clark's *Associative Engines*) that the infamous 'frame problem'—the problem of equipping a cognitive system with the wherewithal to choose appropriate courses of action, in response to internal goals and changing environmental conditions, in real time—is so acute for this approach to cognition.

Department of Philosophy,
University of Adelaide,
Adelaide, South Australia,
Australia.

By Naomi Quinn

MOST cognitive anthropologists' thoughts are far from questions about human evolution. Why this is so I am not sure; perhaps it is closer proximity, historically, to other theoretical traditions and preoccupation with the research questions, intellectual debates and methodological issues they raise, augmented by a distrust of biological explanation (coming as we do from a parent discipline where even psychology is suspect for its truck with 'human nature'). Many of us study cultural understandings, often called *cultural models*, and increasingly now conceived of as shared cognitive schemata; and while a number of us explore what these shared schemata do, it has not much concerned us that they had to have evolved to do it, nor what the consequences of their evolution might be for the form they take. Andy Clark's rich synthesis invites us to begin rethinking our enterprise in terms of culture's coevolution with, and role in, human cognition, and it offers many applicable lessons from the design of robots and the evolution of various organisms. Perhaps the most general lesson I took away—to redeploy a distinction of Clifford Geertz's that is familiar to nearly all anthropologists—is that 'models of' always have their origins as 'models for' or, as Clark would put it, as "local and action-oriented" internal representations (pp. 49, 149). An equally important second lesson is that what these internal representations do is to provide 'scaffolding' that compensates for what the unaided human brain cannot do well (p. 68).

Yet, when Clark writes of internal representations that provide the brain with scaffolding, he is not thinking of culturally provided ones which, for him, are *external.* To be sure, anthropologists will be gratified by the expansive part culture will ultimately prove to play in Clark's

© AAHPSSS, 1998.

Copyright © 2000. All rights reserved.

formulation of the brain-body-world interaction. "The present discussion", he declares, "barely scratches the surface of a large and difficult project: understanding the way our brains both structure and inhabit a world populated by cultures, countries, languages, organisations, institutions, political parties, e-mail networks, and all the vast paraphernalia of external structures and scaffoldings which guide and inform our daily actions" (p. 191; see also pp. 186–7). Nevertheless, by invariably casting culture and the scaffolding it provides as external, as he does in this passage, and by granting the pride of place he does to one external component of culture—"language: the ultimate artefact" (Chapter 10)—his discussion may inadvertently close off a significant part of the project he envisions.

I read this book with a growing sense that it was the book I had been looking for. For more than a year now, I have been musing over the evolutionary implications of some findings of mine, and searching for an evolutionary framework in which to put them. I would like to use my turn in this symposium to bring my material to bear on Clark's treatment of culture, and to see if he will agree to the extension of culture's role in human task solutions that I propose, an extension that seems to me to be very much in the spirit of his argument. What I will describe are unspoken, internal cultural representations that mediate performance of two everyday cognitive tasks. The particular task solutions I report are less important in and of themselves, than for the much more extensive class of such cultural solutions as-yet-unidentified that I believe these two to represent.

By calling these internal representations 'cultural' I mean to convey that they are shared across groups of people and that they come to be shared largely by being learned, just in the way that language is. Indeed language is cultural, too; but it is only a part of our cultural equipment. Although I reconstructed the cultural task solutions I will describe from their use in speech—discourse from extended informal interviews with Americans about their marriages—and although language is certainly implicated in their use as well as vital to their transmission, I want to stress that these internal representations themselves are extra-linguistic. They are not to be equated, either, with inner speech (p. 197). More generally, though they are cultural, they are not external artefacts.[1] Nor are they mental simulations of external-world manipulations (p. 61).

Being automatic and out of awareness, hence unarticulated, and being otherwise unmediated by external artefacts, these culturally shared internal representations are so invisible that other people will be as surprised as I was to discover that we have been using them all along. Which makes it understandable that this class of problem solutions has gone unrecognised and unstudied by comparison to both the tasks that

© AAHPSSS, 1998.
Copyright © 2000. All rights reserved.

language performs, and the artefact-supported and socially distributed task solutions that anthropologist Edwin Hutchins studies—both reported on in *Being There*. I want to argue that the internal representations I will describe are as much a part of what Clark terms (e.g., pp. 32–3) the cultural 'scaffolding' that supports and extends our natural problem-solving abilities, as are language and physical artefacts, the importance of which he rightly emphasises. These internal cultural representations further complicate the mind–world crossings that Clark wants us to consider, by illustrating that, not only has mind "spread itself out into the world" (p. 68), but the world, in the form of culture, has seeped into the mind.

Both tasks are frequent and recurrent, making it understandable why cultural solutions to them have evolved, and why, in Clark's terms (pp. 216–18), these solutions should be 'portable' ones. The first task is a communicative one, that of clarifying what we mean to say to our audience by use of metaphor. To clarify what I mean, I will give the following two wholly typical examples.[2] The first is from an interview in which a husband was describing a moment early in his marriage, when the difficulties his wife and he were confronting made him realise "that my confidence in the everlasting Gibraltar nature of the whole thing was rather naive". This speaker not only captures an expectation he shares with other Americans that marriage be lasting—as we will soon see, an important piece of their internal representation of it; he also underscores how, in his naivete, he overestimated the lastingness of his marriage, and he does so by reference to something that Americans have come to know as an icon of the everlasting—appropriated, reproduced, and widely disseminated as the logo of a national insurance company.

The second example is also a thoroughly American one, culled from the sports page of *USA Today* (May, 1993). Third baseman George Brett, interviewed on the occasion of his retirement from baseball, comments on what has been, in today's game, a remarkably long-term relationship with the Kansas City Royals: "I compare it to a marriage. We've had our problems, but overall, we have had a good relationship. I never, ever want to put on another uniform." Marriage is famous among Americans as something that is meant to endure and that does so (when it manages to do so) because it is rewarding in spite of its difficulties. That is why the metaphor of a marriage gives readers a surer sense of what Brett wants to convey about his relationship with his team.

