S0004-3702(03)00055-9/SSU AID:1985 Vol.●●●(●●●)
ELSGMLTM(ARTINT):m1a v 1.139 Prn:2/04/2003; 12:56
aij1985
P.1 (1-20)
by:ML p. 1

ELSEVIER

Available at
www.ComputerScienceWeb.com
POWERED BY SCIENCE DIRECT®

Artificial Intelligence ●●● (●●●●) ●●●–●●●

Artificial Intelligence

www.elsevier.com/locate/artint

Embodied artificial intelligence

Ron Chrisley

School of Computer Science, University of Birmingham, Birmingham, UK

1. Introduction

Mike Anderson[1] has given us a thoughtful and useful field guide: Not in the genre of a bird-watcher's guide which is carried *in* the field and which contains detailed descriptions of possible sightings, but in the sense of a guide *to* a field (in this case embodied cognition) which aims to identify that field's general principles and properties. I'd like to make some comments that will hopefully complement Anderson's work, highlighting points of agreement and disagreement between his view of the field and my own, and acting as a devil's advocate in places where further discussion seems to be required.

Given the venue for this guide, we can safely restrict the discussion to embodied *artificial intelligence* (EAI), even if such work draws on notions of embodied *cognition* from the fields of philosophy, psychology and linguistics. In particular, I'll restrict my discussion to the impact that embodiment can have on the task of creating artificial intelligent agents, either as technological ends in themselves, or as means to understanding natural intelligent systems, or both.

2. If I said your robot had a body would you hold it against me?

Before moving on to the details of the variety of approaches to embodied cognition, tracing its historical and philosophical roots, or evaluating the arguments for and against it, Anderson rightly decides to consider exactly what is meant by embodiment in the context of AI. Much of this is done in a negative manner, by identifying the crucial planks in symbolic or good old-fashioned artificial intelligence (GOFAI, [26]), and pointing out how an embodied approach differs. This kind of characterisation of EAI is, to a large extent, justified, since much of the work in EAI has identified itself in opposition to a GOFAI target, real or imagined. But it does have the standard problem of definition-in-opposition:

E-mail address: rlc@cs.bham.ac.uk (R. Chrisley).
URL address: http://www.cs.bham.ac.uk/~rlc (R. Chrisley).

[1] "Embodied Cognition: A field guide", this issue.

0004-3702/03/$ – see front matter
doi:10.1016/S0004-3702(03)00055-9

S0004-3702(03)00055-9/SSU AID:1985 Vol.●●●(●●●)
ELSGMLTM(ARTINT):m1a v 1.139 Prn:2/04/2003; 12:56
aij1985
P.2(1-20)
by:ML p. 2

spending valuable space on telling us a lot about what is probably already familiar (symbols, knowledge bases, the sense-model-plan-act cycle, modeling the world, context-free representations, cognitivism, etc.), space which might have been better spent going into more detail on the novel concepts that are definitive of EAI: grounding, selectivity, situatedness, activity, and the notion of the body itself.

For example, different approaches to EAI can be distinguished by the notion of "body" they employ. [14] makes a distinction between four ideas of what it takes to be embodied, from the greater to the less inclusive:

- Physical realisation: The system must merely be realised in some physical substrate or other);
- Physical embodiment: The system must be realised in a coherent, integral physical structure);
- Organismoid embodiment: The physical realisation of the system must share some (possibly superficial) characteristics with the bodies of natural organisms, but need not be alive in any sense; and
- Organismal embodiment: The physical realisation of the system must not only be organism-like, but actually organic and alive.

Further distinctions can be made on the basis of the way that a particular notion of embodiment is employed: Is it claimed that embodiment of a particular kind is necessary for AI? Is it merely necessary for some sensorimotor capacities, or is it required for intelligence in general? Or is it an approach that is merely *preferred*, not required, for the explanatory or technological purposes of a given research group? A general survey of these positions within EAI would be useful.

The point of the cheap gag in the title of this section? One can't begin to figure out how to react to claims that one's AI system is or is not embodied until it is made clear what notion of "body" is being employed by those claims.

3. The trouble with GOFAI and the embodied solution

Perhaps the reason why Anderson spends little or no time discussing different notions of the body[2] is that having a body is not, despite the name of the field, central to the EAI work he reviews. To see why, let's look at how Anderson characterises the field of EAI in opposition to GOFAI. To do so, Anderson must first characterise GOFAI, which he does in the following terms:

(1) An emphasis on explicit, sentential representation.

[2] In his Section 3 ("Embodiment and Grounding") Anderson does discuss four *aspects* of our bodies (physiology, evolutionary history, [involvement in] practical activity, and socio-cultural situatedness), but that discussion assumes an implicit notion of, rather than discusses, what a body actually *is*.

S0004-3702(03)00055-9/SSU AID:1985 Vol.●●●(●●●)
ELSGMLTM(ARTINT):m1a v 1.139 Prn:2/04/2003; 12:56
aij1985
P.3 (1-20)
by:ML p. 3

(2) An emphasis on operations on representations (rules), which are also explicit and sentential, and which operate on representations by virtue of the latter's form or syntax, not their meaning or semantics.
(3) A sense-model-plan-act (SMPA) cycle.
(4) An emphasis on high-level competences such as thought, reason, planning and problem-solving.
(5) A lack of concern with how these higher competences arise out of more primitive competences.

