S0004-3702(03)00054-7/REV AID:1984 Vol.●●●(●●●) aij1984 P.1 (1-40)
ELSGMLTM(ARTINT):m1a v 1.155 Prn:29/05/2003; 10:10 by:ML p. 1

ELSEVIER
Available at www.ComputerScienceWeb.com
POWERED BY SCIENCE DIRECT®
Artificial Intelligence ●●● (●●●●) ●●●–●●●
Artificial Intelligence
www.elsevier.com/locate/artint

Field Review

Embodied Cognition: A field guide

Michael L. Anderson

Institute for Advanced Computer Studies, University of Maryland, College Park, MD 20742, USA

Abstract

The nature of cognition is being re-considered. Instead of emphasizing formal operations on abstract symbols, the new approach foregrounds the fact that cognition is, rather, a *situated activity*, and suggests that thinking beings ought therefore be considered first and foremost as acting beings. The essay reviews recent work in Embodied Cognition, provides a concise guide to its principles, attitudes and goals, and identifies the physical grounding project as its central research focus.

For over fifty years in philosophy, and for perhaps fifteen in Artificial Intelligence and related disciplines, there has been a re-thinking of the nature of cognition. Instead of emphasizing formal operations on abstract symbols, this new approach focuses attention on the fact that most real-world thinking occurs in very particular (and often very complex) environments, is employed for very practical ends, and exploits the possibility of interaction with and manipulation of external props. It thereby foregrounds the fact that cognition is a highly *embodied* or *situated activity*—emphasis intentionally on all three—and suggests that thinking beings ought therefore be considered first and foremost as acting beings.

This shift in focus from Descartes' "thinking thing", and the picture of human being and subjectivity it suggests, to a more Heideggerian approach to being in the world, in which agency and interactive coping occupy center stage, is an extremely important development, the implications of which are only just beginning to be fathomed. Very recently a number of books have appeared which detail this shift, and explore in various ways these implications.[1] I have selected three of them to discuss in detail here: *Cambrian Intelligence* by Rodney Brooks [15]; *Philosophy in the Flesh* by George Lakoff and Mark

E-mail address: anderson@cs.umd.edu (M.L. Anderson).

[1] Books directly related to Embodied Cognition which have appeared just since 1995 include: [3,8,10,13–17, 19,20,22,25,29,35,39–42,55,58,62,65,66,72–74,79,88,91,92,98–100,102,105,111,113,116–120,122,125,126].

doi:10.1016/S0004-3702(03)00054-7

S0004-3702(03)00054-7/REV AID:1984 Vol.●●●(●●●) aij1984 P.2 (1-40)
ELSGMLTM(ARTINT):m1a v 1.155 Prn:29/05/2003; 10:10 by:ML p. 2

Johnson [72]; and *Where the Action Is* by Paul Dourish [29]. Together these works span a fair bit of this extremely active and diverse research area, and I will supplement the discussion with references to other works, as appropriate. In, as it were, summarizing the summaries, I hope thereby to provide a concise guide to the principles, attitudes and goals of research in Embodied Cognition.

In what follows, I will first very briefly outline the basic foundations and philosophical underpinnings of the classical approach to Artificial Intelligence, against which Embodied Cognition (henceforth EC) is in large part a reaction (Section 1). Section 2 outlines the alternative approach to understanding intelligence pioneered by Rodney Brooks. In that section I also discuss in some detail the central criticisms of classical AI which motivate the situated or embodied alternative, and identify the physical grounding project as the central project of EC. This project is envisioned as a comprehensive version of the symbol grounding problem [52,53]—the well-known question of how abstract symbols can acquire real-world meaning—and centrally involves understanding how cognitive contents (however these are ultimately characterized, symbolically or otherwise) must ultimately ground out in (terms of) the agent's embodied experience and physical characteristics. In my view, it is the centrality of the physical grounding project that differentiates research in *embodied* cognition from research in *situated* cognition, although it is obvious that these two research programs are complementary and closely related. Section 3 outlines the different meanings of embodiment to be found in the EC literature, and discusses each one in relation to the physical grounding project, while Section 4 briefly emphasizes the importance of social situatedness to human level intelligence. In Section 5 I review and evaluate some common criticisms of the EC approach. Finally, Section 6 discusses some implications of this new understanding of intelligence and its grounds for the design of human-computer interfaces. I conclude by briefly recounting the central principles of EC research.

1. Cartesianism, cognitivism, and GOFAI

That Descartes is the thinker most responsible for the theoretical duality of mind and body is one of the things that everybody knows; and like most such common knowledge it is not quite accurate. Descartes' arguments for the separation of body and soul are part of a long legacy of dualistic thinking, in which Plato's discussion of the immateriality of the soul and the Christian metaphysical tradition which adopted and preserved that discussion played a central role. Indeed, Descartes himself always insisted that, although body and soul were conceptually, and therefore *ontologically* distinct, they nevertheless formed an empirical unity. What we *have* inherited from Descartes is a way of thinking about our relation to the world—in particular our *epistemological* relation to the world—which serves to support and strengthen this ontological stance [91,92]. Descartes' famous ontological realization that he is a thinking thing is conditioned and tempered by the epistemological admission that all he had accepted as most true had come to him through the senses; yet it is precisely this intrusion of the body between knowledge and world which is in the end unacceptable. The body must be part of the causal order to allow for perceptual interaction, but is therefore both unreliable (a cause of the senses' deception) and, as it were, *too* reliable (driven by physical forces, and so a potential source of unfreedom).[2] Thus, the body is for Cartesian philosophy both necessary and unacceptable, and this ambivalence drives mind and body apart in ways Descartes himself may not have intended. Sensation requires the physicality of the body; the freedom and reliability of human reason and judgment seem to require the autonomy of the soul. We may have no choice about how the world appears to us—we may be to this degree determined by our physical constitution—but we can step back from appearances, and allow reflection and judgment, even if in the end this means adopting the skeptical attitude and simply withholding belief. The postulation of an inner arena of disengaged representation, influenced by experience but governed by reasons and not causes, is surely a natural way to try to account for this possibility.[3]

These tensions clearly play a role in another aspect of the Cartesian world-view, not much emphasized in these contexts, but perhaps of greater importance: the discontinuity between humans and animals. For Descartes, animals are mere mechanisms, complex and interesting to be sure, but physical automata nevertheless [23, Vol. III, pp. 365–366, 374]. They *do* have sensation, because all that is needed for that is the proper organs.[4] However, they lack thought, perhaps the most important evidence for which is the fact that they lack language. This denial that sensing and acting in the world require thinking, and the concomitant identification of thinking with the higher-order reasoning and abstraction paradigmatically displayed in language use is perhaps the true heart of the Cartesian attitude. Indeed, I believe that it is primarily from this inheritance that the central attitudes and approach of cognitivism can be derived.

Simply put, cognitivism is the hypothesis that the central functions of mind—of thinking—can be accounted for in terms of the manipulation of symbols according to explicit rules. Cognitivism has, in turn, three elements of note: representation, formalism, and rule-based transformation. First and foremost is the idea that cognition centrally involves *representation*; cognitivism is committed to the existence of "distinct, identifiable, inner states or processes"—that is, the symbols—"whose systemic or functional role is to stand in for specific features or states of affairs" [20, p. 43]. However, just as is the case in modern logic, it is the *form* of the symbol (or the proposition of which the symbol is a part) and not its meaning that is the basis of its rule-based transformation. To some degree, of course, such formal abstraction is a necessary condition for representation—the token for 'green' in my mental lexicon is not itself green, nor does it necessarily share any of the other properties of green. Indeed, the relation between sign and signifier seems in this sense necessarily arbitrary,[5] and this thereby enforces a kind of distance between the inner

---

[2] The thesis of physical determinism as we have come to formulate it awaited later developments in mechanics and mathematics, and attained its first recognizably contemporary form in the work of Pierre-Simon Laplace (1749–1827); still, mechanism was an important part of Descartes' world-view. See, e.g., [28].

[3] See [49] for a very nice account and critique of Cartesian disengagement, from a Heideggerian perspective.

[4] "Please note that I am speaking of thought, and not of life or sensation. I do not deny life to animals, since I regard it as consisting simply in the heat of the heart; and I do not even deny sensation, in so far as it depends upon a bodily organ. Thus my opinion is not so much cruel to animals as indulgent to human beings—at least to those who are not given to the superstitions of Pythagoras—since it absolves them from the suspicion of crime when they eat or kill animals" [23, Vol. III, p. 366].

[5] But see [107] for some suggestions to the contrary.

S0004-3702(03)00054-7/REV AID:1984 Vol.●●●(●●●) aij1984 P.4 (1-40)
ELSGMLTM(ARTINT):m1a v 1.155 Prn:29/05/2003; 10:10 by:ML p. 4

arena of symbol processing and the external world of meaning and action. Still, as we will see in more detail later, this formal abstraction is nevertheless a matter of degree, and this aspect of cognitivism is separate—and separately criticizable—from the general issue of representation. This brings us to third important aspect of cognitivism: the commitment to explicitly specifiable rules of thought. This commitment follows naturally from the others, for having disconnected the form of a symbol from its meaning, cognitivism rules out the possibility of content-sensitive processing, and so requires formal rules to govern the transformation from one cognitive state to another.[6]

This rather too simple account of cognitivism as the commitment to an account of cognition based on the manipulation of abstract representations by explicit formal rules, brings us to GOFAI (Good Old Fashioned Artificial Intelligence), which has been since its inception, and perhaps still is, the dominant research paradigm in Artificial Intelligence. GOFAI might be loosely defined as any attempt to produce machine intelligence by methods which primarily reflect Cartesian and cognitivist attitudes. The looseness of this definition is meant to emphasize the fact that there are probably no research projects in AI which are purely GOFAI, despite the caricatures which abound in the critical literature. Nonetheless, that many research projects, particularly those from AI's foundational period,[7] display the influences of Cartesian and cognitivist thinking is undeniable.

A particular example will help illustrate these influences, and serve as a reference point for later discussion. Perhaps the most impressive current effort along these lines is Cyc (as in enCYClopedia), the now almost 20-year-old project to create a general-purpose common sense reasoner. The approach is straightforwardly cognitivist:

> The Cyc knowledge base (KB) is a formalized representation of a vast quantity of fundamental human knowledge: facts, rules of thumb, and heuristics for reasoning about the objects and events of everyday life. The medium of representation is the formal language CycL... The KB consists of terms—which constitute the vocabulary of CycL—and assertions which relate those terms [24].

It is the conviction of Cyc's designers that the result of learning, and the precondition for more (and more sophisticated) learning, is the acquisition and retention of many, many, many facts about the world and how it works. The most basic knowledge is the sort of commonsense stuff that we take utterly for granted:

> Consider Dr. Dave Bowman, mission commander of Discovery. His experiences as an engineering or astrophysics student and his astronaut training qualified him to lead the mission. Before that, his high school and undergraduate college prepared him for graduate school, and before that, he learned in elementary and middle school the

[6] In point of fact, the question of how to differentiate symbolic from non-symbolic, rule-based from non-rule-based cognitive systems—how to determine, for instance, whether connectionist implementations of simple reasoning employ representations manipulated by rules—is extremely vexed. See [9,60,123] for a more thorough discussion of these issues.

[7] Dreyfus [32] suggests 1957–1977; Brooks extends this range to about 1981, and suggests further that up until 1991 "there was very little change in the underlying assumptions about the models of thought" [15, p. 153].

fundamentals he needed for high school. And long before—here we get to some very important stuff indeed—his early experiences as a baby and toddler prepared him for kindergarten and first grade. He learned, for instance, to talk and that people generally prepare food in kitchens. He learned that if you leave something somewhere, it often remains there, at least for a while. He learned that chairs are for sitting on and that pouring milk from one glass to another differently shaped glass doesn't change the total volume of the milk. He learned that there's no air in outer space so he'd better not forget the helmet of his spacesuit if he's going walking in space. He learned all this—and a million other things [76, pp. 195–196].

For the purposes of this essay, we might call the supposition that we learn these sorts of things: *if something is not supported it falls*, that learning entails representing that knowledge in this sort of way: $ist((\forall x \neg supported(x) \rightarrow falls(x)), NTP)$[8], and that reasoning is the manipulation of such representations according to formal, logic-like rules, the central hypothesis of GOFAI. It is against this hypothesis that research in EC is principally arrayed.

2. Cambrian intelligence

Although Hubert Dreyfus must be given a great deal of credit for first drawing attention to the limitations of GOFAI [30], I think it can be argued that the single figure most responsible for the new AI is Rodney Brooks. A collection of his seminal papers has recently appeared [15] which together comprise not just a sustained critique of the central hypothesis of GOFAI, but the outlines of a philosophically astute and scientifically principled alternative. The book is split into two sections, *Technology* and *Philosophy*, with the former containing such papers as "A Robust Layered Control System for a Mobile Robot", and the latter "Intelligence without Representation" and "Intelligence without Reason", among others. "Intelligence without Reason", by far the longest paper in the collection (pp. 133–186), is also the most complete statement of Brooks' guiding philosophy, and today stands as the overall best introduction to the principles and motivations of the new AI.

As we have seen, traditional AI is characterized by an understanding of intelligence which foregrounds the notions of thought and reason, and adopts certain conventions for approaching these which centrally involve the creation of representations, and the deployment of high-level cognitive skills such as planning and problem solving. For Brooks, however, such an approach "cannot account for large aspects of what goes into intelligence" (p. 134). In contrast to this high-level or top-down approach to intelligence, Brooks advocates studying intelligence from the bottom up, and specifically urges us to recall our evolutionary lineage. As evolved creatures, human beings are largely continuous with our forebears, and we have inherited from them a substrate of capacities and systems

[8] $ist(x, y)$ means x is true in theory y: in this case $y =$ NTP, the Naive Theory of Physics, one the many microtheories employed in Cyc [48].

S0004-3702(03)00054-7/REV AID:1984 Vol.●●●(●●●) aij1984 P.6 (1-40)
ELSGMLTM(ARTINT):m1a v 1.155 Prn:29/05/2003; 10:10 by:ML p. 6

for meeting our needs in, and generally coping with a given environment. From such considerations follows the perhaps reasonable, but decidedly un-Cartesian thought: "The study of that substrate may well provide constraints on how higher level *thought* in humans could be organized" (p. 135, emphasis in original). As we will see, this tendency to emphasize, on evolutionary grounds, the *continuity* between humans and other animals, and the converse willingness to see in animals instances of intelligent behavior, is an extremely important motivation for the study of EC.