How do usages of metaphor like these actually work? They are not just mappings of one domain on to another, as metaphors are commonly said to be, but particular kinds of mappings. They are references to some point being made about the domain under discussion, in terms of some

© AAHPSSS, 1998.

Copyright © 2000. All rights reserved.

outstanding and unambiguous cultural—that is, guaranteed to be intersubjectively shared between speaker and listener—instance of that feature or property. I call such instances *cultural exemplars.*

This theory of what metaphors do requires that, in order for them to do their work of clarification, members of a speech community must, as they do, share a large stock of cultural exemplars to draw upon. Knowledge of these is accumulated from a variety of experiences, both first- and second-hand. Repeated television viewings of the insurance company advertisement featuring the Rock of Gibraltar are one obvious example of this experience. Crucial is the ongoing experience of hearing and using metaphors in speech, not only because it presents individuals with many more exemplars than could possibly be encountered otherwise, but also because it weeds out more idiosyncratic choices that would be ill-understood by audiences, in favour of more widely agreed-upon cultural exemplars, that communicate well. Through their repeated use as metaphors the latter gain even wider acceptance as good examples, sometimes becoming wholly conventional.

When we need them, these exemplars just come to us, and they do so within the real-time constraints imposed by the action of speaking. Connections, built up from experience, between properties of the world and their known exemplars permit rapid, automatic identification of apposite metaphors. Formulated in this way, the task is one that the human brain is well-equipped to perform. At the same time, the easy accessibility of these cultural exemplars relieves us of the need to specify the meaning of every concept we wish to convey, however unfamiliar, abstract and unlabelled, or otherwise ill-articulated, in its own terms—a cognitive and communicative feat that the brain appears ill-designed to perform. Instead, humans do this task by calibrating the meaning of one thing in terms of another.

The second case I want to try out on Clark is the equally ordinary one of the everyday reasoning we do in our talk. Like clarifying what we mean to say, this reasoning must be accomplished in the real time of speech production.[3] The structures that people use to reason with are idealised event sequences. I will illustrate how such event sequences work by drawing once again on my research on Americans' understandings about marriage. Presumably, such idealised event sequences for reasoning rapidly and readily about significant, recurrent dilemmas have evolved and spread in multiple domains of everyday life.[4]

In order to show how it is used, I must first fill in the specifics of this particular sequence of events; for, as Clark tells us, such internal representations tend to evolve for local purposes and to be "content-bearing" (p. 175). Here, as concisely as possible, is the content of this

© AAHPSSS, 1998.
Copyright © 2000. All rights reserved.

one: Americans expect marriages to be lasting, as we have already seen, and also to be mutually beneficial and shared. Marriages that are not beneficial will not last. This is because Americans understand marriage to be a contractual relationship that must be satisfactory to both parties in order to continue. This potential contradiction between the imperatives of lastingness and benefit poses a central marital dilemma for Americans, and we resolve it in a thoroughly American way: beneficial marriages, and thereby lasting and successful ones, are to be achieved by hard work.

One event is linked to the next by relations of causality, the concatenation of such relations forming a chain of events. This structure is idealised in two ways important for reasoning with it. The first idealisation is that possible events within the structure are highly circumscribed by being limited in number and following each other in invariant order. Marriages that are shared and fulfilling will be mutually beneficial, and marriages that are mutually beneficial will be lasting. Fulfilment is a matter of compatibility, and incompatibility the chief source of marital difficulties which, when unresolved, put marriages at risk of failure and divorce. These difficulties can be overcome, however, with effort, to achieve a fulfilling and hence lasting and successful marriage. Americans share this idealised marital scenario, and use it to reason with.

A second idealisation is that causal relations between pairs of events are themselves highly simplified. In the above reasoning, if a marriage is beneficial it will last, and if it has lasted it is beneficial, categorically. Similarly in other pieces of reasoning, and with causal relations between other events in the sequence. Americans know of complexities, of course, such as unhappy marriages that endure anyway. As well, Americans like these are certainly able to think about their marital relationships in quantitative terms, as the second and third speakers do when they talk about "as much" at stake and "too much" cost. But calculations of quantity are not allowed to complicate their reasoning. Having some unspecified amount at stake in the relationship makes a couple want to keep it going; while some high but unspecified level of cost will leave a spouse unable to stay in it. In the same way, other reasoners ignore all the extenuating circumstances and the shades of degree and probability that they know attend any real-life marital situation, substituting, in place of these complexities, the simplified causality of *plausible inference*.[5] That is, for purposes of reasoning about them, likely events or those susceptible to gradation are treated as if they were always and absolutely true.

This simplification of causality allows speakers to reason as readily in either direction across links in the chain of events—from lastingness to benefit, as does the first speaker, or from benefit to lastingness, as does the second. And it allows them to reason just as readily in the negative—from

© AAHPSSS, 1998.

Copyright © 2000. All rights reserved.

lack of benefit to lack of lastingness, as does the third speaker. Combined with the first idealisation I described, it keeps them from confronting an unmanageable number of potential outcomes, and becoming tangled up in endless complications and nuances. Well-learned, moreover, this task solution affords a further cognitive efficiency. Once all the events in the sequence have become strongly linked to one another, conclusions about the relations between events distant from each other in the causal chain are no harder, and take no more time, for reasoners to reach than conclusions about the relations between events directly causally linked. Thus, for example, speakers reason about the relation between compatibility and lastingness as readily as they reason about benefit and lastingness or compatibility and difficulty. Here, once again, culture seems to work in concert with the brain's connectionist architecture—making real-time reasoning about a familiar problem possible.