Although Anderson gives a tentative definition of GOFAI in terms of Cartesianism and Cognitivism, it is better, as we shall see, to see GOFAI as involving all of the above strands that Anderson mentions (e.g., the "C + C" definition leaves mysterious the relevance of an attack on SMPA to an analysis of GOFAI).

A disadvantage of characterising EAI in opposition to GOFAI makes itself manifest here. This kind of list, necessary for such an approach, polarises most readers into an unproductive standoff. Some see such a list and think: "Yes, that sums up non-embodied AI, and we know what problems such approaches have.[3] So EAI is the only way forward". Others see the list as think "Hey, that's straw-man AI! There's lots of AI work that no one would call 'embodied' which rejects some, most or all of those points.[4] So EAI is not needed; 'non-embodied' AI is already addressing the limitations of the extreme approach presented above".

I think both of these reactions to Anderson's list are based on truths, but their conclusions should be avoided. True, much non-embodied AI work does not fall under the above list very well, so EAI is not the only way to overcome the problems of the approach defined by that list. But it is also true that EAI should not be dismissed just because some of its proponents engage in overzealous rhetoric and unjustifiably claim it is the only way of dealing with those problems. Yes, there might be a way of overcoming the problems of GOFAI within a non-embodied approach. But it might be that there is a way of overcoming the problems of GOFAI by taking an embodied approach. Let a thousand flowers bloom.

Some will think this too conciliatory. They will note that in some situations we can use our powers of reasoning to see in advance that a particular research strategy is doomed, and should thus be abandoned; or promising, and should thus be encouraged. However, I don't think we are in this kind of situation with respect to either embodied or non-embodied AI, at least not on any simple reading of "embodiment".

[3] E.g., the problems of dynamics and relevance, discussed below.

[4] For instance, connectionist AI rejects most of the items on the list; procedural AI would reject 1 and 2; Many reactive and teleo-reactive systems reject 1, 2 and 3; much or most work on AI vision and robotics rejects 2 and 4 (for an example of a GOFAI robot (Freddy) which rejects 3, see [3]); and yet none of these approaches need be particularly more embodied than CYC [34] is, in any contentious sense of the word "embodied".

S0004-3702(03)00055-9/SSU AID:1985 Vol.●●●(●●●)
ELSGMLTM(ARTINT):m1a v 1.139 Prn:2/04/2003; 12:56
aij1985
P.4(1-20)
by:ML p. 4

To support this point,[5] I'll look at two architectural features that are purportedly problems with GOFAI which demand an embodied approach: Formality, and the sense-model-plan-act framework.

3.1. Mere formality

Anderson is right: EAI proponents often criticise non-embodied AI for taking the view[6] that "just as is the case in modern logic, it is the *form* of the symbol. . . and not its meaning that is the basis of its rule-based transformation" (p. 3, original emphasis[7]).

But it's not clear that EAI proponents can use this as a stick with which to beat GOFAI. Brian Cantwell Smith's analysis of what "formal symbol manipulation" might mean concludes that on any reading of that term that actually holds true of the systems that we have taken to paradigmatically fall under it, "formal" cannot mean "manipulation independent of semantics":

> Far from lining up on top of each other, real-world computer systems' physical (i.e., syntactic—R.C.) and semantic boundaries *cross-cut*, in rich and productive interplay. It is not just that computers are involved in an engaged, participatory way with *external* subject matters, in other words, as suggested by some recent "situated" theorists. They are participatorily engaged in the world *as a whole*—in a world that indiscriminately includes themselves, their own internal states and processes [44, p. 37].

He goes on to say that this interdependence of the syntactic and the semantic is "not only architecturally essential, but is also critical, when the time comes, in establishing and grounding a system's intentional capacities". What's important here is that Smith is speaking of real-world computation in general; inasmuch as arch-symbolic AI systems have intentional capacities, inasmuch as they can be understood as having semantics or even syntax (for the latter surely requires the former), inasmuch as they are computers at all; they are *ipsis factis* not "formal", if "formal" implies an independence of syntax and semantics. So if processing symbols in a way which depends on their semantics as well as their syntax is a hallmark of embodiment, then all real-world computing, including most if not all AI systems, are already embodied, in a very non-trivial sense.

If "formal symbol manipulation" doesn't mean "manipulation of symbols independent of semantics", then what *does* it mean? I think the best way to answer that question is to look at the historical context. When behaviourism held sway, anyone adverting to meaningful internal states in their explanation of intelligent behaviour was suspected of an

[5] Actually, I just made four points: (1) EAI hasn't been shown to be the best approach; (2) EAI hasn't been shown to be doomed; (3) non-EAI hasn't been shown to be the best approach; (4) non-EAI hasn't been shown to be doomed. What follows supports point (4) and, to a lesser extent, point (1).

[6] Incidentally, I'm not sure (*pace* [16]) that "Cognitivism" is a good name for this view; surely a Cognitivist is someone who emphasises the role of knowledge in intelligence and mentality, and this is quite distinct from a person (a Formalist? Representationalist? Sententialist?) who takes "representation, formalism, and rule-based transformation" (p. 3) to be central.