But while such thinking may provide justification for pursuing a certain sort of research into cognition, it does not by itself constitute a critique of other approaches. So what, exactly, is wrong with GOFAI? This, of course, is a difficult and vexed question, because arguments over theoretical principles, such as whether humans really represent things in terms of proposition-like mental entities, have so far proved indecisive, there being strong theoretical support on both sides of the issue; so far, neither sufficiently compelling empirical evidence, nor widespread agreement on decisive foundational assumptions has been achieved.[9] Similarly, claims about the deficiencies of particular existing systems can always be deflected, in most cases legitimately, by pointing out that any engineering project is governed by real resource limitations—temporal limitations most significantly, including the need for the graduate students who primarily build such systems to graduate—and that the next iteration will address the noted shortcomings. Still, even without the sort of definitive experiment or event which occasionally marks a scientific revolution [70], it should be possible to look at the *trajectory* and *tendencies* of any given research program, and come to some sort of tentative assessment of its promise. Thus does Brooks approach work in AI and robotics. Surveying some early work, e.g., Skakey [87], CART [84] and Hilare [47], he comments:

> All of these systems used offboard computers ... and operated in mostly static environments. All of these robots operated in environments that at least to some degree had been specially engineered for them. They all sensed the world and tried to build two or three dimensional world models of it. Then, in each case, a planner could ignore the actual world, and operate in the model to produce a plan of action for the robot to achieve whatever goal it had been given.... We will call this framework the sense-model-plan-act framework, or SMPA for short (pp. 136–137).

One of the problems that Brooks identifies with this approach is that it is insufficiently dynamic. Once the system has built its model, it will not notice changes in its actual environment. Then, when it comes time to act (and given the slowness of the systems, this time may be long in coming) the world may well have changed in ways which make the plan obsolete. Depending on the sensitivity of the system, this will either result in the re-initiation of the SMPA cycle, or the execution of an ineffective plan. Neither result is optimal, to say the least. Of course, one way to address this problem is to improve

[9] Indeed, as far as agreement or progress on foundational theoretical issues, there has been little if any progress in the last 10 years. See [68], which presents a set of controversies which must be counted as yet controversial. It is perhaps worth noting, however, that these issues *are* being discussed; it was Kirsh's impression in 1991 that they had been largely ignored.

performance, especially in the extremely expensive process of perception and model construction. But for Brooks and his followers, this is to miss the point that the SMPA framework is by its nature too expensive, and therefore biologically implausible. As an alternative they encourage the adoption of highly reactive models of perception, bypassing the representational level altogether. As he has famously formulated it:

> We have reached an unexpected conclusion (C) and have a rather radical hypothesis (H).
>
> (C) When we examine very simple level intelligence we find that explicit representations and models of the world simply get in the way. In turns out to be better to use the world as its own model.
>
> (H) Representation is the wrong unit of abstraction in building the bulkiest parts of intelligent systems (pp. 80–81).

It is true that despite vast improvements in the speed of microprocessors, and significant advances in such areas as computer vision, knowledge representation, non-monotonic reasoning, and planning there has yet to be an SMPA system that can operate in a complex, real-world environment on biologically realistic time-scales. The twin scale-up of environmental richness and real-time dynamics has so far proved insurmountable. On the other side of the coin, all these areas *are* advancing, and so perhaps the achievement of real-time SMPA intelligence is just a matter of time. Showing that something has not yet happened is a long way from showing it won't.

Yet there are reasons to suspect the latter. Consider first the central problem of planning, that is, of figuring out what to do next, given, say, a certain goal, and a current situation. On the SMPA model, this process involves examining the world model, and mapping on to that model a series of steps which will achieve the desired state. This abstract description hides two kinds of complication, which I will call *dynamics* and *relevance*. The problem of dynamics we have mentioned already; if the world changes, then the plan will have to be likewise adjusted to fit the new situation. Within the SMPA framework, there are two approaches to this problem. The first is to include the dynamics of the world in the model, and to plan in terms of the expected changes in the environment. Naturally, this only pushes the problem back one step, for now we have to monitor whether or not the *changes* are the expected ones, and re-plan when they are not. This suggests a second approach, to include in the initial plans contingent sub-plans to be executed in the case of the various possibilities for the future states of the world. I suspect it is obvious both that this would result in an unmanageable computational explosion, and that real-world planning is nothing like this.

As difficult as the problem of dynamics is, there is yet another problem implicit in the above discussion. For one doesn't want to re-plan in the face of *every* change, only those which are *relevant*, that is, which are likely to affect the achievability of the goal. Thus, for a heavy robot moving across a room the location and dynamics of big, solid objects is likely relevant, but the speed and direction of the draft from the open window is not. Unless the task is to carry a stack of papers. Likewise, the broken and deeply pitted floor tiles make no difference to a robot with big spongy wheels, but might matter to one with different means for locomotion. In general, what counts as a relevant fact worth noticing—

S0004-3702(03)00054-7/REV AID:1984 Vol.●●●(●●●) aij1984 P.8 (1-40)
ELSGMLTM(ARTINT):m1a v 1.155 Prn:29/05/2003; 10:10 by:ML p. 8

say, whether something falls in the general class of obstacle or not—will depend both on the capacities of the agent and the task to be performed. This is obvious enough, to be sure, but turns out to be notoriously difficult to implement in a representational system of any size.

For let us suppose our agent to have modeled a sufficiently complex environment; we would expect this to require many thousands of 'facts'—let's say a large number equal to m. According to the above 'obvious' requirement for agency, our agent would need to know, for each such fact, whether or not it was relevant to the proposed action. And to know *this* requires that the agent know a good deal about each of the actions in its repertoire (more facts), and further know how to tell whether or not a given (environmental) fact was relevant to a given behavioral fact. Given a number of such behavioral facts n, and assuming a straightforward set of relevance rules p could be written, determining the relevance of the various aspects of the environment to any one action will minimally require $m * n * p$ comparisons. Further, assuming that the robot is reasoning about what it knows (not to mention that the environment is itself changing, and perhaps the robot is noticing this), the number m is constantly increasing, as new facts are added to the KB through reasoning and observation. It is this sort of performance problem that is behind Dennett's spoof of the third-generation, newly re-designed robot trying to figure out how to remove its battery from the room it shares with a time-bomb:

> Back to the drawing board. 'We must teach it the difference between relevant implications and irrelevant implications', said the designers, 'and teach it to ignore the irrelevant ones'. So they developed a method of tagging implications as either relevant or irrelevant to the project at hand, and installed the method in their next model, the robot-relevant-deducer, or R2D1 for short. When they subjected R2D1 to the test that had so unequivocally selected its ancestors for extinction, they were surprised to see it sitting, Hamlet-like, outside the room containing the ticking bomb, the native hue of its resolution sicklied o'er with the pale cast of thought, as Shakespeare (and more recently Fodor) has aptly put it. 'Do something!' they yelled at it. 'I am', it retorted. 'I'm busily ignoring some thousands of implications I have determined to be irrelevant. Just as soon as I find an irrelevant implication, I put it on the list of those I must ignore, and...' the bomb went off [27, p. 129].

It may be that the relevance problem is merely practical, but if so, it is a very deep practical problem that borders on the theoretical. At the very least it is a fundamental issue for any representational system, for at the root of the relevance problem is that most basic question for any representation: what should be modeled, and what ignored or abstracted away? One moral which might be derived from the problem of relevance—a weak form of the more radical moral that Brooks actually draws—is that it doesn't make sense to think about representing at all unless one knows what one is representing *for*. The SMPA model envisions the infeasible task of *deriving* relevance from an absolute world model; but if we step back a pace we would realize that no system can ever have such an absolute (or, context-free) world model in the first place. Every representer is necessarily *selective*, and a good representation is thereby oriented toward a particular (sort of) use by a particular (sort of) agent. Just as the problem of dynamics seems to suggest the inevitability of short, incomplete plans, so the problem of relevance pushes us towards adopting more limited representations, closely tied to the particularities, and therefore appropriate to the needs, of a given agent.

For Brooks, the failure of the SMPA model indicates the nature of goal-orientation (which is, after all, the cognitive phenomenon that planning is intended to capture) should be re-thought. The above problems seem to suggest their own solution: shorter plans, more frequent attention to the environment, and selective representation. But the logical end of shortening plan length is the plan-less, immediate action; likewise the limit of more frequent attention to the environment is *constant* attention, which is just to use the world as its own model. Finally, extending the notion of selective representation leads to closing the gap between perception and action, perhaps even casting perception largely in terms of action.[10] Thus, the problems of dynamics and relevance push us toward adopting a more reactive, agent-relative model of real-world action, what I call situated goal-orientation.[11] Brooks writes,

> Essentially the idea is to set up appropriate, well conditioned, tight feedback loops between sensing and action, with the external world as the medium for the loop. . . .
>
> We need to move away from state as the primary abstraction for thinking about the world. Rather, we should think about processes which implement *the right thing*. We arrange for certain processes to be pre-disposed to be active and then given the right physical circumstances the goal will be achieved. The behaviors are gated on sensory inputs and so are only active under circumstances where they might be appropriate. Of course, one needs to have a number of pre-disposed behaviors to provide robustness when the primary behavior fails due to being gated out.
>
> As we keep finding out what sort of processes implement *the right thing*, we continually redefine what the planner is expected to do. Eventually, we won't need one (p. 109).

One is certainly entitled to doubt whether this sort of selective-reactive attunement to the environment can account for the complex tasks routinely faced by the typical soccer mom, as she shuttles her various children to their various activities at the right time, in the right order, meanwhile figuring out ways to work in laundry, dry cleaning and grocery shopping [4]. Likewise, dialog partners seem to engage in complex

[10] This idea, which I think is ultimately very powerful, suggests a notion familiar from phenomenology that the perceptual field is always already an action-field—that the perceived world is always known in terms directly related to an agent's current possibilities for future action. One way of cashing this out is in terms of affordances, the perceived availability of things to certain interventions [46], so that the world, as it were, constantly *invites* action; another is Brooks' notion of a specialized repertoire of behaviors selectively gated by environmental stimuli, which is, in turn, not unrelated to Minsky's notion of the society of mind [83]. More generally, these notions are related to the theory of direct perception, which is usefully and thoroughly discussed in [1] and [19] Chapters 11–12.

[11] It should be noted that the relevant neuroscientific studies suggest evidence for *both* agent- or body-relative representations of action [12,63] *and* objective, or world-relative representations [44,45]. As with many such tensions, the path of understanding is not likely to be either/or, but rather both/and, with the hardest work to be done in understanding the relation between the two kinds of representations.

ARTICLE IN PRESS
S0004-3702(03)00054-7/REV AID:1984 Vol.●●●(●●●) ELSGMLTM(ARTINT):m1a v 1.155 Prn:29/05/2003; 10:10 aij1984 P.10 (1-40) by:ML p. 10

behaviors—remembering what has been said; deciding what to say next in light of overall conversational goals; correcting misapprehensions, and reinterpreting past interactions in light of the corrections, thereby (sometimes) significantly altering the current dialog state—which defy analysis without representations.[12] Indeed, it is a vice too often indulged by scientists working in EC to make the absence of representations a touchstone of virtue in design, and to therefore suppose that, just as do the creatures they devise in the lab, so too must humans display an intelligence without representations. Representation-phobia is a distracting and ultimately inessential rhetorical flourish plastered over a deep and powerful argument. For rather than targeting representations *per se*, the central argument of EC instead strikes at their nature and foundation, thereby forcefully, and to my mind successfully raising the question of whether GOFAI has chosen the right elements with which—the right foundation on which—to build intelligence. For suppose that robotic soccer mom *does* need representations, *does* need an abstract, symbolic planner. Still, if EC is on anything like the right track she cannot live by symbols alone; her representations must be highly selective, related to her eventual purposes, and physically grounded. This strongly suggests that her *faculty* of representation should be linked to, and constrained by, the 'lower' faculties which govern such things as moving and acting in a dynamic environment,[13] without questioning the assertion that complex agency requires both reactive *and* deliberative faculties.[14] The central moral coming from EC is not that traditional AI ought to be given up, but rather that in order to incorporate into real-world agents the sort of reasoning which works so well in expert systems, ways must be found to systematically relate the symbols and rules of abstract reasoning to the more evolutionarily primitive mechanisms which control perception and action. As suggested, this may well involve rethinking the nature and bases of representation, as well as a willingness constrain the inner resources available to automated reasoners (at least those intended for autonomous agents) to items which can be connected meaningfully to that agent's perceptions and actions (either directly or through connected resources themselves appropriately related to perception and action). This returns us to Brooks' evolutionary musings, and a passage apparently so instructive that it appears twice:

> It is instructive to reflect on the way in which earth-based biological evolution spent its time. Single-cell entities arose out of the primordial soup roughly 3.5 billion years ago. A billion years passed before photosynthetic plants appeared. After almost another billion and a half years, around 550 million years ago, the first fish and vertebrates

[12] [123] discusses the general importance of planning to cognitive agents, as part of a critique of research in situated action.

[13] There has been a good deal of interesting work on the relation of higher faculties to lower, from an evolutionary perspective. A good place to start is [108]; see also [88,125].

[14] Research on cognitive agents [43,50,61,77,115] explicitly begins with this both/and attitude. My own work in philosophy and AI is aimed primarily at understanding how deliberative, representing agents can cope with a complex, changing environment. Thus I have been working with Don Perlis on time-situated non-monotonic reasoning, and especially on the problem of uncertainty [4,11,18], with Tim Oates on automated reasoning with grounded symbols [90], and with Gregg Rosenberg on a specification for embodied, action-oriented concepts and representations [106].