The learning of this task solution is overdetermined. Individuals are drawn to it by its utility, by the felt importance of the task, and by the incorporation into its solution of shared understandings about achievement of success through effort which are highly motivating for many Americans.[6] Its appeal also recommends this way of thinking about marriage to those who create more public images of it, which then provide further opportunities for its learning. Even more frequently, people confront this reasoning task in their lives and those of people they know, and talk about it, and hear it talked about. And, in the same way that we must use shared metaphors to be understood, we must reason from a shared framework of assumptions in order to be persuasive.[7]

Both my cases exemplify Clark's dictum (p. 81) that "new cognitive garments seldom are made of whole cloth; usually they comprise hastily tailored amendments to old structures and strategies". In the first case, the old cloth consists of culturally-recognised good examples; in the second case, a culturally shared approach to addressing difficulties of all kinds. In both cases, the 'old cloth' is previously learned cultural understandings.

I have also tried to show how both problem solutions, in their ways, complement the natural abilities of the brain to simplify its work by minimising its internal computational load, just as Clark argues other kinds of mindworld interactions do. More specifically, each of these task solutions would appear to be an example of one of Clark's (p. 167) two classes of "representation-hungry" cases. The first, metaphor, carries information about an otherwise undefined or ill-articulated state of affairs. The second, the event sequence for reasoning, selectively responds to states of affairs that are unified only at an abstract level; what counts as fulfilment, compatibility, or marital difficulty, for example, and how these are experienced, can be wildly various. Says Clark:

© AAHPSSS, 1998.
Copyright © 2000. All rights reserved.

> In the two ranges of cases (the absent and the unruly), the common feature is the need to generate an additional internal state whose information-processing adaptive role is to guide behaviour despite the effective unfriendliness of the ambient environmental signals (either there are none, or they require significant computation to yield useful guides for action). In these representation-hungry cases, the system must, it seems, create some kind of inner item, pattern, or process whose role is to stand in for the elusive state of affairs. These, then, are the cases in which it is most natural to expect to find system states that count as full-blooded internal representations. (p. 168)

Exactly so. And equally so for internal representations that are neither the products of genetic encoding nor the idiosyncratic inventions of individual agents, but have come to be shared by learning.

Department of Cultural Anthropology,
Duke University,
Durham, North Carolina, USA.

1. It is not clear to me whether the symbolic structures representing moon and tide states that successive generations learn in Hutchins and Hazelhurst's simulation, described by Clark on pp. 189–90, would have to be encoded in the external artefacts that Clark seems to think of them as; because of the need for them to be veridical to a complex and detailed state of affairs in the world, they would probably have to be. But this is the closest example I found in the book to what I am talking about.
2. Both of which come from N. Quinn, "Research on shared task solutions", in C. Strauss and N. Quinn, *A Cognitive Theory of Cultural Meaning.* Cambridge: Cambridge University Press, 1997.
3. I am not sure whether the importance of having such internal representations for reasoning about recurrent dilemmas is primarily to do with the real-time constraints of speaking, or whether these structures evolve to support the inner reasoning we do in working out solutions to recurrent problems for ourselves, and are only secondarily applied to the spoken reasoning we do when trying to persuade others of our arguments—or even whether the two contexts can really be separated.
4. Certainly, though, they are not found in all domains. For example, Americans appear to have no shared event structure comparable to that they use to reason about marriage, with which to reason about friendships, when these run into trouble. Apparently, the dilemmas for which solutions evolve are ones that not only recur, but have especial cultural salience and historical longevity.

© AAHPSSS, 1998.

Copyright © 2000. All rights reserved.

5. E. Hutchins, *Culture and Inference: A Trobriand Case Study.* Cambridge, Mass.: Harvard University Press, 1980, p. 56.
6. See N. Quinn, "The case of Americans reasoning about marriage", *Ethos 24*, 1996, and N. Quinn, note 2 above.
7. Think of trying to persuade another American, as one of my interviewees tried to tell me about marriage, "If it's going to work, it's going to work, you know; there's no need going out of your way to do it." (I managed to preserve my interviewer's neutrality, but mentally rolled my eyes.)

By John Sutton

DOLPHINS and bluefin tuna, Andy Clark tells us at the end of this marvellous book, are just not strong enough to swim as fast as they do. Their abilities derive not from great evolved strength but from a capacity to treat their medium as enabling, rather than constraining, their motions. As well as exploiting aquatic swirls and eddies to aid their manoeuvres, they actively create new and useful vortices and pressure gradients in their environment. The world is not a jumble of obstacles against which organisms must struggle, rail, and bump, but a pool of resources with which they interact in a looping, "intricate and iterated dance". When considering the natural, technological, linguistic, and institutional environments in which humans are embedded, Clark argues, we would do well to remember that uncanny fusions of internal and external processes, analogous to the dolphins' use of fluid dynamics, can have a startling productivity which remains invisible to investigation which stops at the boundaries of the skin.

New dynamical approaches to mind and brain are firing strange alliances between roboticists, developmental psychologists, phenomenological philosophers, Artificial Life simulators, neurobiologists, neural network modellers, and followers of the two Gibsons (ecological psychologist J.J. and cyberpunk guru William). Skilfully spying out just what empirical work and which simulations across the disciplines best exemplify the implications, Clark manages all at once to document, temper, justify, and communicate the current excitement. The careful weave of his case, building up and returning to related points across a number of contexts, can obscure the extent to which this is more than a synthesis of the state of play: the book is a sustained argument, running in parallel at many levels, for the view that "advanced cognition depends crucially on our abilities to *dissipate* reasoning . . . Our brains make the world smart so we can be dumb in peace" (p. 180). Clark's developing vision of the mind/brain as a pattern-matching associative engine which in

© AAHPSSS, 1998.
Copyright © 2000. All rights reserved.