[7] In this paper, *all* verbatim quotations are reproduced with their original emphasis, except where indicated otherwise.

S0004-3702(03)00055-9/SSU AID:1985 Vol.●●●(●●●)
ELSGMLTM(ARTINT):m1a v 1.139 Prn:2/04/2003; 12:56
aij1985
P.5(1-20)
by:ML p. 5

implicit dualism, of relying on some kind of ghost in the machine. The computationalist insight was that a machine could be a semantic engine by virtue of being a syntactic engine, and thus without violating any principles of naturalism. The point of modifying "symbol manipulation" with "formal" was to indicate that theirs was not a ghostly, dualistic, question-begging, homuncular kind of representation processing, but rather a kind of manipulation that was naturalistically admissible because it could be characterised entirely in physical, syntactic terms.[8] The point wasn't that semantic properties played no role in processing; it was rather that inasmuch as they did, they did so by virtue of properties that were *not* naturalistically problematic, such as syntactic properties, playing a role.[9]

Seen this way, it is a mistake for EAI to oppose itself to formalism *per se*, since that is tantamount to opposing itself to naturalism, tantamount to saying that there is something in addition to physical properties which plays a causal role in the production of behaviour. That would violate the causal closure of physics. No, it is not the idea that one can have semantic processes by virtue of having processes which operate on representations according to their form which EAI should be taken to be contesting. What is contested, rather, is GOFAI's position on what that form should be; the opponent is a (particular) *formalism*, not a (general) *formality*. Thus, this EAI critique of GOFAI gets us only as far as the connectionist critique of GOFAI did. Except at least the connectionists had a concrete proposal for what should supplant sententialism; the EAI alternative is not so clear. What *is* clear is that opposing sententialism is neither necessary not sufficient for embodiment, in any sense.

3.2. Shoot first, prove propositions later (if at all)

The bulk of Anderson's discussion of the problems of GOFAI which EAI aims to redress focuses on the Sense-Model-Plan-Act (SMPA) framework. Anderson's summary of the criticism is that SMPA is "too expensive, and therefore biologically implausible[10]" (p. 7) because of two problems:

- Problem of dynamics: The world can change after the model is constructed yet before the plan based on that model is executed, possibly requiring a change of plan.

[8] On this reading, "formal" is more general than, say "effective", in that a formal operation need only be physically (non-oracularly) possible; it need not, e.g., consist of a finite number of humanly-executable steps. I should point out that I am not offering a comprehensive definition of "formal"; no doubt it means different things in different contexts. I only seek to clarify what it means in the context of the phrase "formal symbol manipulation".

[9] True, some who embraced the computationalist approach misunderstood these niceties. Some thought, e.g., that since there was a syntactic story about how a behaviour was produced there could not also be a semantic story about how that behaviour (or an action which it realised) was produced. Others thought that the directly naturalisable properties by virtue of which a system also had semantic properties had to be internal, syntactic ones; they ignored the role that, e.g., external, relational properties can play in fixing semantic properties. Nevertheless, I believe the history I have presented illuminates the role that "formal symbol manipulation" and its cognate concepts were playing in the theoretical dialectic of the 50s, 60s and early 70s.

[10] For a discussion of (the possible irrelevance of) biological plausibility, see Section 6.

S0004-3702(03)00055-9/SSU AID:1985 Vol.●●●(●●●)
ELSGMLTM(ARTINT):m1a v 1.139 Prn:2/04/2003; 12:56
aij1985
P.6(1-20)
by:ML p. 6

- Problem of relevance: One could perhaps solve the problem of dynamics if one restricted oneself to reasoning about change only when the change is "likely to affect the achievability of the goal"; but how is this restriction to be achieved? (p. 8).

The problem of dynamics can be addressed within the SMPA framework, Anderson notes, by either (a) modeling the dynamics, and planning "in terms of the expected changes in the environment", or (b) having contingency plans.

The critique of (a) which Anderson presents is: "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" (pp. 7–8). This is indeed a problem, but it doesn't feel like a paradigm-buster, and it certainly isn't the beginning of a regress, as Anderson[11] seems to suggest. A regress would threaten if Anderson had said something like "for the way that the world changes may itself have changed after the time at which we last modeled the world's dynamics yet before the time at which our plan is executed". But he didn't say that for good reason: we don't believe (*pace* [42]!) that the world's dynamics is itself dynamic.

So in effect, the criticism of strategy (a) is only: SMPA may get the dynamics wrong. But this isn't in itself devastating;[12] it only becomes so if it is promoted to something like "SMPA will *always* get the dynamics wrong", or, equivalently, "SMPA can *never* get the dynamics right". Why might we think that is the case?