ARTICLE IN PRESS
S0004-3702(03)00054-7/REV AID:1984 Vol.●●●(●●●) aij1984 P.11 (1-40)
ELSGMLTM(ARTINT):m1a v 1.155 Prn:29/05/2003; 10:10 by:ML p. 11

> appeared, and then insects 450 million years ago. Then things started moving fast. Reptiles arrived 370 million years ago, followed by dinosaurs at 330 and mammals at 250 million years ago. The first primates appeared 120 million years ago and the immediate predecessors to the great apes a mere 18 million years ago. Man arrived in roughly his present form 2.5 million years ago. He invented agriculture a mere 10,000 years ago, writing less than 5000 years ago and 'expert' knowledge only over the last few hundred years.
>
> This suggests that problem solving behavior, language, expert knowledge and application, and reason, are all pretty simple once the essence of being and reacting are available. That essence is the ability to move around in a dynamic environment, sensing the surroundings to a degree sufficient to achieve the necessary maintenance of life and reproduction. This part of intelligence is where evolution has concentrated its time–it is much harder (p. 81 also pp. 115–116).
>
> This is the physically grounded part of animal systems ... these groundings provide the constraints on symbols necessary from them to be truly useful (p. 116).

This is perhaps the most important *theoretical* upshot of Brooks' work: the knowledge-is-everything approach of GOFAI, exemplified especially by Douglas Lenat and the others involved in Cyc, can never work, for such a system lacks the resources to ground its representations [52,53]. More importantly, the structure and type of representation employed in Cyc, being unconstrained by the requirements any grounded, agency-oriented components, may well turn out to be unsuited for direct grounding. Despite the "implicit assumption that someday the inputs and outputs will be connected to something which will make use of them" (p. 155) the system in fact relies entirely on human interpretation to give meaning to its symbols, and therefore implicitly requires the still mysterious—and no doubt highly complex—human grounding capacity to serve as intermediary between its outputs and real-world activity. This is one reason why EC researchers tend to prefer a bottom-up approach to grounding, and often insist on working *only* with symbols and modes of reasoning which can be straightforwardly related to perception and action in a particular system. A system like Cyc which relies on human faculties to give meaning to its symbols may well require a fully implemented human symbol grounder, including human-level perceptual and behavioral capacities, to be usefully connected to the real world.

Grounding the symbol for 'chair', for instance, involves both the reliable detection of chairs, and also the appropriate reactions to them.[15] These are not unrelated; 'chair' is not a concept definable in terms of a set of objective features, but denotes a certain kind of thing for sitting. Thus is it possible for someone to ask, presenting a tree stump in a favorite part of the woods, "Do you like my reading chair?" and be understood. An agent who has grounded the concept 'chair' can *see* that the stump is a thing for sitting, and is therefore (despite the dearth of objective similarities to the barcalounger in the living room,

[15] Consider, in this regard, the difference between the perceptual judgments "There's a fire here" and "There's a fire there", or "There's a dollar here". Surely appropriate grounding has not been achieved by a system which does not react differently in each case.

and despite also being a tree stump) a chair.[16] Simply having stored the fact that a chair is for sitting is surely not sufficient ground for this latter capacity. The agent must know what sitting *is* and be able to systematically relate that knowledge to the perceived scene, and thereby see what things (even if non-standardly) afford sitting. In the normal course of things, such knowledge is gained by mastering the skill of sitting (not to mention the related skills of walking, standing up, and moving between sitting and standing), including refining one's perceptual judgments as to what objects invite or allow these behaviors; grounding 'chair', that is to say, involves a very specific set of physical skills and experiences.

A further problem is the holistic nature of human language and reasoning: 'chair' is closely related to other concepts like 'table'. One is entitled to wonder whether knowing what sort of chairs belong at a table is part of the mastery of 'table' or 'chair'. It is unlikely that clear boundaries can be drawn here; knowing one partly involves knowing the other. Likewise, 'chair' is related to 'throne', so that it is not clear whether we should say of someone who walked up and sat in the King's throne that she failed to understand what a throne was, or failed to understand what a chair was (and wasn't). Given that these concepts are semantically related, that there is a rational path from sentences with 'chair' to sentences with 'table' or 'throne', any agent who hopes to *think* with 'chair' had better have grounded 'table' and 'throne', too.

The point is not that Cyc's symbols can never be grounded or given real-world meaning. Rather, the basic problem is that the semantic structures and rules of reasoning that Cyc uses are rooted in an understanding of what these concepts mean *for us*, and therefore they can be usefully (and fully) grounded only for a system which is likewise very much like us. This includes not just basic physical and perceptual skills, but also (as the throne example demonstrates) being attuned to the social significance of objects and situations, for seeing the behaviors that an object affords may depend also on knowing who one is, and where one fits not just in the physical, but also the social world. An android with this sort of mastery is surely not in our near future, but a lesser robot using Cyc for its knowledge base is just wasting clock cycles, and risks coming to conclusions that it cannot itself interpret or put to use—or worse, which it will misinterpret to its detriment. For EC, the only useful reasoning is grounded reasoning.[17]

The notion that grounding is at the root of intelligence, and is the place to look for the elusive solution to the relevance problem—for grounding provides the all-important

[16] Lying behind this claim is the theory of direct, or ecological, perception [46], which for reasons of space has not been discussed in detail here. In short, the idea is (1) at least some perception does not involve inference from perception to conception, but is rather direct and immediate, and (2) perception is generally action-oriented, and the deliverances of the perceptual system are often cast in action-relative, and even, as in the case of affordances, action-*inviting* terms. In illustration of claim (1), Clancey [19] suggests that one's identification, while blindfolded, of the object in one's hand as an apple is unmediated and direct; it is part of one's immediate perceptual experience that the object is an apple. He contrasts this with the case where one counts the bumps on the bottom of the apple to infer that it is a Red Delicious. The latter involves inference from perceptually gathered information; the former does not (pp. 272–273). Claim (2) we have already discussed, in, e.g., note 10, and immediately above. The theory of direct perception is usefully and thoroughly discussed in [1] and [19] Chapters 11–12.

[17] Note that this doesn't mean that Cyc is not a useful tool. The point is that Cyc is *only* useful for an agent that can ground its symbols. Humans are currently the only such agents known.

constraints on representation and inference with which the purely symbolic approach has such trouble—Brooks calls "the physical grounding hypothesis" (p. 112).

> [Cyc employs] a totally unsituated, and totally disembodied approach ... [110] provides a commentary on this approach, and points out how the early years of the project have been devoted to finding a more primitive level of knowledge than was previously envisioned for grounding higher levels of knowledge. It is my opinion, and also Smith's, that there is a fundamental problem still, and one can expect continued regress until the system has some form of embodiment (p. 155).

As interesting and revolutionary as Brooks' contribution to AI and robotics has been, the ultimate success of the research program depends on whether and how the physical grounding hypothesis can be cashed out. This will mean, among other things, coming to a concrete understanding of phrases like "some form of embodiment", and specifying in detail how embodiment in fact acts to constrain representational schema and influence higher-order cognition. As I have already mentioned, I take this concern with physical grounding to be a central defining characteristic of research in EC, differentiating it from research in *situated* cognition. Thus it is to recent contributions to this project—which I will call the physical grounding project—that we now turn.

3. Embodiment and grounding

It is, of course, obvious that embodiment is of central interest to EC. Yet it is also clear that the field has yet to settle on a shared account of embodiment: indeed, there is little explicit discussion of its meaning, as if it were a simple term in little need of analysis. When considered in the abstract, primarily as an indication of the area of focus in contrast to that indicated by an interest in 'symbols', this is no doubt good enough. But having thus focused attention on embodiment, it is incumbent on the field to say something substantial about its meaning. In this project, the resources of phenomenology should not be overlooked, especially [56,81,82], despite their difficulty.[18] Naturally, a single section of a field review is not the place to try to give the required account. Instead of trying to provide a unified theory, I will lay out the general terrain, and say something about the various meanings of embodiment and their relation to the physical grounding project.

Perhaps the most basic, but also the most theoretical, abstract way to explain what it means to focus attention on the embodiment of the thinking subject comes from Merleau-Ponty and Heidegger.

> Heidegger, for instance, shows—especially in his celebrated analysis of being-in-the-world—that the condition of our forming disengaged representations of reality is that

[18] [31,59,121] are all, in part, attempts to capture the insights of phenomenology in a way directly useful to cognitive science. [117] is a more strictly philosophical attempt to give a thorough account of embodiment and its relation to mind. [51] relates representations to intervention in the context of scientific practice. More work along these various lines is needed.

> we be already engaged in coping with our world, dealing with the things in it, at grips with them. . . . It becomes evident that even in our theoretical stance we are agents. Even to find out about the world and formulate disinterested pictures, we have to come to grips with it, experiment, set ourselves to observe, control conditions. But in all this, which forms the indispensable basis of theory, we are engaged as agents coping with things. It is clear that we couldn't form disinterested representations any other way. What you get underlying our representations of the world—the kinds of things we formulate, for instance, in declarative sentences—is not further representations but rather a certain grasp of the world that we have as agents in it [114, pp. 432–433].

Likewise, Merleau-Ponty argues that perception and representation always occur in the context of, and are therefore structured by, the embodied agent in the course of its ongoing purposeful engagement with the world. Representations are therefore 'sublimations' of bodily experience, possessed of content already, and not given content or form by an autonomous mind; and the employment of such representations "is controlled by the acting body itself, by an 'I can' and not an 'I think that' " ([59, pp. 108–109], see also [33]). A full explanation of the significance of this claim would require an excursion into the bases of the categories of representation and the transcendental unity of apperception in Descartes and Kant, for which there is no room here (but see [117]). However, the most immediate point is fairly straightforward: the content and relations of concepts—that is, the structure of our conceptual schema—is primarily determined by practical criteria, rather than abstract or logical ones. Likewise, experience, which after all consists of ongoing inputs from many different sources, is unified into a single object of consciousness by, and in terms of, our practical orientation to the world: "the subject which controls the integration or synthesis of the contents of experience is not a detached spectator consciousness, an 'I think that', but rather the body-subject in its ongoing active engagement with [the world]" [59, p. 111].

Considering the problem at this level of abstraction is an important ongoing philosophical project, for re-interpreting the human being in terms which put agency rather than contemplation at its center will have eventual implications for many areas of human endeavor.[19] Continued engagement with phenomenology and related areas is essential to this re-understanding. However, for the more narrow project of coming to grips with the details of physical grounding, it is perhaps better to turn our attention elsewhere.

One extremely important contribution to the physical grounding project has been the ongoing work of George Lakoff and Mark Johnson. Since at least 1980, with the publication of *Metaphors We Live By* [71], they have been arguing that the various domains of our mental life are related to one another by cross-domain mappings, in which a target domain inherits the inferential structure of the source domain. For instance, the concept of an argument maps on to the concept of war (Argument is War) and therefore reasoning about arguments naturally follows the same paths.

> It is important to see that we don't just *talk* about arguments in terms of war. We can actually win or lose arguments. We see the person we are arguing with as an opponent.

[19] Not the least important of which is Ethics. See, e.g., [122].

We attack his positions and defend our own. We gain and lose ground. We plan and use strategies. If we find a position indefensible, we can abandon it and take a new line of attack. . . . It is in this sense that the Argument is War metaphor is one that we live by in this culture; it structures the actions we perform in arguing [71, p. 4].

Lakoff and Johnson's most recent contribution along these lines is *Philosophy in the Flesh* [72], which is the strongest, most complete statement to date of the claim that all cognition eventually grounds out in embodiment. *Philosophy in the Flesh* is in many ways a flawed work, for reasons stemming primarily from the authors' apparent disdain for the large amount of related work in philosophy, artificial intelligence, and cognitive science. I have briefly detailed some of these limitations in [93]. However, because of the importance and promise of its central argument and approach, I will focus here on the positive contribution of the book.

As an illustration of how a given example of higher-order cognition can be traced back to its bodily bases, consider the metaphorical mapping "Purposes are Destinations", and the sort of reasoning about purposes which this mapping is said to encourage. We imagine a goal as being at some place ahead of us, and employ strategies for attaining it analogous to those we might use on a journey to a place. We plan a route, imagine obstacles, and set landmarks to track our progress. In this way, our thinking about purposes (and about time, and states, and change, and many other things besides) is rooted in our thinking about space. It should come as no surprise to anyone that our concepts of space—up, down, forward, back, on, in—are deeply tied to our bodily orientation to, and our physical movement in, the world. According to Lakoff and Johnson, every domain which maps onto these basic spatial concepts (think of an upright person, the head of an organization, facing the future, being on top of things) thereby inherits a kind of reasoning—a sense of how concepts connect and flow—which has its origin in, and retains the structure of, our bodily coping with space.

Like most work in EC, *Philosophy in the Flesh* has little explicitly to say about what the body *is*. Nevertheless, it is possible to identify four aspects of embodiment that each play a role in helping shape, limit and ground advanced cognition: physiology; evolutionary history; practical activity; and socio-cultural situatedness. We will discuss each in turn.

3.1. Physiology

According to Lakoff and Johnson, the mind is inherently embodied not just because all its processes must be neurally instantiated, but also because the particulars of our perceptual and motor systems play a foundational role in concept definition and in rational inference. Color concepts, for instance, are characterized by a "center-periphery" structure, with certain colors being "focal" and others conceptualized in terms of the focal hue. In the category "red" there is a central red, as well as peripheral hues tending toward the purple, pink, and orange. "The center-periphery structure . . . is a result of the neural response curves for color in our brains. Focal hues correspond to frequencies of maximal neural response", with the peripheral structure being determined by the overall shape of the neural response curve. "An adequate theory of the conceptual structure of red, including an account of why it has the structure it has . . . cannot be constructed solely from the spectral properties of surfaces. It must make reference to color cones and neural circuitry" [72, p. 24].

A somewhat different take on the contribution of physiology to the content of abstract representations can be found in [94]. This account emphasizes the fact, familiar from Merleau-Ponty, that our sense organs are themselves dynamic instruments of exploration. Thus, the immediate output of the visual system is not a static picture, but a series of sensory changes related to both the movements of the eye and head, and the motions of the object (or quality) perceived.

> Under the present view of what seeing is, the visual experience of a red color patch depends on the *structure of the changes* in sensory input that occur when you move your eyes around relative to the patch, or when you move the patch around relative to yourself. For example, suppose you are looking directly at the red patch. Because of absorption by the macular pigment, the stimulation received by the color sensitive retinal cones will have less energy in the short wavelengths when you look directly at the red patch, and more when you look away from the patch. Furthermore, since there is a difference in the distribution and the density of the different color-sensitive cones in central vs. peripheral vision, with cone density dropping off considerably in the periphery, there will be a characteristic change in the relative stimulation coming from rods and cones that arises when your eyes move off the red patch. What determines the perceived color of the patch is the *set of such changes* that occur as you move your eyes over it [94, Section 5.7].