"parasitiz[ing] the external world" itself leaks out, the mind "mingling shamelessly with body and with world" (pp. 73, 53), will in turn seep out of its natural confines in the philosophy of psychology and influence theoretical frameworks which begin elsewhere.

Cultural fears of the alleged dehumanising effects of cognitive science are driven by the mistaken assumption that the search for mechanisms of mind must inevitably neglect or (worse) reduce the puzzling, wonderful complexity of psychological life. Wary, wise humanists have been sceptical because of three broad characteristics of most work in cognitive science. First, misplaced physics-envy in many domains of the psychological sciences has ruled out curious or complicating sensitivity to history and culture: the vast gulfs between brains and society are felt in practice even, or especially, by those who deny them in principle. Second, the central assumption of psychological atomism, with independent mental items dully interacting by principles of seventeenth-century mechanics until pulled out of storage by some homuncular central executive, has seemed to rule out the fusions and confusions between thoughts or memories which are pervasive in experience. Third, even when this is challenged by semantic arguments to the effect that meaning can't be in the head, a residual methodological atomism has left much cognitive science treating the individual mind as the largest possible unit of study, with the environment (natural and cultural) treated only as a series of obstacles and a source of input.

Clark actively repudiates each of these aspects of previous cognitive science, and (better yet) demonstrates by example how rich are the alternative explanatory frameworks already in place. Insensitivity to culture is no longer necessary, Clark shows, giving examples from cognitive anthropology (studying group thinking, distributed across agents and tools), and from economics (the role of institutional structures as scaffolding in the rare cases when consumer decisions do approximate the idealised rationality of classical theories): his own sassy pop-culture sensibility exemplifies the new literary norms of elegant science writing, each chapter seguing neatly into the next. Psychological atomism is replaced by connectionists' overlapping *distributed* representations, and by firm rejection of any separation between (passive) data and (active) processor. And, in the key advance Clark garners from the new dynamicism, both body and environment are seen as intrinsically linked or *coupled* with the individual mind/brain, in loops of "continuous reciprocal causation", with brain, body and world as responsive to each other's changing behaviour as are jazz improvisers to each other and to their instruments (p.165). Theorists hostile to cognitive science who previously thought that it must *inevitably* neglect embodiment, or lived

© AAHPSSS, 1998.

Copyright © 2000. All rights reserved.

social experience, or the inarticulable forms of life in which minds are embedded, or the ready-to-hand natural or technological objects with which minds interact, cannot thus just assimilate this latest dynamical trend to the history of computer-driven speculation: those who resist the encroachment of science on mind, fearing that it will swamp human dignity and historical awareness, have perhaps been attacking only a 1950s ghoul, a dreich behaviourism without laughter.

Yet there is also considerable continuity across this apparent epistemological rupture: on a second front, Clark seeks (against radical dynamicists like Port and van Gelder—see review symposium on *Mind as Motion* in *Metascience 11*) to retain a general (functional) notion of *representation* which includes personalised, idiosyncratic, context-specific, action-oriented and distributed representations. There is indeed little point in *defining* 'representation' so as to include only the static, context-independent, action-neutral items postulated in pre-connectionist theory: but I want to pick up on a small point in Clark's argument for 'minimal representationalism' (Chapter 8). He accepts that there may well be some "cases in which the web of causal influence grows so wide and complex that it becomes practically impossible to isolate any 'privileged elements' on which to pin specific information-carrying adaptive roles" (p. 166). But he argues that some such privileged elements *must* be involved in two kinds of "representation-hungry problems", to understand our ability temporarily to *decouple* from the world: we can think about absent, nonexistent, or counterfactual states of affairs (in remembering or imagining, or in considering future action), and about 'unruly' classes of things which aren't naturally grouped together in perception ("all the valuable items in a room . . . [or] all and only the goods belonging to the pope"). Here, Clark wagers, representation-talk will be justified and we will make "relatively fine-grained assignments of inner vehicles [complex brain states] to information-carrying adaptive roles" (p. 168).

But take an example of thinking about the absent: in autobiographical memory, the bits of the world with which I'm in causal contact in an episode of (willed or unwilled) recollection or reminiscence are far, far away. It's not clear that in remembering hot afternoons, or an old anger, I must have created, as Clark suggests, "some kind of prior and identifiable [internal] stand-ins for the absent phenomena" (p. 167). Rather than the distinct 'inner surrogates' he describes, may not the past be carried in just as complex a web of causal influence? *Some* memories of the personal past *may* be stored fully formed, the past preserved in aspic to trouble us in reminiscence, though we currently have poor evidence for such fixity and no idea of plausible mechanisms. But many autobiographical memories

© AAHPSSS, 1998.
Copyright © 2000. All rights reserved.

are, to a greater or lesser extent, condensed summaries of a number of different portions of the past: why should we expect all the psychological work of editing and condensing to have been done already, *before* the episode of remembering, and neatly packaged in a single prior item, rather than occurring in the present in a causal conspiracy between different distributed traces and context-specific cues in current input? I don't think it unlikely, as Clark does, that continuous interplay between internal and external factors, and a concomitant complexity of inner dynamics, may also characterise some of these "cases for which the representational approach is . . . most compelling" (p. 175). No one has a finished account of the metaphysics of distributed representations, in particular of how to individuate the many superposed dispositional implicit representations which can coexist in one representing system. But I wonder if Clark's minimal representationalism might not still be too strong here, if the isolability requirement might only be satisfied at the time of the occurrent remembering, when many quite different fragmentary traces are pooled or actualised together.