Complexity. In a real-world situation, there are just too many variables to be taken into account, and their interaction yields a combinatorial explosion. This is also why (b) won't work either: It would require too many contingency plans. "But", common sense interjects, "most of the variability of the world is irrelevant to most tasks; couldn't (a) and (b) have a chance if they restricted themselves only to the variables and contingencies which might affect the achieveability of the goal?" Perhaps, but now the problem of relevance appears: How is such a restriction to be accomplished? Here it does seem that a thorough-going SMPA approach encounters a regress. Control of inference in general, and determination of relevance in particular, can be considered to be species of "internal action". Therefore, the problems of dynamics and relevance that arise for SMPA-generated action as a whole also arise for the action of "relevance-determining" which was meant to solve said problems.[13]

This is a familiar situation: An analysis of some aspect A of rationality is offered, and its prerequisites identified. It is then argued that A can only be rational if its prerequisites are rationally attained. And yet this, circularly, requires the application of A itself, which requires the prerequisites for A, *ad infinitum*. We see this kind of paradox in "What the Tortoise Said to Achilles" [9] (A = deduction), Hume's [30] and Goodman's [22] riddles (A = induction), Wittgenstein's [48] private language argument(s) (A = rule-following), and debates on the Language of Thought [18] [21] (A = rule-following again).

[11] And perhaps T.S. Eliot: "In a minute there is time\\For decisions and revisions which a minute will reverse", *The Love Song of J. Alfred Prufrock*.

[12] In fact, it would be a devastating blow against SMPA theory as a model of human cognition if such theory implied that SMPA systems were never wrong, since clearly we humans often are!

[13] Note that it isn't the details of SMPA (e.g., the relative order of the constituent steps) that is causing the problem here, but rather the deliberative nature of SMPA.

S0004-3702(03)00055-9/SSU AID:1985 Vol.●●●(●●●)
ELSGMLTM(ARTINT):m1a v 1.139 Prn:2/04/2003; 12:56
aij1985
P.7 (1-20)
by:ML p. 7

There seem to be four ways out of this conundrum:

(1) The application of *A* to the prerequisites of *A* is a different task than the applications of *A* to external actions, and therefore might not cause a regress. In the case of planning, for example, it is possible (though perhaps not plausible) that the problems of dynamics and relevance do not arise for the action of determining relevance itself, or if they do, not in an intractable way.
(2) The prerequisites of *A* may be justified by a rational procedure distinct from *A*. In the case of planning, for example, one might try to *deduce* or *induce* the relevancy relations directly, rather than planning "relevance-determining" actions, thus providing a rational basis for planning without regress. This solution will not work, however, if, as seems likely, *all* modes of rationality have a regress problem.
(3) One could give up on foundationalism with respect to rationality, and instead hold that *A* can be rational even if it is based on non-rational (unexamined) assumptions. In the case of planning, a system might just have some arbitrary, unjustified, unreflected ways of determining relevance; the rationality of the system is rationality relative to those pre-given, fixed parameters.[14] On this view, rationality is a matter of how one deals with what one is given.[15]
(4) An extreme, final option is to give up on rationality altogether: Accept that *A* is not strictly rational, yet deny that such strict rationality is a requirement for a working AI. This option will be attractive to those AI workers whose primary interest is in building a working system, rather than in adherence to a rationalist ideology. In the case of planning, what would justify the SMPA framework is not some (were it possible) *a priori* establishment of its rationality, but an *a posteriori* establishment of its utility.

So here we have four ways in which the SMPA framework might be defended, and an embodied approach made unnecessary. If pressed to speculate as to which of these four possible solutions are the most promising, I'd have to plump for 3 and 4; 1 and 2 seem implausibly optimistic given the history of attempts to formalise reason. Yet there is an interesting sense in which 3 and 4, as patches to the SMPA framework, themselves require a kind of embodiment. A distinction can be made between those aspects of ourselves which can be understood in rational and conceptual terms, and those aspects which cannot. For example, my reasons for coming to work today, and my belief that today is Monday, are rational/conceptual aspects of myself; my mass, volume, heart rate, and the positions of my limbs are not. The thorough-going SMPA fantasy is that competent real-world action can be achieved by a system solely by virtue of its conceptual aspects. But if 3 or 4 is the correct response to the problems of dynamics and relevance, a thorough-going SMPA framework is not possible: the SMPA strategy, in order to work in the real world, *must* be based on aspects of the system for which no rational/conceptual description, analysis, or

[14] And is thus comparable to the notions of bounded or minimal rationality put forward by Simon [43] and Cherniak [10], *inter alia*.

[15] Thus, robots could do worse than pray the Serenity Prayer (which apparently dates back to Boethius), suitably modified: "Maker, grant me rationality to change the things I can change, embodiment to effect the things I cannot, and wisdom to know the difference".

S0004-3702(03)00055-9/SSU AID:1985 Vol.●●●(●●●)
ELSGMLTM(ARTINT):m1a v 1.139 Prn:2/04/2003; 12:56
aij1985
P.8(1-20)
by:ML p. 8

justification is available. If a robot is going to be able to act in the real world, it will only partly (if at all; cf. solution 4) be because of the rational virtues of its SMPA strategies; it will also necessarily be a matter of *just being built the right way*.[16] That is, it will partly be a matter of having an unreflective (and therefore non-rational) disposition to, e.g., take certain properties and not others to be relevant to particular actions. If we reconstruct the claims of the embodiment theorists to be claims about the limitations of the *purely* conceptual, rational subject; if we take "embodiment" to mean those aspects and abilities of a system which cannot be analysed in purely rational terms, but rather must be understood causally or according to some norm other than individual rationality; if (but only if) we take "embodiment" to mean the aspects of a system which are not just cognitively impenetrable [40], but which are not the outcomes of processes which can be construed as rational; if these changes to our thinking are made, then we will have reason to believe that real world intelligence must (in this sense) be embodied.