Thus does the physiology and design of the visual system have a rather direct effect on the contents and overall structure of the representations (or, more generally speaking, the useful abstractions) which can emerge from it. One part of the physical grounding project, then, is spelling out such direct physiological constraints on higher-level abstractions.

3.2. Evolutionary history

As we have already seen, another important dimension of embodiment, and an important part of the physical grounding project, is the evolutionary history of the agent. Although, of course, the evolutionary history of an agent is physiologically *stored*, it expresses its effects in a somewhat less direct manner.

First the sentiment, as expressed by Lakoff and Johnson:

> Reason is evolutionary, in that abstract reason builds on and makes use of forms of perceptual and motor inference present in "lower" animals. The result is a Darwinism of reason, a rational Darwinism: Reason, even in its most abstract form, makes use of, rather than transcends, our animal nature. The discovery that reason is evolutionary utterly changes our relation to other animals and changes our conception of human beings as uniquely rational. Reason is thus not an essence that separates us from other animals; rather, it places us on a continuum with them (p. 4).

Apparently what Lakoff and Johnson have in mind here is the sort of thing we mentioned already above, that the rules of reason can often (if not always) be traced back to more primitive modes of inference, via the connections provided by metaphorical mapping across the relevant domains.

> This is not just the innocuous and obvious claim that we need a body to reason; rather, it is the striking claim that the very structure of reason itself comes from the details of our embodiment. The same neural and cognitive mechanisms that allow us to perceive and move around also create our conceptual systems and modes of reason. Thus, to understand reason we must understand the details of our visual system, our motor system, and the general mechanisms of neural binding (p. 4).

Here the *method* of constraint is straightforwardly physiological—indeed, Lakoff and Johnson often insist that direct neural connections underlie the mappings in question—but the *mode* of constraint is one step more abstract than this. For what is inherited from the source domains, however they are instantiated, is not the details of this instantiation, but the set of inference rules instantiated therein, thus giving rise to an abstract structural or dynamic similarity between the two domains. Most of Lakoff and Johnson's work has been aimed at this aspect of the physical grounding project: spelling out the myriad structural and dynamic similarities between cognitive domains, and tracing many of them back to very basic spatial, perceptual and motor capacities.

There is, however, also an alternate way in which the evolutionary history of an agent might be included in an explanation of the bases of complex behavior, and that is via emergence. We can illustrate emergence with a simple example from [112].

Steels describes a simple robot, which is programmed with two distinct behaviors: (1) it takes a zig-zag path toward any light source, and (2) whenever it encounters an obstacle, it turns before moving again. In the environment contrived for this robot, it must feed itself by going between two poles at the recharging station whenever the light at the station turns on. This recharging behavior is nowhere programmed into the system, but it nevertheless emerges when the robot is placed in the described environment.

Emergence, then, constitutes a third way to relate complex behavior to its physical, or evolutionary grounds. In this case any explanation of the recharging behavior must include reference to the simple behaviors with which the robot was endowed, but the recharging behavior itself is not directly attributable to these, but only to the dynamic interaction between the basic capacities of the robot and its given environment.

3.3. Practical activity

The issue of dynamic agent-world interaction brings us to yet another aspect of embodiment: the practical activity of an agent, and its relation to thinking, problem solving, and symbol grounding.

> [W]e should acknowledge an important role for what Kirsh and Maglio [67] nicely term 'epistemic actions': namely, actions whose purpose is not to alter the world so as to advance physically toward some goal (e.g., laying a brick for a walk), but rather to alter

S0004-3702(03)00054-7/REV AID:1984 Vol.●●●(●●●) aij1984 P.18 (1-40)
ELSGMLTM(ARTINT):m1a v 1.155 Prn:29/05/2003; 10:10 by:ML p. 18

> the world so as to help make available information required as part of a problem solving routine. Examples of epistemic actions include looking at a chessboard from different angles, organizing the spatial layout of a hand of cards ... laying out our mechanical parts in the order required for correct assembly, and so on [21, p. 511].

Consider an instance of (fairly) simple tool use: using a paper clip to retrieve a life-saver from between the car seats. What we *don't* do is exactly pre-compute the distance of the goal, the required shape the paper clip must take, and the precise trajectory of approach necessary to hook the life-saver. Instead we unbend the clip according to a rough estimation of the requirements of the task, try it out, bend it a bit differently and try again. We may also change the angle at which we are viewing the candy, or the direction from which we approach it. We re-position ourselves in these ways to reduce the complexity of the visualization problem; likewise we employ repeated trials to allow us to better 'see' what shape the paper clip must be to be most useful in this context. The point is not that we do no calculation or pre-computation, but that our problem solving routines involve intensive cooperation between computations and internal representation on the one side, and repeated environmental interactions on the other.

Here the cognitive strategies employed are constrained and shaped by the performance characteristics of our body as a whole in the given circumstances, and also by our brain's limited computational resources. The shape, size and flexibility of our fingers first rules out the possibility of a tool-free approach (and we may decide this based on a rough visual calculation, or discover this by first trying to reach the candy with the bare hand). Then the size and configuration of the surrounding space becomes a further constraint, for it, in combination with our overall bodily dimensions and characteristics, determines the range of possible bodily attitudes possible in the space while still allowing at least one hand to be brought usefully to bear in retrieving the life-saver (anyone who has installed a car radio will have vivid recollections of negotiating this issue). This, in turn, limits the range of possible views of the object, thus determining the difficulty of the visualization required, and the kinds of cognitive and motor strategies that will be needed to solve the problem. Finally, having determined some provisionally acceptable definition of the problem space, we (re-)shape the clip and go about our iterative work. Here cognition is bound up with the shape, size, and motor possibilities of the body as a whole, and further relies on the possibility of repeated interactions with the environment. Thus we uncover yet another level of connection between cognition and physical embodiment.

At this level, too, one can begin to give an account of how an agent grounds its *own* symbols, that is, tell not just a story about how various cognitive *capacities* are bound up with and shaped by the details of an agent's embodiment, but the complementary story about how practical activity plays a role in giving meaning to the particular experiences of, or perhaps the representations generated by, a given individual agent.

Consider a classic experiment by Held and Hein [57]. In this experiment, two sets of kittens were raised in the dark, and only exposed to the light under very controlled conditions. When they were in the light, one set of kittens was allowed to roam freely, although each kitten from this first group was fitted with a harness, itself attached to a basket in which a given member of the second group of kittens was placed in such a way that only its head was free to move about. Because the kittens were raised mostly in the dark, both groups developed the same motor capacities and physical repertoire. Likewise, kittens from both groups were exposed to the same amount and sort of visual stimuli. However, only kittens from the first group were in a position to move about and see at the same time. The results were quite striking, and indicated that the kittens from the second, constrained group had not developed an appreciation of the physical significance of their visual experience. For instance, when a normal kitten is held and brought close to the floor or a table-top, it will reach out its paws in anticipation of the contact. The constrained kittens did not display such reaching behavior. Likewise, kittens from the constrained group were more likely to bump into walls and fall off the edges of things, apparently not recognizing the significance of the relevant visual input.

A similar account can be given of the conditions under which it is possible to ground abstract symbols, as in the case of linguistic reference. [34,92] provide the beginnings of such an account, and argue in part that practical activity, understood as a mode of epistemic access to the world, is a necessary underpinning of our general referential capacity.

At the highest level, what is at issue here is the fact that practical, bodily activity can have cognitive and epistemic *meaning*; it can be a part of a particular problem solving routine, as with the use of paper in long division, or be involved in ongoing cognitive development, as with the use of environmental interactions and manipulations to learn to recognize objects, or to come to appreciate of their significance and meaning.

> [T]he flow of thoughts and the adaptive success of reason are now seen to depend on repeated and crucial interactions with external resources. The role of such interactions, in the cases I have highlighted, is clearly computational and informational: it is to transform inputs, to simplify search, to aid recognition, to prompt associative recall, to offload memory, and so on. In a sense, then, human reasoners are truly distributed cognitive engines: we call on external resources to perform specific computational tasks, much as a networked computer may call on other networked computers to perform specific jobs [20, pp. 68–69].

3.4. Socio-cultural situatedness

The last aspect of embodiment we will consider here is also the most complex, and the level at which the division between *embodied* and *situated* cognition no longer makes much sense. We have seen already the ways in which practical activity and interaction with the environment can itself be a cognitive strategy, thus in a very general way grounding cognition in agency. But for higher mammals, and especially humans, these interactions are always themselves situated in a broader social and cultural context. This means at least two things, the complex interrelation of which I have neither the space nor, probably, the acuity to detail here: (1) the interactions can take place not just with individual objects or artifacts, but also with persisting structures, which may be cultural and social, concrete and abstract (2) actions themselves can have not just immediate environmental effects, but social or cultural ones; that is, actions have meanings which must, of course, play a role in their deployment.

In this light, consider the concept of an altar: no set of objective features defines an altar, but rather the altar's place in a complex web of religious, social, institutional and comportmental relations. The proper way to approach an altar, and the sorts of behavior that are appropriate near and toward an altar, both define and demonstrate what it is.

It is here that cognitive science most directly borders fields like sociology and cultural studies. The fact that action has (social) meaning, and that agency takes place within a web of cultural structures not directly under the control of the actor, which we have come to (because of our bottom-up approach) only after many pages, is for these fields the ground truth on which their analysis depends. The ultimate success of the physical grounding project will require that these various domains be made consistent with one another (keeping in mind that a *reduction* of one to the other is the crudest and least likely possible form that such consistency might take); and this in turn is going to require far more interaction and collaboration between these fields than is currently common. *Philosophy in the Flesh* does attempt an excursion into this territory, in the form of several chapters on basic philosophical, ethical, and social concepts, and the ways in which they must be reconsidered in light of research on EC. Given the condescending tone and careless thinking which often marks these chapters it is probably best that they be simply ignored. [75,105,118,120] appear to offer some more promising beginnings on which such collaboration might build.

4. Neolithic intelligence

In considering the various aspects of embodiment most frequently employed in EC research, and their relation to what is perhaps the central project of EC, the investigation and specification of the physical grounding project, we have been able, I think, to discern—at least in outline—a continuous path from lower cognition to higher, from an intelligence that Brooks has called Cambrian to one we might therefore call Neolithic. In each case we have had our attention called to the myriad ways in which cognition depends upon the physical characteristics, ability, and situatedness of the agent. For humans, it seems, it is this last which is most crucial. That is, it may be the case that the unique reach and power of human—Neolithic—intelligence is a result not so much of a unique ability to perform complex, symbolic cognition in abstraction from the environment, but is rather due in large measure to the remarkable richness of the environment in which we do our thinking. Thus, although it is beyond my ability to say much more about socio-cultural situatedness *per se*, it is, I think, worth emphasizing the central role played by persisting institutions and practices in supporting the possibility of high-level cognition. In cognitive science such structures are called scaffolds; a scaffold, in this sense, occurs when an epistemic action results in some more permanent cognitive aid—symbolic, or social-institutional.

Clark [20] notes that we do very complex things, e.g., building a jumbo jet or running a country "only indirectly—by creating larger external structures, both physical and social, which can then prompt and coordinate a long sequence of individually tractable episodes of problem solving, preserving and transmitting partial solutions along the way" (p. 186).

These structures include language, especially written language, and indeed all physically instantiated representations or cognitive aids, such as maps, road signs, and the arrangement and labeling of supermarket aisles. Such scaffolds allow us to break down a complex problem into a series of small, easy ones, as when we write down intermediate results in long division, or memorize only as many abstract spatial directions as we know are necessary to get us to where we can employ the simpler strategy of following the road signs to downtown.

Not just symbol systems, but social structures and procedures can sometimes fill a similar role. The roles, hierarchy, and pre-set plays of a football team offer one simple instance. For any given play, each member of the team generally has a single, simple task—the blocking of a single player in a particular direction, for instance—and from the coordinated effort of the various players emerges a solution to the extremely complex and otherwise largely unconstrained problem of advancing the ball. Even considered from the viewpoint of a single player, the running back, this problem can be an extremely difficult one to solve given the rapidly changing and hostile environment of the problem space. If the back had to *decide* on each occasion where to run, it is likely that the extra time thereby required would make a solution impossible. The play acts to significantly reduce (pre-specify) the range of possible solutions (run off left tackle), and the role of the blockers is thereby dual: (1) to clear the physical path of the runner, and (2) to open the visual space so the back need not decide, but can *see* exactly where to run. In addition, there are procedures to change plays, and rules for who can do so, under what circumstances. In a well-run team, the hierarchy of coaches, assistant coaches, player coaches, and players, as well as the arrangement and contents of the play-book, allow each individual to focus on solving a limited, locally-specified problem, passing that solution on to the right person (or enacting it at the right time), thereby providing the means for others to solve their own tractable set of immediate problems.

From such examples, and a confidence that there is a specifiable continuity between such simple cases, and much more complex ones like the Boeing Corporation or the Government of the United States, Clark concludes:

> The mature cognitive competencies which we identify as mind and intellect may thus be more like ship navigation than capacities of the bare biological brain. Ship navigation emerges from the well-orchestrated adaptation of an extended complex system comprising individuals, instruments, and practices. Much of what we commonly identify as our mental capacities may likewise turn out to be properties of the wider, environmentally extended systems of which human brains are just one (important) part [20, p. 214].

Clark primarily emphasizes the role played in cognition by repeated interactions with a highly structured environment. But there is another aspect to these examples that is worth bringing to the fore: the first-personal perspective of the agent. In each of the examples above, the agent's behavior comes about only in so far as he is aware of the socially instantiated role he is playing (running back, or Chief Quartermaster) and regulates his behavior with respect both to the perceived situation and this understood role. Drawing attention to this subjective aspect of cognition is one important contribution of William Clancey's well-known book, *Situated Cognition* [19]. For Clancey, understanding the functional interactions between agent and environment

S0004-3702(03)00054-7/REV AID:1984 Vol.●●●(●●●) ELSGMLTM(ARTINT):m1a v 1.155 Prn:29/05/2003; 10:10 aij1984 P.22 (1-40) by:ML p. 22

> requires a special notion of *goal-driven* that involves a kind of *subjectivity*. This subjectivity is not realized as possessing a subset of facts about the world or misconceptions, as in descriptive models; rather it is a form of feedback between how the world is perceived and how the person conceives his or her identity. Conceptualizing situations, problems, and alternate actions inherently involves an aspect of self-reference in the perceptual-conceptual mechanism. That is, a person's understanding of 'What is happening?' is really 'What is happening *to me* now?' (p. 27, emphasis in original).