The applicability of such an isolability requirement has already been questioned at another level of explanation in Clark's further demonstration of the power of the concept of distribution in the philosophy of neuroscience. In a typically suggestive deployment of empirical evidence, Clark cites challenges to the homuncular vision of the monkey brain as a somatotopic map in which localised groups of neurones control, for example, individual fingers. It turns out that movement of a single finger requires *more* activity across all the neuronal groups of the relevant motor area of the brain, where the localisationist story predicts activity only in the specific neurones which 'control' that finger. Isolated digit movements are not, as classical neuroscience assumed, the basic unit from which others are built up, but the complex case, requiring extra resources to *prevent* the movements of other digits (p. 131). Evidence like this perhaps challenges the very idea of local representation in the brain: of course neurones don't disappear when not 'active', and differences in *any* neuron in a system (not just those thought to 'code' for some outcome) may subtly influence the course of processing. One part of the monkey brain, it seems, may have come to superimpose control on other parts, with the kind of cognitive discipline necessary for the control of isolated movements being an evolutionarily late mechanism of approximating local representation in a fundamentally superpositional, distributed system. Clark repeats the lesson that cognitive order is an achievement, not a given, which emerges rarely and by roundabout means: like Philo in Hume's *Dialogues*, he cannot see "why an orderly system may not be spun from the belly as well as from the brain".

© AAHPSSS, 1998.

Copyright © 2000. All rights reserved.

When we do think straight or plan sequences of action successfully, it's often not because of vast internal computational power, but by relying on the many varieties of external *scaffolding* which serve to transform or break down problems which are too hard for our meek associative processes. Water, kitchen shelves, logical symbols, words, stock exchanges, parents, friends and computers can all aid in cognitive reorganisation, holding in different ways the strategies and training schemes as well as the stored knowledge gained in cultural as well as individual learning. Clark returns repeatedly to questions about the relations between our (partial, contextualised) mental representations and the variety of the external representations with which minds interact. Setting such questions firmly at the centre of cognitive science, seeing interaction with the world as fundamental rather than intrinsic to mentality, is itself a great step towards difficult, exciting interdisciplinarity: the particular histories, to name a few, of cryptograms and codes, of perspective, of autobiographical genres, of diagrams and graphs, of photography, of artificial memory techniques, of map-making, of clothes, of laboratory practices, or of religious ritual now become integral data for a historical and comparative cognitive science rather than humanistic curiosities.

So, we augment our relatively fluid, context-specific mental representations with relatively stable, reusable, context-independent external representations, among which linguistic representations play a key role. But in addition, we then internalise text-like codes to 'minimise contextuality' in our own minds:

> by 'freezing' our own thoughts in the memorable, context-resistant, modality-transcending format of a sentence, we thus create a special kind of mental object—an object that is amenable to scrutiny from multiple cognitive angles, is not doomed to alter or change every time we are exposed to new inputs or information, and fixes the ideas at a high level of abstraction from the idiosyncratic details of their proximal origins in sensory input. (p. 210)

Clark is, however, consistently clear that such static, frozen representations are fundamentally alien to our mind/brains, carefully distancing himself from Dennett's view that the continual bombardment of our neural nets with serial public linguistic input produces a deep transformation in the forms of cognitive processing (pp. 197–8, cf. 63–6).

To put Clark's less comfortable picture in different terms, civilising processes, in both culture and individual, require a kind of self-oppression, in which we have to achieve mastery over our own brains by assimilating

© AAHPSSS, 1998.
Copyright © 2000. All rights reserved.

symbolic 'props and pivots' of a form which is, in a sense, profoundly unnatural. Like the medieval monks who laboriously forced strange architectural memory palaces into their minds so as to keep stored items distinct, to guarantee immunity from the melding characteristic of 'natural' memory, we all impose (an approximation of) rigidity and inflexibility on our own mental representations. As the dolphins teach us, of course, supplements to our bare biology are responsible for many wonderful extensions to our capacities: but Clark's stress on the generality of (at least some of) our learning mechanisms reminds us that the specific cognitive trajectories along which our particular cultural and institutional learning aids allow us to go are, in a way, deeply contingent. Clark's version of dynamical cognitive science foregrounds the action-oriented and path-dependent nature of 'mind on the hoof' (p. 35), and it opens up vast theoretical terrain in which it may be possible to attend to brains and contexts at once.

School of Philosophy, University of Sydney,
Sydney, New South Wales,
Australia

Author's Response

By Andy Clarke

THERE can be few things more satisfying than reading friendly, constructive engagements with one's own work.[1] I thank the four reviewers for their patient and penetrating comments, and for the truly marvellous overviews of the project. The pieces by Hooker and Sutton distil the essence of the project with great and enviable clarity, while all four reviewers push, probe and extend the work in challenging yet helpful ways.

The general idea of *Being There* was to weave a variety of sometimes unlikely looking components into a coherent (but somewhat non-standard) view of natural intelligence: a view in which basic organism/environment coupling is fundamental and in which advanced cognition emerges as deeply continuous with these roots. A major

© AAHPSSS, 1998.

Copyright © 2000. All rights reserved.

element of the story, as noted by several reviewers, was a highly generalised notion of 'scaffolding'—of bodily and environmental structures (including linguistic and cultural artefacts) that re-shape the space of individual reason and thus enable us to press maximal benefit from fragmentary, pattern-completion styles of internal computational organisation.

Such a view, although not mainstream, is certainly not novel. Hooker's own work on control theory, the very substantial literatures of 'new robotics', artificial life and dynamical systems theory, and the more philosophical frameworks of Varela, Lakoff, Johnson and others, are all clear examples of closely related views mentioned in the text. Work in connectionism, cognitive anthropology, education and economics is also invoked and a major goal of the book was to try to coax these various elements together to isolate some unifying themes, and to highlight some problematic issues.

The coaxing together seems to have been largely successful, and the reviewers' appreciative comments warmed my heart on a cold morning. One reviewer (Quinn) goes on to suggest an interesting extension to the set of core elements—a proposal I will return to later. For the most part, however, the critical comments focused on three of the more troublesome issues raised by the text. First, the unexplicated notion of agent autonomy; second, the problematic suggestion that mind might somehow leak out into the surrounding world; and third, the vexed role of internal representation in the explanation of intelligent behaviour. I shall take these in turn, then end by discussing Quinn's proposed extension and some possible future developments.