Note, however, that this falls a long way short of establishing the more extreme claims of EAI proponents. In particular, it allows for the possibility of artificial intelligence being grounded in a sub-rational or sub-conceptual substrate in a manner quite different from the way that natural, organic intelligence is. Of course, AI can benefit from an understanding of how the body accomplishes this grounding in the natural case. But a slavish copying of nature may be unnecessary and in some cases unhelpful (as it was in the achievement of artificial flight [5,8,49]; see Section 6).

So the only target which most EAI-talk has any chance of hitting is the purely rational, thorough-going SMPA system. But the notion of a "purely" rational AI system is a distraction. It's questionable whether any implemented AI system has ever been purely rational, or whether such would even be possible.[17] A physical system in which every physical difference makes an intentional difference may be *logically* possible, but it beggars the imagination, and surely has little to do with the best course for AI research. This recalls, and perhaps explains, the infamous remark from Drew McDermott [36] which Anderson quotes: "*no working AI program has ever been bothered at all by the frame problem*" (p. 25). However, I would replace "at all" with "insurmountably". Thorough-going SMPA is just as perverse as thorough-going deduction (cf. the Tortoise and achilles again).[18]

[16] Of course, having an efficient and effective planner is also a matter of "being built the right way", but embodiment is *only* that; there is not, in addition, a rational characterisation/justification, as there may be for a planner's decisions.

[17] The system would have to be such that all physical processes and states would, at all times, instantiate some rational process or state. Mere hair growth would be out, as would mere mass increase/decrease, mere battery voltage drop, or two molecules merely exchanging positions. These events could only be allowed in a "purely" rational system if something conceptually or rationally normative supervened on them.

[18] Thus, non-thorough-going SMPA is, in some sense, more rational than thorough-going SMPA—at least you're not *guaranteed* to get eaten by the tiger. But that is not the notion of rationality I have been employing in this discussion. A process can be *externally rational* (i.e., is pragmatically useful, gives the right results for survival) even though it is not *internally rational* (based on a procedure which can be understood to further the agent's desires int he light of its beliefs). The implicit claim which I am rejecting is that SMPA must strive to be thorough-goingly internally rational.

S0004-3702(03)00055-9/SSU AID:1985 Vol.●●●(●●●)
ELSGMLTM(ARTINT):m1a v 1.139 Prn:2/04/2003; 12:56
aij1985
P.9 (1-20)
by:ML p. 9

Having a body in any robust sense, then, is not a requirement for dealing with the problems of SMPA and pure rationality. But even if there is no knock-down argument in favour of EAI, any different, novel way of responding to the planning challenge is worth considering. What does EAI have to offer?

The EAI response is best identified by giving examples. Don't use complex geometric reasoning to calculate where the centre of the corridor is—by the time you finish, your position will probably have changed, if you are trying to act in the world in a continuous, real-time way. Instead, merely apply differentially more torque to your right wheel if the dark spot in front of you is to the left of your visual field, and the converse if the spot is on the right [29]. Or, if you're a cricket: Don't create or learn complex pattern detection algorithms and use spatial planners/deliberators to figure out how to move toward your mate—your mate will have moved by the time you finish all that, and the problem of dynamics sets in. Instead, merely have an ear which is built so that when your mate's calls are heard, the shape of the ear results in the proper signals being sent to the legs to result in mate-directed movement [47].

These do seem like fresh alternatives to the GOFAI approach. And despite the well-known problem of how such systems could scale up to handle higher-level cognitive tasks (Anderson gives some good references for both sides of that debate), it certainly seems worthwhile for at least some AI researchers to investigate empirically what such systems can do. But it is a misleading misnomer to take their characteristics of planless reaction, constant attention and selective representation to be central to "embodied" AI. Having a body is neither necessary (e.g., consider Agre and Chapman's Pengi [2]) nor sufficient (unless one trivialises the issue by defining "body" so that every GOFAI robot has one) for having those three features.

So it turns out that characterising EAI in opposition to formality, or in opposition to the SMPA framework, isn't that helpful. Where else can we turn?

4. You're grounded!

One of Anderson's important points is this: Although the GOFAI virtues of deliberation and reasoning can and should be retained, EAI's contribution is to point out the importance of *grounding*, even (or especially) for such high-level operations as planning, reasoning, deliberating and symbol processing. I'm inclined to agree with this statement, with the proviso that I might disagree with Anderson and/or many EAI proponents on what grounding amounts to; more on that later. First: What notion of grounding is EAI offering, according to Anderson?

Neither Anderson nor [24], which introduced the symbol grounding problem (but to whom Anderson does not refer), give an explicit definition, but the meaning of "grounding" seems to be captured by something like:

> A symbol is grounded if it has its meaning or content by virtue of its causal properties and relations to the referent or subject matter of the symbol; or is appropriately causally related to (defined in terms of) grounded symbols.