Like the other authors we have been discussing, Clancey questions the central hypothesis of GOFAI: the identification of knowledge with explicit, stored descriptions, and of cognition with the rule-based manipulation of said descriptions. He, too, emphasizes the importance of tacit, inarticulable 'knowledge' closely related to regulating behaviors and coordinating perceptual-motor interactions. But unlike some EC researchers, he continually focuses on understanding human-level cognition, which means also attending to, and casting some light on, human subjectivity and language from a situated perspective. While not incompatible with the main thrust of EC research, this approach nevertheless foregrounds slightly different themes, and ultimately suggests a richer notion of 'situated' than is sometimes evident in cognitive science literature. We saw, above, one example of how somewhat different conclusions can be drawn from the same interaction between agent and socially structured environment, depending on whether one emphasizes the role of the environment, or that of the agent's *understanding* of that environment (and the ways in which her self-conception influences that understanding).[20] The same dynamic can be seen in Clancey's approach to human language. Whereas above we emphasized the external role language can play, by conceptually structuring an environment or allowing for external storage of intermediate results in a complex cognitive process, Clancey emphasizes the *internal* importance of language as a cognitive aid.

> If human knowledge doesn't consist of stored descriptions, what then is the relation of what we say to what we do? Speaking must be seen not as bringing out what is already inside, but as a way of changing what is inside. Speaking is not restating what has already been posted subconsciously inside the brain, but is itself *an activity of representing*. Our names for things and what they mean, our theories, and our conceptions *develop in our behavior* as we interact with and re-perceive what we and others have previously said and done. This causal interaction is different from the linear 'describing what I perceive' or 'looking for what I conceive'. Instead, the processes of looking, perceiving, understanding, and describing are arising together and shaping each other. This is called the *transactional perspective* (p. 3, emphasis in original).

Thus, for Clancey, the 'situation' within which agency must be understood, and through which cognition takes place, has both internal and external dimensions. Indeed, an agent's 'situation' is itself not static, but arises in the course of dynamic transactions between

[20] We might call this Situated Self-Consciousness [6,96].

internal and external resources and structures which are to a large degree mutually constitutive. Who the agent perceives herself to be plays a role in what situation she sees herself in; likewise the perceived situation may affect which role an agent chooses to adopt.

> To summarize, cognition is *situated*, on the one hand, by the way conceptualizing relates to sensorimotor coordination and, on the other hand, by the way conceptualization, in conscious beings, is about the agent's role, place, and values in society. Thus, situated cognition is both a theory about *mechanism* (intellectual skills are also perceptual-motor skills) and a theory about *content* (human activity is, first and foremost, organized by conceptualizing the self as a participant-actor, and this is always with respect to communities of practice) (pp. 27–28, emphasis in original).

5. How embodied *is* cognition?: Caveats and criticisms

As I thought appropriate for a field review, and in accordance with my own judgments as to its worth and promise, I have so far been giving a largely sympathetic account of EC research. It will come as no surprise to the reader, however, that EC has received its share of criticism. One informal criticism—informal because although the objection has come up in nearly every discussion of EC-related research I have been involved in, I have not seen it worked out in detail in print—is that EC cannot be true because the physically disabled are obviously able to learn, acquire concepts, and communicate. If concepts were, indeed, related to, or their contents depended upon the specific physical capacities of embodied agents, then there ought to be detectable differences in the conceptual repertoire of differently abled individuals; yet there are no such differences.

It is perhaps worth saying, first of all, that no researcher in, or theorist of, embodied cognition has ever suggested that physical handicaps imply cognitive deficits. Nor, if there *were* differences in the conceptual contents or structures of differently abled individuals, would one expect them to be detectable at the level of the linguistic mastery displayed in conversational interaction (the usual evidence offered by those who object to EC along these lines).[21] Language and linguistically available concepts are highly abstract phenomena; one would therefore expect the criteria for participation in a linguistic community to be likewise somewhat abstract. Thus, the concept of 'walking', in so far as it is logically and semantically related to various concepts of movement, and given that examples of walking exist in, and can be easily seen in the environment, ought to be easily acquirable by an individual who cannot, and who perhaps never could, walk. The concept can be placed in a logical and semantic network which is on the whole grounded, even given that there is no specific experience of walking which directly grounds the concept. Everyone is able to understand things which they have not directly experienced, through imagination, analogy, demonstration, and testimony; the physically disabled are in this regard no different.

[21] IQ assessments of children with spinal muscular atrophy [124] bear this supposition out.

S0004-3702(03)00054-7/REV AID:1984 Vol.●●●(●●●) aij1984 P.24 (1-40)
ELSGMLTM(ARTINT):m1a v 1.155 Prn:29/05/2003; 10:10 by:ML p. 24

This suggests two further thoughts, which are worth airing. First, following out the theme of abstraction, it is worth recalling that the metaphorical groundings suggested by Lakoff and Johnson—for instance, the Purposes are Destinations mapping (Section 3)—rely not on specific experiences of particular modes of movement (the feeling of walking is like *this*) but the more abstract experience of moving through space. Concepts will often be grounded in these higher-order derivations of experience, in this case the experience of motility itself, rather than the (generally unconscious, but potentially attended to) feelings which accompany the particular mechanics underlying that motility. However, and this is the second thought, there *ought* to be differences in the particularities of grounding for differently abled individuals. We have seen an example of this already, in the kitten experiment [57] mentioned in Section 3; this experiment suggests that a sufficiently severe congenital disability might indeed be accompanied by difficulties in perceptual grounding, in attaching meaning to one's perceptual experience. Less dramatically, the experiment does suggest that there would be subtle differences in the perceptual experiences of differently abled individuals, at least in so far as the *affordances* presented by the environment would necessarily differ. Stairs afford something quite different to a person in a wheelchair, and to a person who can walk. I have no immediate suggestion as to how to test for this difference,[22] but it does seem as if EC is committed to there *being* such a difference; thus is EC committed to a set of presumably testable empirical claims.

More formal objections to EC are of two general sorts: objections to EC's objections to GOFAI (thus suggesting that EC is not well-motivated), and objections to EC research itself. In an instance of the first sort of criticism, Drew McDermott [80] argues that "*no working AI program has ever been bothered at all by the frame problem*[23]" (p. 116, emphasis in original). This may well be true, but insofar as it is, it is the result of building into the program (in one way or another) the unchanging context within which the reasoning will take place, and accepting the performance or domain limitations that this context-dependence implies. This is no criticism of McDermott's point, for as he himself asserts:

> The only practical solution to the prediction problem is to select a model of the situation and use it to make inferences. ... Any model must omit much of what will actually happen, such as the fate of the drain water or the consumption of electricity. Hence any real inference maker will tend to overlook potentially useful inferences. Occasionally a too-detailed model will be selected, and too many inferences will be made. Obviously

[22] But see, e.g., [64].

[23] The frame problem, as originally introduced [78] refers to the narrow issue of how to succinctly state, for a given action, not just what *does* change as a result of the action, but what does *not*. Very little changes as the result of a typical action, and one doesn't want to have to have to explicitly state for every fact that doesn't *that* it doesn't. (See [54] for further discussion.) But although this is the original meaning and scope of the frame problem, it has come in the literature to stand for a much broader class of problems. As a way of engaging with critics of AI, McDermott therefore expands the meaning of 'the frame problem', and breaks the problem into two parts. The first part, the 'inertia problem', is identical to the original frame problem. The second part, the 'prediction problem' is very closely related to the relevance problem, which I cited as a major motivation for the EC approach. McDermott suggests that the inertia problem has been essentially solved, and that the prediction problem is not really a problem.

S0004-3702(03)00054-7/REV AID:1984 Vol.●●●(●●●) aij1984 P.25 (1-40)
ELSGMLTM(ARTINT):m1a v 1.155 Prn:29/05/2003; 10:10 by:ML p. 25

> there are interesting technical questions about how the right model is selected and how it is cranked to generate results. But there is no interesting question about how to avoid generating an infinite number of irrelevant inferences (p. 119).

In other words, once we *decide* on a relevant model, the "frame problem" disappears. Perfectly true, except that the hard part of the problem appears to have been shifted to the issue of deciding on a model; for the question becomes: which model is relevant?[24]. (As Pylyshin suggests is typical of solutions to the frame problem, a crucial part of the mechanism is "on order" [104].) One way of reading EC research, on which it is perfectly compatible with McDermott's suggestion here, is as offering one approach to the question of how, from the bottom-up, to constrain both the models themselves, and an agent's choice between them. It is a basic tenet of EC research that when an agent *does* utilize explicit representations, those representations are localized, task-specific, and rooted in more basic methods of environmental coping. Thus, the EC approach satisfies the constraints facing any proposed solution to the frame problem. This is, of course, not the same as *providing* a solution; it means only that it cannot be determined *a priori* that such an approach will fail. There are many hard questions to be faced, such as how representations can be derived from perceptual experience, how these representations are to be related to action, and how it might be possible to reason with representations generated in this manner. The physical grounding project is difficult, complex, and far from complete.

This brings us to critiques of EC itself, the most important of which is the claim that no complex, intelligent creature can actually get by entirely without representations. David Kirsh's commentary is typically enlightening:

> Situationally determined activity has a real chance of success only if there are enough egocentrically perceptible cues available. There must be sufficient local constraint in the environment to determine actions that have no irreversibly bad downstream effects. Only then will it be unnecessary for the creature to represent alternative courses of actions to determine which ones lead to dead ends, traps, loops, or idle wandering.
>
> From this it follows that if a task requires knowledge about the world that must be obtained by reasoning or by recall, rather than by perception, it cannot be classified as situation determined. Principle candidates for such tasks are:
>
> (1) Activities which involve other agents, since these often will require making *predictions* of their behavior.
> (2) Activities which require response to events and actions beyond the creature's sensory limits, such as taking precautions now for the future, avoiding future dangers, contingencies, idle wandering—the standard motive for internal look ahead.
> (3) Activities which require understanding a situation from an objective perspective such as when a new recipe is followed, advice is assimilated, a strategy from one

[24] None of this is to mention the ever-lurking problem of what to do when one's model proves inaccurate. EC is committed to "the ubiquitous biological design principle: oversimplify and self-monitor" [26, p. 4]

S0004-3702(03)00054-7/REV AID:1984 Vol.●●●(●●●) ELSGMLTM(ARTINT):m1a v 1.155 Prn:29/05/2003; 10:10 aij1984 P.26 (1-40) by:ML p. 26

context is generalized or adapted to the current situation. All these require some measure of conceptualization.

(4) Activities which require some amount of problem solving, such as when we wrap a package and have to determine how many sheets of paper to use, or when we rearrange items in a refrigerator to make room for a new pot.

(5) Activities which are creative and hence stimulus free, such as much of language use, musical performance, mime, self-amusement.

These activities are not isolated episodes in normal human life. Admittedly, they are all based on an underlying set of reliable control systems; but these control systems are not sufficient themselves to organize the *global* structure of the activity [69, p. 171].

As I have indicated already (Sections 2 and 3) I think that a great deal of work in EC is in fact vulnerable to this criticism, and significant research effort ought to be devoted to the study of hybrid systems, which not only have both reactive and deliberative elements, but in which these elements are systematically connected and able to cooperate in guiding the behavior of artificial agents. Such research is in its infancy; as mentioned above open issues include such basic questions as understanding in detail what embodied representations *are* [94,106], how they might be generated from perceptual and proprioceptual data [89], and, once generated, how one might usefully reason with them [5,90]. Further, there is the related but independent question of how (architecturally or otherwise) to coordinate the activity of logic-based reasoning components with other ‘non-symbolic’ reasoning components (say, bayes-nets) and also with more basic, perhaps reactively-oriented control layers. Brooks’ subsumption architecture [15, Chapter 1] offers some promising insights as to how to coordinate disparate functional components without requiring a globally shared representational scheme (which would cause the problems we have been discussing to re-surface), but it is not entirely clear how one would (or whether one could) extend his architecture to include layers which reason with explicit representations.

Like the informal criticism mentioned above, such formal criticisms of EC—at least those of which I am aware—are valid and useful, not because they invalidate EC as a viable research program, but because, in identifying the limits of certain narrowly conceived *versions* of the program (for there are purists in every field), they suggest new avenues for future exploration. A somewhat more fanciful critique of situated cognition can be found in [95]. That article, offered in the form of a fable, argues that situated cognition threatens to throw the baby out with the bathwater. In a clever and subtle use of symbolism, the baby (RepBaby) represents representation, and situated cognition is played by a nanny (SitNanny) under the influence of disreputable European ideas. Unfortunately, the central rhetorical strategy of the fable is to offer the most uncharitable possible interpretation of research in situated cognition (SC), reject it, and claim that the more reasonable interpretation of SC was an element of AI research all along. Thus:

> [O]ne can agree that AI should pay more attention to real-time activities that are intimately involved with the immediate environment, without feeling a need to reject the entire framework of empirical science, embrace radical social constructivism, or reject the idea of mental representation (p. 17).

ARTICLE IN PRESS
S0004-3702(03)00054-7/REV AID:1984 Vol.●●●(●●●) aij1984 P.27 (1-40)
ELSGMLTM(ARTINT):m1a v 1.155 Prn:29/05/2003; 10:10 by:ML p. 27

Indeed, I should hope so. Much of my own earlier work, summarized in [92], aims to show that the way to avoid certain radical anti-realist conclusions is precisely to take up into epistemology some of the ideas underlying work in situated and embodied cognition. As far as I am concerned, therefore, the attempted association of EC/SC research with "radical social constructivism" and the rejection of empirical science is a non-starter, and I will not discuss it further here. The issue of whether SC means to "reject the idea of mental representation", however, is somewhat more complicated, and although I have already indicated my ultimate position on the matter, discussing it in the current context will help clear up some perhaps common concerns and misconceptions.