Autonomy

Cliff Hooker's stylish and engaging commentary highlights an important question—one that is, I confess, not even addressed in the book. I make extensive use of the popular term 'autonomous agent' but say nothing about the nature of the autonomy itself. Worse still, the examples I give of real-world artificial 'autonomous agents' are, Hooker suggests, not really autonomous agents at all, although they do "share some of the same general functional features as autonomous systems".

Hooker's view, as I understand it, is that genuine autonomy involves a special kind of intelligent control of action, what he calls adaptable, anticipative control. Autonomous, Adaptable, Anticipative Systems (AAA Systems) are ones that modify their own responses and routines so as to create and sustain a life (or functionality) preserving coupling with their

© AAHPSSS, 1998.
Copyright © 2000. All rights reserved.

environments. A robot such as Herbert (the soda can collecting robot described in the early pages of *Being There*) is not an AAA System, as its activity is not adaptably geared to maintaining its own functionality. AAA Systems, Hooker suggests, display a type of organisation that goes beyond "mere dynamical pattern formation". If we identify cognitive systems as AAA Systems, then we can see, rather concretely, why cognition involves a special kind of agent–environment coupling.

This strikes me as a good way to go. The strong sense of autonomy that Hooker defines does allow us to mark some important discontinuities in the design space that is being explored by contemporary work in robotics and artificial life. My own guess, however, is that the notion of anticipative, adaptable response is itself still too broad and disunified to mark any rigid boundary between cognitive and non-cognitive routes to adaptive success. Indeed, part of the thrust of *Being There* is to suggest that the cognitive/non-cognitive distinction is itself too coarse a tool to bear real scientific weight. Certain kinds of simple insects and maybe even some plants may well fit the basic image of an AAA System, exhibiting both some degree of learning and of self-modification geared to survival. What we will probably find then (and I have no reason to think that Hooker disagrees with this) is that a lot depends on the different ways in which anticipative, adaptable response is supported. (In a later section, I will comment on one such way: the use of inner circuits to emulate agent/environment dynamics).

On the topic of autonomy, I would also flag Tim Smither's interesting work (e.g., Smithers, ms) which seems to dovetail nicely with Cliff Hooker's. Smithers argues that true autonomy requires a process of "self law-making", not just self-regulation. An example of this would be systems which actively create the kinds of environment (both internal and external) they need in order to function efficiently. Such a notion of autonomy also fits well with the observation, central to *Being There*, that intelligent behaviour often depends on the creation and exploitation of 'external scaffolding'—environmental structures that simplify and reconfigure the tasks confronting biological brains.

In sum, I agree that *Being There* works with a broad and unanalysed notion of "autonomous agent". In my defence, I note that so do most real-world robotics laboratories and that the broad notion (of embodied, usually mobile devices capable of simple real-world real-time activity) does pick out an interesting class of systems. But I agree that a stronger notion of autonomy may help identify important discontinuities in design space (see Sloman 1994). And much of my current work is indeed concerned to fine-tune the story in just these kinds of way (see especially Clark, in press; Clark and Grush, submitted).

© AAHPSSS, 1998.

Copyright © 2000. All rights reserved.

Seepage

Gerard O'Brien approaches me from a different angle, with a deft blow to an acknowledged weak spot: the consistency organ. O'Brien worries about the idea (pursued gently in the book and more vigorously in Clark and Chalmers 1995) that mind may sometimes seep outside the traditional envelope of skin and skull, inhering instead in extended systems comprising the biological brain and selected aspects of the body and local environment. The reason why this doesn't happen, he argues, lies in the different ways in which external and internal components store and organise information: differences that ought to have been especially clear to the author of two books (Clark 1989, 1993) contrasting connectionist and classical modes of information storage and retrieval. (Hence the threat to the consistency organ.)

More precisely, O'Brien argues that external information stores (such as the well-maintained and constantly available notebook featured in Chapter 10 of the book and in Clark and Chalmers 1995) are not plausibly seen as functionally isomorphic to biological long-term memory, at least as depicted by connectionist theory. Such a notebook might indeed be somewhat similar to a classical vision of an inner data-base. But the connectionist vision, with its stress on superpositional information storage (and on associated properties such as free generalisation, content addressability and graceful degradation) paints a quite different picture. If the connectionist story is (as it seems to be) closer to the natural facts than the classical one, then there is indeed a world of difference between the passive discrete symbol structures found in the typical external store and the active inexplicit representations found in the head.

O'Brien depicts my suggestion that mind might seep out into the world as based entirely on a principle of functional isomorphism: if some element outside the head is contributing to behavioural success in a way that is functionally isomorphic to the contribution of some inner, standardly cognitive resources, then it should be seen as part of the cognitive system too. But I think he reads too much into the (perhaps ill-advised) locution of 'functional isomorphism'. For the isomorphism is said to hold only in respect of the explanatory role of the external elements in a commonsense account of the agent's behaviour. The basic idea (developed more fully in Clark and Chalmers 1995) is that the notebook entries explain the same kinds of very broad patterns of purposive behaviour as does knowledge stored in biological memory. To that, O'Brien will reply (I suppose) that the kinds of pattern provided for are really subtly different, perhaps in respect of properties such as generalisation and the like. To which we will reply that these differences leave

© AAHPSSS, 1998.
Copyright © 2000. All rights reserved.

intact a more fundamental similarity concerning the appeal to stored information in the explanation of purposive action.