S0004-3702(03)00055-9/SSU AID:1985 Vol.●●●(●●●)
ELSGMLTM(ARTINT):m1a v 1.139 Prn:2/04/2003; 12:56
aij1985
P.10 (1-20)
by:ML p. 10

Thus, a symbol is ungrounded if it is not ultimately completely definable in terms of symbols which have their content by virtue of their causal relations with their subject matter. One way for this to occur is for the meaning of the symbol to be merely a matter of some (other?) agent ascribing the symbol that meaning. Anderson gives an example for the symbol "chair" in an agent: it is grounded only if its use by that agent is governed by an ability to reliably detect (and recognise?) chairs, and if it gives rise to appropriate chair behaviour.

This kind of pragmatist theory of meaning has familiar problems. For example, it seems to imply that if I act inappropriately toward something, I *ipso facto* cannot be thinking about that something, since appropriate behaviour toward it is a requirement for representing it. Consider Anderson's example of a non-monarch sitting on a king's throne; Anderson concludes that such a person must fail to grasp either the concept "chair" or "throne". But if we generalise this point, it looks as if we could never be wrong, say, in mathematics. If I say "$2 + 2 = 5$", then on the pragmatist line, I must not grasp one or more of the concepts normally involved in that expression. But if so, then I haven't said something false about 2, addition and 5; rather, I have either said something true about something else, or I have said nothing at all. Neither of these consequences appeal.

In any case, the strongest form of the EAI grounding argument goes something like: People and EAI use grounded symbols; GOFAI does not. Therefore, EAI, and not GOFAI, _____ (fill in the blank as required: "can provide good models human cognition", "is possible", "can produce true understanding", etc.).[19]

It might be that grounding is required for cognition, understanding, intelligence, etc. but the above argument form is not a good one.[20] In any event, I'm not as convinced as Anderson seems to be that symbol grounding is at the heart of EAI. EAI is much more radical: It is not a thesis about how to ground symbols, but the idea that symbols, grounded or not, play only a small role in intelligence as a whole, and that there are many aspects of mentality in which they may play no role whatsoever. When Thelen and Smith [45] tell us that changes in the mass of the infant's leg play a crucial role in the acquisition of the walking skill, when Webb [47] uses a robot to show us how the shape of the cricket ear allows them to achieve phonotaxis, when Beer [6] analyses the robust gait of his six-legged robots using dynamical systems theory, when Breazeal and Scassellati [7] show us how putting eyebrows on Kismet ensures the proper carer-robot dynamic necessary for Kismet to learn to track objects visually; when these researchers make their contributions, they are not doing so primarily, if at all, by way of showing how symbols are grounded. The issue just doesn't come up. Symbol grounding may be an important issue, even to some EAI people, but it just doesn't seem to hit the nail on

[19] Anderson seems to allow that perhaps not *all* of my representations are grounded in this way, but may be grounded by other humans. E.g., perhaps many of my symbols denoting technical concepts get their meanings from the content assigned to them by the experts to which I defer. But the claim of the proponent of embodiment would be that at least *some* of my symbols are grounded in my own causal relations to things in the world (other than experts), and the symbols that do have deferential content must be so grounded for the experts (either individually, or as a group) to which I defer.

[20] At least not without some kind of assistance; perhaps that could be provided by the points Dreyfus makes concerning intelligibility; see Section 6.

S0004-3702(03)00055-9/SSU AID:1985 Vol.●●●(●●●)
ELSGMLTM(ARTINT):m1a v 1.139 Prn:2/04/2003; 12:56
aij1985
P.11 (1-20)
by:ML p. 11

the head when attempting to characterise the field, or at least how the field characterises itself.

To be fair, when Anderson talks of the "physical grounding project" as being the central project of the embodied approach, he might have something more general in mind than grounding symbols in the manner just discussed. Take, for example, his explication of the project as "centrally involving understanding how cognitive contents (however these are ultimately characterised, symbolically *or otherwise*) must ultimately ground out in (terms of) the agent's embodied experience and physical characteristics" (emphasis added). But this still places an emphasis on representational content, as if EAI agreed with GOFAI that is where the action is.[21]

There is also the drawback that symbol grounding is neither necessary nor sufficient for having a robust form of embodiment. One could be doing AI that placed embodiment centre stage and yet not be producing robots with grounded symbols (e.g., your robot might not have symbols at all); and conversely one could be doing non-embodied, yet grounded (in video cameras and disembodied robot arms), AI. That might explain why the EAI people, who presumably are interested primarily in embodiment, don't talk about symbol grounding much.

Anderson is right in saying that embodiment should be distinguished from situatedness. However, I'm not so sure that "it is the centrality of the physical grounding project [to embodied cognition] that differentiates research in embodied cognition from research in situated cognition" (emphasis removed). Inasmuch as I understand what is meant by grounding, it would seem that being embedded in a world in the proper way has as much to do with it as, if not more than, having a body does. (More on situatedness in Section 5.)

The kind of grounding that I think is important to AI, is, instead, the kind of grounding we saw in Section 3.2 to be essential to overcoming the (theoretical) problem of (pure) rationality. This kind of grounding comes for free:[22] any symbolic planner that doesn't get into an infinite loop can only be so because it is grounded in the non-rational.

5. Getting situated

As I just stated, Anderson is right to distinguish embodiment from situatedness: the former concerns the way that intentional (computational, representational, mental, etc.) properties are dependent on (realised in, made possible through, constrained by) a non-

[21] Perhaps what needs to be grounded is something even more general than representational content (such as "behaviour", "activity", "features", "competences", etc.), the project becomes something so amorphous as to be amenable to the staunchest supported of GOFAI. Who but a dualist would deny that a system's behaviour should be grounded in its physical characteristics? Surely it is the physical constitution of a PC which makes it the case that it can be understood as running Open Office under Linux? What is new here, exactly?