The authors' main substantial worry seems to be that:

> The situationalists are attacking the very *idea* of knowledge representation—the notion that cognitive agents think about their environments, in large part, by manipulating internal representations of the worlds they inhabit (p. 17, emphasis in original).

What keeps me from denying this characterization of SC outright is the phrase "in large part". Theorists in SC probably *do* deny that representations play a *large* part in cognition, whereas Hayes, et al. apparently believe that representations play a *very* large part:

> Internal representations might not be consciously available to introspection, might utilize ontological frameworks that are determined by social or other contexts, might be involved with (and have their use involved in) social practices or any other kind of human activity, and might be involved in perceptual and motor skills at any cognitive level. None of these are in any way at variance with the representationalist hypothesis.

If representation can mean all this, then it seems clear that it plays a large part in cognition, indeed. In fact, it isn't clear to me, especially given the possibility of unconscious representations and those "involved in perceptual or motor skills at any level", what sort of cognitive or environmentally reactive mechanism *couldn't* be construed as utilizing internal representations. Does the coil in my thermostat represent the temperature? Sure, I guess. But of what use is such a broad notion of representation? I myself, given a box of this size, would start to make divisions, say: representations governed by direct causal coupling (thermostat coils and Watt governors); distributed representations like the activation patterns propagating through a simulated neural network; images or picture-like representations; and descriptions or language-like representations. With a little thought, and a bit of reading in the relevant literature ([9,36–38]; [101] offers a fairly comprehensive bibliography) I might come up with more distinctions, and perhaps an overall account of these varieties of representation, and the cognitive importance of each type. Depending on my conclusions, I might even be tempted to stop calling some of these things representations, reserving that term for more specialized use. Indeed, as Hayes, et al. report in a wondering tone, some SC researchers have done just that:

> In fact, several of the new critics (some having noisily jumped ship) take representation to mean something like a text or a picture, and object consciously manipulated by talking to oneself or visualizing internal imagery, and identical in nature to external

> representations such as writing or diagrams. . . . But RepBaby was not brought up this way (p. 19).

While it is true that the original notion of a 'symbol'—as proposed in the physical symbol systems hypothesis [7,85,86]—was quite broad, given the distinctions suggested above, the decision to restrict the use of 'representation' to a smaller sub-set of inner states seems reasonable and motivated, and has the added virtue of conforming to common AI practice (as illustrated in the central hypothesis of GOFAI, Section 1). Hayes et al. may have good reasons for continuing to advocate a broader usage, but if so those reasons are not evident here, and the reader is left wondering if its main virtue is allowing the authors to accept the central claims of SC without appearing to have shifted ground.

The authors treat their other potentially substantive issue, of whether, colloquially put, representations are in the head, in a similar fashion.

> Let's take a simple example. A plumber and a customer who knows nothing about plumbing are together in a kitchen. Where is the knowledge of plumbing? RepBaby has an almost childishly simple answer to this question: It's in the plumber's head. Is RepBaby right? (p. 19).

Well, naturally it depends on what is meant by "in the head". Thus the authors move to make this notion as broad as possible, in two ways. First:

> Memories provide an even more vivid illustration of beliefs being in the head. We all have memories that no one else shares. Surely, these are not located anywhere but in our own heads (or at most our own bodies) (p. 20).

Apparently, "in the head" can also mean "in the body", despite the fact that accepting the notion that memories, or other internal representations, might be somehow distributed in, or otherwise require the capacities of, the body as a whole suggests a fairly radical revision of traditional, mentalistic cognitive science. (See, e.g., [121] for a discussion.) But never mind that detail. Having thus broadened our heads physically, the authors move to expand them socially, as well:

> One might argue that although there are things in the head we call representations, any account of the *meaning* of these things (and hence any account of why they should be regarded as representations) must be inherently social. Our beliefs have meaning only by virtue of their role in a society of which we are a part. Wittgenstein argued that the idea of a language spoken by a single person is incoherent: languages are systems of communication between agents. . . . The old AI idea of common sense might fit easily here: what more natural place to locate the sense common to the members of a society than in that society itself? We are quite sympathetic to this view, properly understood" (p. 20).

Far from being a criticism of SC, this looks to me like the sort of thing that a theorist of SC—or, more precisely, an externalist about meaning—might actually believe, although,

S0004-3702(03)00054-7/REV AID:1984 Vol.●●●(●●●) aij1984 P.29 (1-40)
ELSGMLTM(ARTINT):m1a v 1.155 Prn:29/05/2003; 10:10 by:ML p. 29

unlike RepBaby, he would be inclined to conclude that meaning properly understood, is therefore *not* in the head [97,103]. But let's not mind that detail, either. With this idea claimed for traditional AI, and our representing heads suitably expanded to encompass the requisite physical and social territory, what is left for SitNanny to believe is "that the representational tokens themselves aren't in the head or that representational tokens can only have an external, social existence, or even that there isn't any representation at all" (p. 20). SitNanny may believe this, but I don't know of a single SC researcher who does. I conclude from the absence of a citation for this claim that the authors don't know of one, either.

Rather more substantial versions of the central concerns of Hayes et al. can be found in [123]. Vera and Simon see no necessary incompatibility between the approach to cognition advocated by proponents of situated action and that advocated by proponents of physical symbol systems; more than this, they see the former as an instance of the latter.

> The proponents of SA, in contrasting it with the established symbolic viewpoint in cognitive science, have at one time or another made the following claims:
> (1) SA, unlike symbolic computation, requires no internal representations.
> (2) SA operates directly between the agent and the environment without the mediation of internal plans.
> (3) SA postulates direct access to the "affordances" of the environment. That is, the actor deals with the environment, and with his or her own actions, in purely functional terms.
> (4) SA does not use productions.
> (5) The environment, for purposes of SA, is defined socially, not individually or in physicalist terms.
> (6) SA makes no use of symbols [123, p. 38].

Vera and Simon deny each of these claims, but in so doing they interpret the disputed concepts in a way that is, to my mind, quite favorable to the situated, or embodied, research program. Thus, for instance, their understanding of an internal representation is extremely broad, and allows for representations which are very minimal, action-oriented, and functionally or indexically characterized. It is this understanding of internal representation which is central to their denial of items (1), (3), and (6), above:

> In some situations, an actor's internal representations can be extremely simple, but no one has described a system capable of intelligent action that does not employ at least rudimentary representations. Perhaps the barest representation encompasses only goals and some symbolization of a relation between goal and situation, on the one hand, and action in the other.... In systems like Pengi [2] and the creatures of Brooks [15], often taken as paradigmatic examples of applied SA, there are substantial internal representations, some of them used to symbolize the current focus of attention and the locations of relevant nearby objects, others used to characterize the objects themselves in terms of their current functions (pp. 38–39).

Or, as they say earlier: "That the symbols in question are both goal-dependent and situation-dependent does not change their status. They are genuine symbols in the traditional information-processing sense.... 'The-bee-that-is-chasing-me' is a perfectly good symbol" (p. 37). Since one main aim of research in EC has been to understand and advocate functionally characterized and action-oriented inner states, this clarification by the central proponents of the symbolic approach to AI is no doubt very welcome.

In a similar vein, the authors accept the centrality of affordances to guiding agency, but insist that these affordances be understood symbolically, i.e., in terms of internally encoded states.

> We have already seen that when people are dealing with familiar situations, using habitual actions, their internal representations, at the conscious level, may be almost wholly functional, without any details of the mechanisms that carry out these functions. The "affordances" of the environment, represented internally, trigger actions.
>
> Of course, the absence of consciousness of mechanisms implies neither that mechanisms are absent nor that they are non-symbolic. To acquire an internal representation of an affordance, a person must carry out a complex encoding of sensory stimuli that impinge upon eye and ear. And to take the corresponding action, he or she must decode the encoded symbol representing the action into signals to the muscles (p. 41).

Given the broad view of symbols adopted by the authors, as any internal pattern, however instantiated, that designates or denotes, it seems likely that detecting and reacting to affordances will, indeed, turn out to involve internal symbols. We should note, however, that on this understanding the symbols involved may include not just unconscious states, but processes taking place primarily in the central nervous system as a whole, and perhaps only minimally involving the brain. Whatever the ultimate utility of calling such processes symbolic, we should at least be aware of the great distance between the view of cognition put forward here by Vera and Simon, and that summarized in the central hypothesis of GOFAI, as defined in Section 1.

Finally, with regard to understanding the importance of repeated interactions with the environment, and the role of external scaffolds in supporting intelligence, the authors write:

> If the SA approach is suggesting simply that there is more to understanding behavior than describing internally generated, symbolic, goal-directed planning, then the symbolic approach has never disagreed: "The proper study of mankind has been said to be man. But ... man—or at least the intellective component of man—may be relatively simple; ... most of the complexity of his behavior may be drawn from his environment, from his search for good designs [109, p. 83]" (p. 23).

Speaking just for myself, I lose interest in a debate between two theoretical approaches when it devolves into deciding who first suggested the position which both sides agree to be correct.

What should be heartening about these articles, at least to the EC/SC community, is the tacit suggestion that the main ideas behind EC are reasonable and widely accepted, even by those who count themselves among its opponents. Indeed, it may perhaps be said that, but for those like Lenat and the developers of Cyc who apparently cling to the notion that intelligence can be axiomatized *all the way down*, as it were, the main tension between those who identify themselves as supporters of the EC/SC approach and those who identify with more traditional AI is over the issue of just exactly *how much* (and which parts) of intelligence can be accounted for by representation and computation, and how much by reaction and interaction.

What is troubling is that, in their enthusiasm to maintain the compatibility of traditional AI with EC research, Hayes et al. and Vera and Simon have gathered some pretty disparate ideas under the same big tent, thereby potentially submerging some important tensions and difficulties. For in point of fact, there are significant challenges facing the study of (embodied) cognition, and the outlines of these can be discerned in the authors' over-generalizations: How will the various items in their grab-bag labeled 'representations' be related to one another? How can a conscious, explicit representation encode information about or from an unconscious, inexplicit one? Under what circumstances might such a relation be needed? What does it mean for a representation—say, a motor representation of an action—to be encoded *in the body* rather than in the head? These are hard questions, and answering them will involve not just technical, but also conceptual work, at least some of which is likely to uncover basic confusions and thereby require revisions of the EC program. Given the credentials of the authors, I look forward to seeing from them articles—even, or especially, ones critical of EC/SC—which substantively explore these issues. It is through such work that the physical grounding project will be advanced.

6. Leveraging intelligence: Scaffolding and design

As we have seen, one extremely important method for increasing the reach and power of simple human intelligence is through the use of external (and internal) scaffolds, both natural and designed. Yet the scaffolds by and through which we use computing machines to enhance human cognition have been designed largely under the influence of an outdated and likely incorrect view of the nature of that cognition. Thus is the extremely important area of Human-Computer Interaction in need of its own overhaul; or so argues Paul Dourish in *Where the Action Is* [29].

Dourish begins by noting the fact that "computer science is based entirely on philosophy of the pre-1930's" (p. vii), but his motivation for re-thinking this state of affairs, for wishing to include in computer science the insights about human cognition and being-in-the-world coming from such thinkers as Heidegger and Wittgenstein, is eminently practical. "We need", he writes, "new ways of interacting with computers, ways that are better tuned to our needs and abilities" (p. 2). Anyone who has had trouble using their desktop PC may readily agree, but Dourish is after something more universal than fine-tuning the traditional desktop interface. Instead, he notes the near ubiquity of computer processors in our lives—microwaves, cell phones, automobiles—and asks the basic question of how we might more fruitfully employ all this largely unused computer power. More precisely, he asks which sets of human skills computing devices should be designed to exploit, for he sees the history of computing in terms of a succession of answers to this question. He divides that history into six eras, four past: Electrical; Symbolic; Textual; and Graphical and two, he hopes, future: Tangible and Social computing; and Embodied Interaction. We will briefly explain the first four, but like Dourish, will spend the bulk of our efforts coming to understand the last two.

In the Electrical era, interacting with a computer required a thorough understanding of electronic design. At the time, a computer could do one thing, and that thing was dictated by the configuration of its circuits. To get it to do something else meant re-configuring those circuits, and so required knowing which plugboards and patch cables were required in what arrangements to achieve the desired end.

A move to more Symbolic form of interaction represented a step forward. In this era, computer hardware began to become more standardized, and could be understood more abstractly in terms of capacities like registers and accumulators, the particular implementations of which therefore mattered less to the user. This development allowed for the introduction of assembly languages, "essentially symbolic forms of machine language ... so that a sequence of instruction codes such as 'a9 62 82 2c' is rendered as a symbolic expression such as 'mov1 (r1+), r2' " (p. 7).

Expanding the range and power of symbolic interaction resulted in the Textual era, in which the model for human computer interaction came to be written language. Although we are still a long way from true natural language interaction, still command-line interfaces like Unix do employ a large vocabulary and a simple grammar, thus "drawing on our abilities to create meaningful sentences" (p. 10) in the course of our interaction.

Although the command-line interface does allow room for a wider range of human capacities to be brought to bear in computing, perhaps the more important upshot of the Textual era was the way in which it "brought the idea of 'interaction' to the fore" (p. 10).

> Textual interaction drew upon language much more explicitly than before, and at the same time it was accompanied by a transition to a new model of computing, in which a user would actually sit in front of a computer terminal, entering commands and reading responses. With this combination of language use and direct interaction, it was natural to look on the result as a 'conversation' or 'dialogue'. These days, this idea of dialogue is central to our notion of 'interaction' with the computer, replacing configuration, programming, or the other ideas that had largely characterized the interplay between users and systems in the past (pp. 10–11).

Of course, the era with which we are most familiar is the present Graphical era. Like the eras before it, the Graphical era opened the way for us to bring still more human capacities to bear in interacting with our computers. Most importantly, by providing a two-dimensional virtual space, the graphical user interface allows us to manage information by managing that virtual space.