Such an exchange, however, only gets us so far. A better response to O'Brien's critique is, I think, to see it as identifying a potential tension between two components of the extended mind story itself. One component (the one he focuses on) stresses the way that extra-neural elements can play a role similar to internal ones (as in talk of external memory, etc.). But a second component, which was repeatedly highlighted in the text, turned on the way external elements may play a role different from, but complementary to, the inner ones. It is this vision that is invoked in the discussion of Hutchins' work on the role of maps, compasses and so on in an extended (multi-agent and artefact) ship navigation system: a discussion I explicitly cite (p. 214) in introducing the topic of the extended mind. This same complementarity is foregrounded by the claim that the user–artefact relationship may be as close and intimate as that of the spider and the web (p. 218) and by the analogy (ch. 11) with the tuna's active creation of water-bound eddies and vortices so as to improve its aquatic performance.

Given this second line of argument (the one stressing complementarity), it is best to see functional isomorphism as at most part of a sufficient condition for cognitive extension, rather than as a necessary feature. The more interesting and plausible argument, I feel, is the one which describes the seepage of mind into the world by stressing that "the brain's brief is to provide complementary facilities that will support the repeated exploitation of operations upon the world [and] to provide computational processes (such as powerful pattern completion) that the world, even as manipulated by us, does not usually afford" (*Being There*, p. 68).

It should be clear enough, from this last quote, that I have certainly not forgotten the lessons that connectionism taught us. The argument for the extended mind thus turns primarily on the way disparate inner and outer components may co-operate so as to yield integrated larger systems capable of supporting various (often quite advanced) forms of adaptive success. The external factors and operations, in this model, are most unlikely to be computationally identical to the ones supported directly in the wetware—indeed, the power of the larger system depends very much on the new kinds of storage, retrieval and transformation made possible by the use of extra-neural resources (see also the tale of John's Brain told in the appendix). These new operations, however, may often be seen as performing kinds of tasks which, were they but done in the head, we would have no hesitation in labelling cognitive. This is because they contribute to behavioural success by for example storing and manipulating information, and by reconfiguring problem spaces. This kind of higher-

© AAHPSSS, 1998.

Copyright © 2000. All rights reserved.

level functional isomorphism is, I think, quite compatible with the idea (stressed by both O'Brien and myself) that there exist deep and important differences between e.g., active biological and passive symbolic modes of storage and retrieval.

Representation (and computation)

Both Sutton and Hooker would like to see a more fully worked-out story about how to factor internal representation and computation into the larger, ecumenical package of *Being There*. So would I. As it stands, the chapter that tackles these topics (Chapter 8: "Being, Computing Representing") is both the largest and the most frustratingly 'unfinished' one in the book. In it, I argue for what I call 'minimal representationalism': the view that we need to combine dynamical and ecological analyses with the search for in-the-head states and processes that both encode contents (albeit, often fragmentary, action-specific kinds of content) and that exploit computational routines so as to systematically transform one content into another. Such states and processes, I argue, are most strongly implicated in episodes in which we reason about absent, counterfactual or imaginary states of affairs.

Sutton queries the point about thoughts concerning the absent. Instead of persisting inner surrogates for what is not present-to-hand, Sutton proposes that we create such surrogates on the spot, out of the whole cloth of a complex web of inner and outer dynamics. But I have no problem with such an account. All it means (if true) is that the inner surrogate comes into being as and when it is needed. This is fine by me: what matters is (still) that on-going behaviour, in such cases, is explained by appeal to identifiable inner content-bearers. The stability and long-term persistence of such items is not an issue on which I have to take a stand.

That said, I should concede the more general substance of Sutton's worry. For it is true that it is not inconceivable that complex, evolving inner states, of some kind which does not succumb to any fine-grained content-ascribing decomposition, might somehow support behaviour which is coordinated with respect to distal, absent or non-existent states of affairs. We cannot rule this out *a priori*, and some researchers in Artificial Life and real-world robotics are already trying to solve such coordination problems without making any prior commitments to the use of internal representation (e.g., Beer 1996).

My own view, however, is that the most practical and efficient mechanisms for coordinating complex behaviour with what is absent, imaginary and counterfactual will involve the use of systems of inner states

© AAHPSSS, 1998.
Copyright © 2000. All rights reserved.

and processes whose functional role is to stand-in for the 'missing' states of affairs—in short, internal models and internal representations. In recent (post-*Being There*) work, I have pursued this idea using some of the apparatus mentioned by Hooker who asks "could off-line emulation be the intended source of Clark's representation?". Very briefly, the idea (pursued at length in Clark and Grush, submitted; and also in Clark, in press) is that internal representation, strongly conceived, gets its foot in the door of biological cognition when on-line, real-time behaviour requires a system to adjust certain parameters on the basis of information that is not available fast enough to allow direct control by environmental feedback. It is speculated, for example (see Ito 1984, Kawato *et al.* 1987, Dean *et al.* 1994) that the control of reaching requires proprioceptive feedback to be deployed before real signals from the sensory peripheries could be exploited. A solution is to train on-board circuitry to mimic the dynamics of the larger system and to generate a prediction of the real signal that can then be used to fine-tune the reaching. The emulator circuit thus acts as a stand-in for the real-world system itself. Although I mention this work in the book (pp. 22–3), it is not there developed into a general story about (strong) internal representation. The development (again, see Clark and Grush, submitted) involves noting that such an emulator, though originally invoked to fine-tune actual reaching, may be run off-line so as support motor imagery without real-world action (see Grush 1995). In such cases we can actively isolate the precise aspects of the processing that correspond to different target events and states of affairs (in the reaching case, to different arm motion parameters). Our suggestion is that a creature uses full-blooded internal representations if and only if it is possible to identify within them specific states or processes whose functional role is to act as de-coupleable surrogates for specifiable (usually extra-neural) states of affairs.[2] Motor emulation circuitry, we think, provides a clear, minimal and evolutionarily plausible case in which these conditions are met. And it is shows how internal representations might first originate in systems whose 'goal' is merely to maintain close and fluent behavioural contact with the world around them.