[22] For free? Metaphysically, yes; Epistemologically, no. It is a considerable intellectual and engineering achievement to transform, e.g., a formal specification of a computation (a purely rational system if anything is) into a working, physical implementation. (T.S. Eliot would no doubt say I understate the case: "Between the idea\\And the reality\\…Between the conception\\And the creation…Between the potency\\And the existence…Falls the Shadow" *The Hollow Men*).

S0004-3702(03)00055-9/SSU AID:1985 Vol.●●●(●●●)
ELSGMLTM(ARTINT):m1a v 1.139 Prn:2/04/2003; 12:56
aij1985
P.12 (1-20)
by:ML p. 12

intentionally characterised physical substrate; the latter concerns how such properties are dependent on a (possibly intentional and non-physically characterised) surround. Despite this distinction, much of Anderson's review of the field of embodied cognition mentions or involves situatedness; the above section claimed that Anderson's notion of grounding is probably a situatedness requirement, and his discussions of social embeddedness and exploitation of external representations and "props" (tools, language, etc.) are explicitly about what situated AI might look like.

Anderson looks at a couple of papers from the traditional AI camp [27,46] that attempt to assimilate situatedness into GOFAI. That is, they claim that there is nothing about situatedness that is incompatible with traditional AI theory (e.g., the Physical Symbol System Hypothesis) and methodology/practice. (After all, we are often reminded, the Simon who gave us the Physical Symbol System Hypothesis is the same one who gave us Simon's ant.) While sympathetic to (roughly) a hybrid of GOFAI with embodied/situated AI, Anderson is not impressed with the claims that such a union is business as usual for the GOFAI camp.

In particular, Anderson thinks that Hayes et al. [27] defend GOFAI against the situated onslaught only by redefining GOFAI (and thus, by contrast, situated AI) in a way that trivialises the debate. This trivialisation is meant to be shown by the fact that the resulting definition of situated AI is so implausible and extreme as to be no one's position. Hayes et al. are defending themselves against an imaginary straw man:

> With... our representing heads suitably expanded to encompass the requisite physical and social territory, what is left for SitNanny [the situated cognition theorist—RC] to believe is 'that the representational tokens themselves aren't in the head or that representational token can only have an external, social existence, or even that there isn't any representation at all'. SitNanny may believe this, but I don't know of a single [situated cognition] researcher who does. I conclude from the absence of a citation for this claim that the authors don't know of one, either [4, p. 30], citing [27, p. 20].

But Hayes et al. are not hallucinating an opponent; there are several people who hold views that they ascribe to SitNanny. If Michael Morris, the author of "Why there is no such thing as mental representations" [37] doesn't count because he isn't a situated cognition researcher, one can instead turn to [25], [35], or [32], among many others.

Anderson prefers Vera and Simon's [46] comparative analysis of GOFAI and situated representation, but still thinks they overstate the differences:

> [O]n 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 [4, p. 32].

S0004-3702(03)00055-9/SSU AID:1985 Vol.●●●(●●●)
ELSGMLTM(ARTINT):m1a v 1.139 Prn:2/04/2003; 12:56
aij1985
P.13 (1-20)
by:ML p. 13

But Anderson's "loose" definition of GOFAI in terms of Cartesianism and a "too simple account" of Cognitivism seems to allow plenty of room for Vera and Simon's view.[23] It's the *physical* symbol system hypothesis, after all, not the *cerebral* or even *neural* symbol system hypothesis. One cannot simultaneously lambaste GOFAI for being Cartesian in that it maintains a strong multiple-realisation thesis and an autonomy of the mental from the physical, while simultaneously claiming that its Cognitivism assumes that the symbols of mental processing are specifically cerebral, and not merely neural, bodily, or worldly. If GOFAI theorists are as big on disembodiment as the Cartesian epithet suggests, in that they are the ones that are claiming that robots, nay computers, made of silicon can be intelligent, if their view is the one that implies that properly strung together beer cans [41] can realise thinking, then surely they have no strong claim as to what physical stuff underlies *human* mentality. They certainly won't be baffled by such questions as "What does it mean for a representation… to be encoded *in the body* rather than in the head" [4, p. 33]. If one understands how bodystuff in the head can instantiate or implement ("encode") a representation, one *ipso facto* understands how bodystuff not in the head might do so.