This move from a linguistic to a spatial orientation brings us to tangible computing, one of the two current research programs in HCI which for Dourish pre-sage the era of Embodied Interaction. Theorists of tangible computing call attention to the physicality of the things that we typically interact with, even when we are interacting primarily with symbols (say, reading a medical chart) or interacting symbolically (having a conversation). For the physical features of an information object may themselves carry information which is not necessarily symbolically encoded, or may dictate how one should interpret the symbolically coded information. So, for instance, in a medical chart the use of pencil rather than ink may indicate that the information is tentative; the state of the chart itself may indicate how often it has been used; and of course the handwriting can reveal which personnel have been involved in the patient's treatment. Further, possession of the chart, being exclusive (there being only one chart to possess) is itself significant, and is bound up not simply with the question of information availability, but also with issues of power and control. Thus not just the information, but the physical instantiation of that information, shape medical practice as it pertains to the patient.

Dourish details such cases where the physical attributes of information-bearing objects play an important role in how that information is interpreted and incorporated into information-intensive practices like delivering medical care, controlling air traffic, and the like. The point, understood as a critique of current computer interfaces, is that we in fact have much more complicated ways of dealing with information and information-carrying objects than is allowed for in a typical computer set-up. So, for instance, my PC has a single mouse with which I can do a single thing. My interaction with the virtual desktop is in this way highly focused, and therefore limited. In contrast, my interaction with my actual desktop often utilizes both hands at once, and I employ a more complex spatial arrangement of objects than is possible on my computer screen. For instance, I have special-purpose areas (bills, pending work, etc.) and use stacking and cross-stacking to indicate what needs attention, and in what order.

> Not only is there not a single point of interaction, there is not even a single device that is the object of interaction. The same action might be distributed across multiple devices, or, more accurately, achieved through the coordinated use of those artifacts. Imagine sitting at your desk to write a note. The writing comes about through the coordinated use of pen, paper, and ink, not to mention the desk itself and the chair you sit in; you might write on the page with your dominant hand while the nondominant hand is used to orient the page appropriately. These are all brought together to achieve a task; you act at multiple points at once (p. 51).

In general, tangible computing seeks to bring computation and information more fully into the physical world, with which we were evolutionarily equipped to easily cope. A nifty example of tangible information design is the marble answering machine. For each message received, the machine associates that message with a marble, which is then rolled down a track. When you arrive home, the number of marbles in the track indicates the number of messages, and dropping a marble into the slot marked 'listen' will play the message. The beauty of the design is that the typical procedures which one would have to (frustratingly) learn how to perform with a typical answering machine are now quite natural: to play any given message, select a marble from the track and drop it in the play slot. Delete a message by putting the marble back in the general pile; save it by putting it back in the track (or putting it in your pocket!); reorder the messages (perhaps to indicate their relative importance) by reordering the marbles. The marble answering machine thus models our interaction with information on the intuitive procedures we have developed for coping with the physical world.

It is important to notice the ways in which tangible computing is arrayed against another powerful trend in HCI: virtual reality. The object of virtual reality is to bring the user more completely into the realm of the computer, leaving behind the tangible physical world. In its more extreme versions, it is guided by the notion that humans will—or must—take a quasi-evolutionary next step and adjust to the requirements of computing machinery [55]. No doubt our adjustment to the ways of computers is to some degree inevitable, but Dourish suggests that the demand is rooted in the thinking of a long-past era when computer time was much more expensive than human time. Under these conditions, it made sense to ask the computer to do only what it did fastest and best, even if that meant burdening the user with an awkward and difficult interface. VR maintains this assumption that we have to meet the computer on its own terms, and hopes to design interfaces which will make this meeting as smooth as possible. In contrast, tangible computing is dedicated to re-considering the nature and uses of computation, and the design of computing devices, so that we can bring the computer more fully into *our* world.

> The tangible computing work attempts to capitalize on our physical skills and our familiarity with real-world objects. It also tries to make computation manifest to us in the world in the same way as we encounter other phenomena, both as a way of making computation fit more naturally with the everyday world, and as a way of enriching our experiences with the physical. It attempts to move computation and interaction out of the world of abstract cognitive processes and into the same phenomenal world as our other sorts of interactions (pp. 102–103).

For Dourish, the natural end of tangible and social computing (the latter of which we have not specifically discussed, but which is concerned with understanding, and designing in the light of, the social situatedness of information, and the social significance of information-using practices) is computation as Embodied Interaction. As is indicated above, this will involve making computational devices which are capable of receding sufficiently into the background that it is possible to achieve our purposes with and through them without very much *noticing* them, the way a good hammer allows one to pound nails without paying any attention to the tool itself. It also means making computation into a scaffold as ubiquitous and easily utilized as street signs.

According to Dourish, the only way to make this possible is to better understand the nature of our world, emphasis on *our*, that is, the lived world of our experience. Consistent with the other writers we have been discussing, Dourish believes that the phenomenological tradition—in which he includes such thinkers as Husserl, Heidegger, Schutz, Merleau-Ponty and even Wittgenstein—is an important resource for anyone hoping to understand this lived world.

At the center of phenomenology is the notion of embodiment, and the basic themes should be by now familiar: embodiment means not just having, and acting through, some physical instantiation, but recognizing that the particular shape and nature of one's physical, temporal and social immersion is what makes meaningful experience possible. As Dourish puts it, "Embodiment is the property of our engagement with the world that allows us to make it meaningful", and thus "Embodied Interaction is the creation, manipulation, and changing of meaning through engaged interaction with artifacts" (p. 126).

For Dourish, Heidegger is the first theorist of Embodied Interaction, for it was he who rejected the lingering Cartesianism in Husserl, the "abstract, disconnected intentionality ... of a Cartesian homunculus peering out at the world and seeing what's there" (p. 108), and replaced it with a radical re-thinking of how the world reveals itself through our encounters with it. Heidegger stressed that we do not encounter an alien world of abstract properties, or a meaningless stream of sense data which we must work to make meaningful and coherent; rather, in our everyday experience, the world is already structured in familiar and meaningful terms. For Heidegger, the most important element of our encounter with the world, which accounts for a good deal of the structure of experience, is practical interaction.

> The kind of dealing which is closest to us is as we have shown, not a bare perceptual cognition, but rather that kind of concern which manipulates things and puts them to use [56, p. 95].

But as we noted earlier, embodiment has not just practical but also social significance, for the construction of meaning, of the terms through which we encounter the world, is not generally private, but is rather a shared and social practice. Although this notion of being-with-others is also an important part of Heidegger's work, it is to Alfred Schutz that Dourish turns for help in understanding this aspect of our embodiment.

> [Schutz] brought phenomenological reasoning to the problems of sociology.... Taking the life-world as a focus recasts the problems of sociology. It means turning away from the idea of general laws that operate outside the immediate purview of the actors whose behavior they regulate. Instead, it characterizes sociological problems as practical, mundane ones routinely encountered—and solved—by social actors in the course of their day-to-day activities [29, p. 113].

This implies, among other things, a methodology which includes close attention to the details of actual practice (as opposed to the formalizations or abstract procedures that might be found in a corporate handbook). This, combined with attention to the ways in which physical objects do—or don't—fit easily into these practices, forms the center of Dourish's design philosophy. The upshot is straightforward, if difficult to implement in practice: artifacts meant to help humans with information-intensive tasks must be designed in light of the facts of human cognitive practices, and this means in particular understanding that human cognition grows out of and relies on a very practical, interaction-dependent orientation to the world.

Although Dourish competently and engagingly summarizes the main philosophical ideas behind EC, there are, for the more theoretically minded, better, more detailed expositions [20,31,49,117]. However, I have yet to encounter a better work for the practical-minded computer scientist who, having heard—from a distance, as it were—the noise of the most recent cognitive revolution, wonders what it might mean for her.

7. Conclusion

Research in Embodied Cognition offers the outline of a new, comprehensive approach to explaining, and perhaps eventually reproducing, Neolithic intelligence. Although related to and continuous with situated cognition, EC takes the physical grounding project as its central research focus. This project calls for detailing the myriad ways in which cognition depends upon—is grounded in—the physical characteristics, inherited abilities, practical activity, and environment of thinking agents.

Against the Cartesian claim that we are radically distinct from animals, uniquely possessed of a soul and its attendant powers of abstract reason, EC maintains our evolutionary continuity. We, like all animals, are essentially embodied agents, and our powers of advanced cognition vitally depend on a substrate of abilities for moving around in and coping with the world which we inherited from our evolutionary forbears.

Against the cognitivist claim that cognition is the rule-based manipulation of abstract representations, EC maintains that there is much more to cognition than mental representation. Cognition exploits repeated interaction with the environment, not only using the world as its own best model, but creating structures which advance and simplify cognitive tasks. The explicit representations cognition does employ are generally limited, physically grounded and oriented toward the specific needs of a given agent.

Along with research in situated cognition, EC further suggests that intelligence lies less in the individual brain, and more in the dynamic interaction of brains with the wider world—including especially the social and cultural worlds which are so central to human cognition—and therefore suggests that fields like sociology and cultural studies can themselves be important resources for (and in some guises are part of) the cognitive sciences.

And finally, we have seen that this re-conception of human cognition has implications not just for the project of creating artificial intelligence, but for the related project of harnessing computation to enhance human intelligence. Whatever the next step is to be in human cognitive progress, it ought to be based on a better and more thorough understanding of intelligence than we have so far managed. Research in EC promises one important component of that eventual understanding.

Acknowledgements

I would like to thank Athena Coleman, Art Graesser, Ken Hennacy, Terry Horgan, Cara Johnson, Roger Kreuz, Max Louwerse, Tim Oates, Don Perlis, Gregg Rosenberg, and John Tienson for helpful discussions of these issues. This work was supported in part by grants from AFOSR and ONR.

References

[1] Behavioral Brain Sci. (Special issue on Direct Perception) 3 (1980).
[2] P.E. Agre, D. Chapman, Pengi: An implementation of a theory of activity, in: Proceedings AAAI-87, Seattle, WA, AAAI Press, Menlo Park, CA, 1987, pp. 268–272.

[3] P. Agre, Computation and Human Experience, Cambridge University Press, Cambridge, 1997.
[4] M.L. Anderson, D. Josyula, Y. Okamoto, D. Perlis, Time-situated agency: Active logic and intention formation, in: Workshop on Cognitive Agents, 25th German Conference on Artificial Intelligence, 2002.
[5] M.L. Anderson, T. Oates, How an agent learns to use its head, in preparation.
[6] M.L. Anderson, D. Perlis, Situated self-consciousness, in preparation.
[7] M.L. Anderson, D.R. Perlis, Symbol systems, in: L. Nadel, D. Chalmers, P. Culicover, B. French, R. Goldstone (Eds.), Encyclopedia of Cognitive Science, Macmillan Publishers, London, 2002.
[8] R.C. Arkin, Behavior-based Robotics, MIT Press, Cambridge, MA, 1998.
[9] W. Bechtel, Connectionism and the philosophy of mind, in: W.G. Lycan (Ed.), Mind and Cognition: A Reader, Basil Blackwell, 1990, pp. 252–273.
[10] J. Bermúdez, A.J. Marcel, N. Eilan, The Body and the Self, MIT Press, Cambridge, MA, 1995.
[11] M. Bhatia, P. Chi, W. Chong, D.P. Josyula, M. O'Donovan-Anderson, Y. Okamoto, D. Perlis, K. Purang, Handling uncertainty with active logic, in: Proceedings AAAI Fall Symposium on Uncertainty in Computation, 2001.
[12] E. Bizzi, S.F. Giszter, E. Loeb, F.A. Mussa-Ivaldi, P. Saltiel, Modular organization of motor behavior in the frog's spinal cord, Trends in Neuroscience 18 (1995) 442–446.
[13] P. Bordieu, Practical Reason, Stanford University Press, Palo Alto, CA, 1998.
[14] C. Breazeal, Designing Sociable Robots, MIT Press, Cambridge, MA, 2002.
[15] R. Brooks, Cambrian Intelligence: The Early History of the New AI, MIT Press, Cambridge, MA, 1999.
[16] R. Brooks, Flesh and Machines: How Robots Will Change Us, Pantheon, New York, 2002.
[17] J. Cassell, J. Sullivan, S. Prevost, E. Chuchill, Embodied Conversational Agents, MIT Press, Cambridge, MA, 2000.
[18] W. Chong, M. O'Donovan-Anderson, Y. Okamoto, D. Perlis, Seven days in the life of a robotic agent, in: Proceedings of the GSFC/JPL Workshop on Radical Agent Concepts, 2002.
[19] W.J. Clancey, Situated Cognition: On Human Knowledge and Computer Representations, Cambridge University Press, Cambridge, 1997.
[20] A. Clark, Being There, MIT Press, Cambridge, MA, 1996.
[21] A. Clark, Embodied, situated and distributed cognition, in: W. Bechtel, G. Graham (Eds.), A Companion to Cognitive Science, Basil Blackwell, 1999, pp. 506–517.
[22] G. Cohen, Memory in the Real World, Erlbaum Press, Hillsdale, NJ, 1996.
[23] Cottingham, Stoothoff, Murdoch, Kenney, The Philosophical Writings of Descartes, Cambridge University Press, Cambridge, UK, 1991.
[24] Cycorp, The cycorp website, 2002, http://www.cyc.com/products2.html.
[25] A. Damasio, The Feeling of What Happens: Body and Emotion in the Making of Consciousness, Harvest Books, New York, 2000.
[26] D. Dennett, Producing future by telling stories, in: K.M. Ford, Z.W. Pylyshyn (Eds.), The Robot's Dilemma Revisited: The Frame Problem in Artificial Intelligence, Ablex Publishing, Norwood, NJ, 1996, pp. 1–7.
[27] D.C. Dennett, Cognitive wheels: The frame problem of AI, in: C. Hookaway (Ed.), Minds, Machines and Evolution, Cambridge University Press, Cambridge, 1984, pp. 129–151.
[28] R. Descartes, The world, or treatise on light, in: M. Cottingham, M. Stoothoff, M. Murdoch, M. Kenney (Eds.), The Philosophical Writings of Descartes, Cambridge University Press, Cambridge, 1991.
[29] P. Dourish, Where the Action Is: The Foundations of Embodied Interaction, MIT Press, Cambridge, MA, 2001.
[30] H.L. Dreyfus, What Computers Can't Do: A Critique of Artificial Intelligence, Harper and Row, New York, 1972.
[31] H.L. Dreyfus, Being-in-the-World: A Commentary on Heidegger's Being and Time, Division 1, MIT Press, Cambridge, MA, 1991.
[32] H.L. Dreyfus, What Computers (Still) Can't Do: A Critique of Artificial Reason, MIT Press, Cambridge, MA, 1992.
[33] H.L. Dreyfus, Intelligence without representation: Merleau-Ponty's critique of mental representation, Phenomenol. Cognitive Sci. 1 (4) (2002) 367–383.
[34] G. Evans, The Varieties of Reference, Oxford University Press, Oxford, 1982.
[35] G. Fauconnier, M. Turner, How We Think: Conceptual Blending and the Mind's Hidden Complexities, Basic Books, New York, 2002.