The Future

Naomi Quinn, in her richly suggestive and multi-layered commentary, offers a fascinating counterpoint to my tendency to depict cultural scaffolding as external and as heavily linguistic. Quinn's emphasis, by contrast, is on the "unspoken, internal cultural representations that mediate performance of . . . cognitive tasks". These involve, as I understand

© AAHPSSS, 1998.

Copyright © 2000. All rights reserved.

it, shared culture-specific ideas and metaphors that, although often unconscious and unarticulated, serve to structure our understanding, judgement and responses. Quinn depicts, in persuasive detail, the content of (to take one example) a shared cultural representation of marriage as a lasting, yet fundamentally contractual and mutually beneficial, relationship. Such shared conceptions make it possible to construct arguments and discourses whose flow depends crucially on unstated, invisible premisses and assumptions. The presence of such a shared backdrop reduces cognitive load and scaffolds problem-solving: yet the scaffolding consists neither in external structures nor in linguistic productions, inscriptions or rehearsals.

I think Quinn is right to depict this as a kind of cognitive scaffolding and as a way in which culture seeps into the mind. Such internal scaffolding helps to enforce a kind of mental hygiene by both restricting and propelling our reasoning and inference. (Sutton's lovely description of the role of linguistic rehearsal has a natural extension to this kind of unarticulated, schematic case: the culturally inherited schemes act as a kind of pivot for linguistic and interpersonal reason.)

My only fear, in all this, is that the notion of scaffolding could one day grow too broad. It would not do, for example, if every aspect of cognition could be seen as performing a scaffolding function. We need to maintain a sense that the scaffolding involves elements that are in some hard-to-pin-down sense external to the most basic processes of biological reason. I think, however, that the case of internal cultural representations probably qualifies, insofar as we are there dealing with inner states whose shape, content and role are fixed by some quite specific social and collective practices which seem to reconfigure on-board reason in ways not predictable from a more individualistic stance. But however we describe them, Quinn is surely right to flag an important dimension of analysis ignored in my original treatment.

There are other directions, also, in which I hope to extend the original project. One is to look more closely at the question of biological implementation; to ask whether neural computation might be pressing important functionality out of 'mere implementation details' such as the low-level physics of the hardware (see e.g., Thompson 1996). Another is to look at the 'double life' of beliefs and ideas, being on the one hand mental entities ascribed to individual agents and, on the other hand, entering into larger, collective dynamics that have properties all their own (think of the way ideas and beliefs interact and snowball in financial markets—see Arthur 1997). Accommodating this 'double-aspect' of beliefs and ideas is, I suspect, going to prove crucial to the understanding of many forms of cultural scaffolding. In addition (and as we saw), the

© AAHPSSS, 1998.
Copyright © 2000. All rights reserved.

respective explanatory roles of dynamics, computation and representation are still somewhat up for grabs. Terms of art such as 'emergence' and 'scaffolding' probably require more work. And the whole issue of the mind's (putative) extension into the world is begging for further work and reflection. So there is plenty to do!

I would like to end, however, on a truly positive note. It has been a striking (and tremendously gratifying) feature of the response to *Being There* that it has found favour amongst a truly wide diversity of disciplines and readers. In particular, I am greatly excited by the response from the social sciences, cultural anthropology, education, business and economics, as well as philosophy and the traditional cognitive sciences. There is, in the current climate, a real opportunity (or so it seems to me) to now draw together a rich, diverse and highly multi-disciplinary base in pursuit of a truly integrated science of the mind: a science that confronts cognition on its home turf, as the activity of social agents locked in the enabling embrace of culture, artefact and world.

Department of Philosophy,
Washington University,
St Louis, Missouri, USA.

1. I just thought of seven.
2. It is a nice question whether there is a coherent weaker sense of internal representation applicable to cases where the 'de-coupleability' criterion is not met. For an attempt to pin down such a weaker sense, see Wheeler and Clark (in progress).

References

Arthur, B. (1997) "Beyond rational expectations" in J. Drobak and J. Nye (eds), *The Frontiers of the New Institutional Economics.* London: Academic Press.

Beer, R. (1996) "Towards the evolution of dynamical neural network for minimally cognitive behavior". *Proceedings of the Society for Adaptive Behavior.*

Clark, A. (1989) *Microcognition.* Cambridge, Mass: MIT Press.

Clark, A. (1993) *Associative Engines.* Cambridge, Mass: MIT Press.

Clark, A. (in press) *The Dynamical Challenge.*

Clark, A., and Chalmers, D. (1995) "The Extended Mind". *PNP Research Report.* Washington University, USA.

Clark, A., and Grush, R. (submitted) "Towards a Cognitive Robotics". *Adaptive Behavior.*

© AAHPSSS, 1998.

Copyright © 2000. All rights reserved.

Dean, P., Mayhew, J., and Langdon, P. (1994) "Learning and Maintaining Saccadic Accuracy: A Model of Brainstem-Cerebellar Interactions". *Journal of Cognitive Neuroscience.*

Grush, R. (1995) "Emulation and Cognition". PhD Dissertation, University of California.

Ito, M. (1984). *The Cerebellum and Neural Control.* New York: Raven Press.

Kawato, M., Furukawa, K., and Suzuki, R. (1987) "A hierarchical neural network model for the control and learning of voluntary movement". *Biological Cybernetics.*

Sloman, A. (1994) "Explorations in design space". Paper presented at the 11th European Conference on AI (ECAI), Amsterdam.

Smithers, T. (ms) "Autonomy in robots and other agents".

Thompson, A. (1996) "Unconstrained evolution and hard consequences". *Cognitive Science Research Report* (CSRP), University of Sussex, UK.

Wheeler, M. and Clark, A. (in progress) "Genic representation: Reconciling content and causal complexity".

© AAHPSSS, 1998.

Copyright © 2000. All rights reserved.