S0004-3702(03)00054-7/REV AID:1984 Vol.●●●(●●●) aij1984 P.38 (1-40)
ELSGMLTM(ARTINT):m1a v 1.155 Prn:29/05/2003; 10:10 by:ML p. 38

[36] H. Field, Mental representation, Erkenntnis 13 (1978) 9–61.
[37] J. Fodor, Representations, MIT Press, Cambridge, MA, 1981.
[38] J. Fodor, A Theory of Content and Other Essays, MIT Press, Cambridge, MA, 1990.
[39] S. Franklin, Artificial Minds, MIT Press, Cambridge, MA, 1995.
[40] W.J. Freeman, Societies of Brains, Lawrence Erlbaum, Hillsdale, NJ, 1995.
[41] W.J. Freeman, How Brains Make Up Their Minds, Weidenfeld and Nicolson, London, 1999.
[42] W.J. Freeman, R. Núnez, Reclaiming Cognition: The Primacy of Action, Intention and Emotion, Imprint Academic, Thorverton, UK, 1999.
[43] D.M. Gabbay, R. Nossum, M. Thielscher, Agents in proactive environments, J. Logic Comput. 9 (1) (1999) 25–45.
[44] A.P. Georgopoulos, R.E. Kettner, A.B. Schwartz, Primate motor cortex and free arm movements to visual targets in three-dimensional space. II. Coding the direction of movement by a neuronal population, J. Neurosci. 8 (1988) 2928–2937.
[45] A.P. Georgopoulos, A.B. Schwartz, R.E. Kettner, Neuronal population coding of movement direction, Science 233 (1992) 1416–1419.
[46] J.J. Gibson, The Ecological Approach to Visual Perception, Houghton Mifflin, New York, 1979.
[47] G. Giralt, R. Chatila, M. Vaisset, An integrated navigation and motion control system for autonomous multisensory robots, in: Brady, Paul (Eds.), Robotics Research 1, MIT Press, Cambridge, MA, 1984, pp. 191–214.
[48] R.V. Guha, D.B. Lenat, Cyc: A midterm report, AI Magazine 11 (3) (1990) 33–59.
[49] C.B. Guignon, Heidegger and the Problem of Knowledge, Hackett, Indianapolis, IN, 1983.
[50] J.-S. Gutmann, W. Hatzack, I. Herrmann, B. Nebel, F. Rittinger, A. Topor, T. Weigel, The CS Freiburg Team: Playing robotic soccer based on an explicit world model, AI Magazine 21 (1) (2000) 37–46.
[51] I. Hacking, Representing and Intervening, Cambridge University Press, Cambridge, 1983.
[52] S. Harnad, The symbol grounding problem, Phys. D 42 (4) (1990) 335–346.
[53] S. Harnad, Symbol grounding problem, in: L. Nadel, D. Chalmers, P. Culicover, B. French, R. Goldstone (Eds.), Encyclopedia of Cognitive Science, Macmillan Publishers, London, 2002.
[54] P.J. Hayes, What the frame problem is and isn't, in: Z.W. Pylyshyn (Ed.), The Robot's Dilemma: The Frame Problem in Artificial Intelligence, Ablex Publishing, Norwood, NJ, 1987, pp. 123–137.
[55] N.K. Hayles, How We Became Posthuman: Virtual Bodies in Cybernetics, Literature, and Informatics, University of Chicago, Chicago, IL, 1999.
[56] M. Heidegger, Being and Time, Harper Collins, New York, 1962.
[57] Held, Hein, Adaptation of disarranged hand-eye coordination contingent upon re-afferent stimulation, Perceptial Motor Skills 8 (1958) 87–90.
[58] H. Hendriks-Jansen, Catching Ourselves in the Act: Situated Activity, Interactive Emergence, Evolution and Human Thought, MIT Press, Cambridge, MA, 1996.
[59] D. Hilditch, At the heart of the world: Merleau–Ponty and the existential phenomenology of embodied and embedded intelligence in everyday coping, PhD Thesis, Department of Philosophy, Washington University, St. Louis, MO, 1995.
[60] T. Horgan, J. Tienson, Rules, in: W. Bechtel, G. Graham (Eds.), A Companion to Cognitive Science, Basil Blackwell, Oxford, 1999, pp. 660–670.
[61] M.N. Hums, M.P. Singh, Cognitive agents, Internet Comput. (November/December 1998) 87–89.
[62] M. Jeannerod, The Cognitive Neuroscience of Action, Blackwell, Oxford, 1997.
[63] W.J. Kargo, S.F. Giszter, Rapid correction of aimed movements by summation of force-field primitives, J. Neurosci. 21 (1) (2000) 409–426.
[64] M.P. Kaschack, A.M. Glenberg, The role of affordances and grammatical constructions in sentence comprehension, J. Memory and Language 43 (2000) 508–529.
[65] J.A.S. Kelso, Dynamic Patterns: The Self-Organization of Brain and Behavior, MIT Press, Cambridge, MA, 1995.
[66] J. Kennedy, Swarm Intelligence, Morgan Kaufmann, New York, 2001.
[67] D. Kirsh, P. Maglio, On distinguishing epistemic from pragmatic action, Cognitive Sci. 18 (1995) 513–549.
[68] D. Kirsh, Foundations of AI: The big issues, Artificial Intelligence 47 (3) (1991) 3–30.
[69] D. Kirsh, Today the earwig, tomorrow man, Artificial Intelligence 47 (3) (1991) 161–184.
[70] T.S. Kuhn, The Structure of Scientific Revolutions, 2nd edn., University of Chicago, Chicago, IL, 1970.

[71] G. Lakoff, M. Johnson, Metaphors We Live By, University of Chicago, Chicago, IL, 1980.
[72] G. Lakoff, M. Johnson, Philosophy in the Flesh: The Embodied Mind and Its Challenge to Western Thought, Basic Books, New York, 1999.
[73] G. Lakoff, R.E. Núnez, Where Mathematics Comes From: How the Embodied Mind Brings Mathematics into Being, Basic Books, New York, 2000.
[74] C. Langton, Artificial Life: An Overview, MIT Press, Cambridge, MA, 1995.
[75] J. Lave, Cognition in Practice, Cambridge University Press, Cambridge, UK, 1988.
[76] D.B. Lenat, From 2001 to 2001: Common sense and the mind of HAL, in: D.G. Stork (Ed.), Hal's Legacy: 2001 as Dream and Reality, MIT Press, Cambridge, MA, 1998, pp. 193–208.
[77] H. Levesque, F. Pirri, Logical Foundations for Cognitive Agents, Springer-Verlag, New York, 1999.
[78] J. McCarthy, P.J. Hayes, Some philosophical problems from the standpoint of artificial intelligence, in: B. Meltzer, D. Michie (Eds.), Machine Intelligence 4, Edinburgh University Press, Edinburgh, 1969, pp. 463–502.
[79] R. McClamrock, Existential Cognition, University of Chicago, Chicago, IL, 1995.
[80] D. McDermott, We've been framed: Or, why AI is innocent of the frame problem, in: Z.W. Pylyshyn (Ed.), The Robot's Dilemma: The Frame Problem in Artificial Intelligence, Ablex Publishing, Norwood, NJ, 1987, pp. 113–122.
[81] M. Merleau-Ponty, The Structure of Behavior, Duquesne University Press, Philadelphia, PA, 1989.
[82] M. Merleau-Ponty, The Phenomenology of Perception, Routledge, New York, 2002.
[83] M. Minsky, The Society of Mind, Simon and Schuster, New York, 1986.
[84] H.P. Moravec, The Stanford Cart and the CMU Rover, Proc. IEEE (1983) 872–884.
[85] A. Newell, Physical symbol systems, Cognitive Sci. 4 (1980) 135–183.
[86] A. Newell, H.A. Simon, Human Problem Solving, Prentice Hall, Englewood Cliffs, NJ, 1972.
[87] N.J. Nilsson, Shakey the robot, Technical Report 323, Stanford Research Institute AI Center, 1984.
[88] S. Nolfo, D. Floreano, Evolutionary Robotics, MIT Press, Cambridge, MA, 2000.
[89] T. Oates, Grounding knowledge in sensors: Unsupervised learning for language and planning, PhD Thesis, University of Massachusetts, Amherst, MA, 2001.
[90] T. Oates, M.L. Anderson, Learning to reason with grounded symbols, in preparation.
[91] M. O'Donovan-Anderson, The Incorporated Self: Interdisciplinary Perspectives on Embodiment, Rowman and Littlefield, Lanham, MD, 1996.
[92] M. O'Donovan-Anderson, Content and Comportment: On Embodiment and the Epistemic Availability of the World, Rowman and Littlefield, Lanham, MD, 1997.
[93] M. O'Donovan-Anderson, Review of philosophy in the flesh, Rev. Metaphys. (June 2000).
[94] J.K. O'Regan, A. Noë, A sensorimotor account of vision and visual consciousness, Behavioral Brain Sci. 24 (5) (2001).
[95] K.M. Ford, P.J. Hayes, N. Agnew, On babies and bathwater: A cautionary tale, AI Magazine 15 (4) (1994) 14–26.
[96] D. Perlis, Consciousness as self-function, J. Consciousness Stud. 4 (1997) 509–525.
[97] A. Pessin, S. Goldberg, 'The Twin Earth Chronicles: Twenty Years of Reflection on Hilary Putnam's 'The Meaning of Meaning', M.E. Sharpe, New York, 1996.
[98] J. Petitot, F.J. Varela, B. Pachoud, J.-M. Roy, Naturalizing Phenomenology, Stanford University Press, Palo Alto, CA, 2000.
[99] R. Pfeifer, C. Scheier, Understanding Intelligence, MIT Press, Cambridge, MA, 1999.
[100] R. Picard, Affective Computing, MIT Press, Cambridge, MA, 1998.
[101] D. Pitt, Mental representation, in: E.N. Zalta (Ed.), Stanford Encyclopedia of Philosophy, Fall, 2000, http://plato.stanford.edu/archives/fall2000/entries/mental-representation/.
[102] R. Port, T. van Gelder, Mind as Motion: Explorations in the Dynamics of Cognition, MIT Press, Cambridge, MA, 1995.
[103] H. Putnam, The meaning of meaning, in: H. Putnam (Ed.), Mind, Language, and Reality, Cambridge University Press, London, 1975.
[104] Z.W. Pylyshyn, The frame problem blues: Once more, with feeling, in: K.M. Ford, Z.W. Pylyshyn (Eds.), The Robot's Dilemma Revisited: The Frame Problem in Artificial Intelligence, Ablex Publishing, Norwood, NJ, 1996, pp. xi–xviii.

S0004-3702(03)00054-7/REV AID:1984 Vol.●●●(●●●) aij1984 P.40 (1-40)
ELSGMLTM(ARTINT):m1a v 1.155 Prn:29/05/2003; 10:10 by:ML p. 40

[105] B. Rogoff, J. Laveit, Everyday Cognition: Development in Social Context, Harvard University Press, Cambridge, MA, 1999.
[106] G. Rosenberg, M.L. Anderson, Content and action, in preparation.
[107] C. Sample, Living words: Physiognomy and aesthetic language, in: M. O'Donovan-Anderson (Ed.), The Incorporated Self: Interdisciplinary Perspectives on Embodiment, Rowman and Littlefield, 1996, pp. 113–126.
[108] M. Sheets-Johnstone, The Roots of Thinking, Temple University Press, Philadelphia, PA, 1990.
[109] H.A. Simon, The Sciences of the Artificial, MIT Press, Cambridge, MA, 1969.
[110] B.C. Smith, The owl and the electric encyclopedia, Artificial Intelligence 47 (1–3) (1991) 251–288.
[111] B. Smith, On the Origin of Objects, MIT Press, Cambridge, MA, 1996.
[112] L. Steels, The artificial life roots of artificial intelligence, Artificial Life 1 (1994) 75–110.
[113] L. Steels, R. Brooks, The Artificial Life Route to Artificial Intelligence: Building Embodied, Situated Agents, Lawrence Erlbaum, Hillsdale, NJ, 1995.
[114] C. Taylor, Overcoming epistemology, in: Baynes, et al. (Eds.), After Philosophy: End or Transformation?, MIT Press, Cambridge, MA, 1987.
[115] M. Thielscher, The qualification problem: A solution to the problem of anomalous models, Artificial Intelligence 131 (1–2) (2001) 1–37.
[116] E. Thompson, Colour Vision: A Study in Cognitive Science and the Philosophy of Perception, Routledge, London, 1995.
[117] S. Todes, Body and World, MIT Press, Cambridge, MA, 2001.
[118] M. Tomasello, The Cultural Origins of Human Cognition, Harvard University Press, New York, 2000.
[119] M. Turner, The Literary Mind, Oxford University Press, New York, 1995.
[120] M. Turner, The Cognitive Dimensions of Social Science, Oxford University Press, New York, 2001.
[121] F. Varela, E. Thompson, E. Rosch, The Embodied Mind, MIT Press, Cambridge, MA, 1991.
[122] F.J. Varela, Ethical Know How: Action, Wisdom, and Cognition, Stanford University Press, Stanford, CA, 1999.
[123] A.H. Vera, H.A. Simon, Situated action: A symbolic interpretation, Cognitive Sci. 17 (1993) 7–48.
[124] A. von Gontard, K. Zerres, M. Backes, C. Laufersweiler-Plass, C. Wendland, P. Melchers, G. Lehmkuhl, S. Rudnik-Schoneborn, Intelligence and cognitive function in children and adolescents with spinal muscular atrophy, Neuromuscular Disorders 12 (2002) 130–136.
[125] F.P. Wilson, The Hand, Random House, New York, 1998.
[126] M. Wrathall, J. Malpas (Eds.), Heidgegger, Coping and Cognitive Science, MIT Press, Cambridge, MA, 2000.