The fact is, situatedness in computational systems is not a new phenomenon. When I, as a recent graduate, was programming a robot to navigate the corridors of Xerox PARC, my algorithms didn't calculate and represent the direction that the robot needed to head in to continue after taking a "sensor-readings-collection" break. Rather, because the robot's heading was unlikely to change during the break, the robot's position itself stored the information as to the heading to take after the break. Around the same time, I was writing programs in LISP on a Macintosh to generate music on an external synthesizer connected by a serial link. In order to know what bytes to send to the serial port to achieve a particular change in the synthesizer, I needed to know the existing value at the address I was about to write to (I could only send bit "masks", not actual bit values). After briefly considering the "GOFAI" idea of constantly maintaining a model of the entire synthesizer, I instead decided to precede all changes with a value query that would tell me the current value. In both cases, I was using the world as its own best model; I was exploiting the relations between the computational system and its environment as a way of offloading (avoiding expensive computational operations). But what's important about this is not that I was doing anything different than anybody else before, at the time, or since. What's important is that what I was doing is so pervasive, banal even, in real-world, traditional computation. Nothing should be read into the fact that my personal examples only cite exploitation of the physical/computational surround. The example of the complex system underlying ship navigation [31], which situated theorists like to cite, is just as much a model for how actual computational systems function by exploiting a complex web of socially-mediated relations, as it is a model for how humans do so.[24]

True, the role of situatedness in cognition has been relatively neglected by traditional AI. True, there is a lot to be done to improve our understanding of the ways that

[23] As it should; if one's construal of GOFAI implied that Herbert Simon wasn't doing it, that's as close as a *reductio ad absurdum* of that definition as one might hope for.

[24] For a study of how the success of mundane computation is achieved only through an extensive context of carefully registered and maintained social embedding, see [1].

S0004-3702(03)00055-9/SSU AID:1985 Vol.●●●(●●●)
ELSGMLTM(ARTINT):m1a v 1.139 Prn:2/04/2003; 12:56
aij1985
P.14 (1-20)
by:ML p. 14

human embodiment facilitates certain forms of situatedness, and our understanding of the means by which artificial systems can exploit their environment as a way of offloading computational tasks. My points are only that (1) situatedness is or can be an important aspect of non-robustly embodied systems (situatedness does not require embodiment), and (2) accommodating situatedness does not require a radical, discontinuous change in the conceptual or theoretical framework of representational AI (although it might require a markedly different *methodology* from, say, that of CYC [34]). For example, perhaps understanding and designing systems with the selective representations of which Anderson speaks (and which are already present in traditional AI systems such as Cassandra [39]) requires a dramatically novel notion, that of non-conceptual content [12,23]. But even if so, it is a notion which will slot into and modify an existing theoretical framework of representations, computation, information processing, etc., not demand starting from a blank (or worse: neurophysiological) theoretical slate.[25]

I do not wish to downplay the need to explore situatedness. Not only do we need better to understand how current computers crucially exploit relations to the world in order to get their jobs done, but we will benefit from understanding how natural systems, including humans, use different forms of situatedness to achieve their goals. What's most striking about the natural cases is that while current computational (as opposed to robotic) situatedness is (almost?) entirely mediated by symbols, the most basic forms of natural situatedness exploit an unconceptualised environment directly. Therefore, discovering ways for computational systems likewise to exploit their non-conceptual surround is an exciting project. But while it's clear that to do so, such systems will have to have substantial non-symbolic and non-conceptual physical aspects, it is not clear that these aspects will sum up to anything that we would recognise as a *body*.

EAI proponents rightly extol the virtues of situatedness, but are less forthcoming in their analysis of its vices. The fact is, there's a trade-off between speed and generality. The more one optimises one's methods to exploit contingencies of the environment, the more one's success is bound to those contingencies; change the context only a little and the situated routine fails, usually in a spectacularly stupid way. A challenge, which Anderson acknowledges (p. 27), is to understand how a system can exploit its situated routines when appropriate, but resort to other means when the context changes (possibly to one which is of a type which has not been encountered before). There seem to be at least five kinds of solution:

- Have a *stored* routine for the new context as well. This has the advantage of being fast and clear how to implement, but the number of routines would be combinatorially prohibitive (compare the "contingency plans" in Section 3.2).

[25] Even Clancey [15, p. 113], in his response to Vera and Simon's paper, concedes: "Certainly, it isn't necessary (or perhaps possible) to break "completely from traditional … theories" (p. 46) but instead to reconsider the relation of our models to the cognitive phenomena we sought to understand. Symbolic models, as tools, will always be with us. Yet, already the shift has begin from viewing them as intelligent beings, to but the shadow of what we must explain.

S0004-3702(03)00055-9/SSU AID:1985 Vol.●●●(●●●)
ELSGMLTM(ARTINT):m1a v 1.139 Prn:2/04/2003; 12:56
aij1985
P.15 (1-20)
by:ML p. 15

- Give up on situatedness for the new context and employ abstract, internal deliberation. This also has the advantage of being clear how to implement, but it has the familiar speed disadvantage of deliberation.
- One important idea that comes out of the study of situatedness is that complex problems (such as dividing a 10-digit number by a 5-digit one) can, through the use of external symbols, be reduced to the iterated application of abilities which were selected for in our evolutionary history (such as pattern matching and association). Thus there is the possibility of replacing the difficult task of abstractly reasoning (in the head) about a novel situation, with abstractly reasoning (using external symbols) about that same situation. But the same complexity considerations apply to both; if internal reasoning about novel situations is intractable, so will external reasoning be.
- A less deliberative application of situatedness would be something like: Suppose we manage to structure our world in such a way that it is highly likely that a novel context will be encountered only immediately after encountering the (familiar) contexts to which it is most similar. In such a case, (fragments of) the most recently active routines (or their recombinations) will have a good probability of being applicable in the new situation.
- A related idea is to *compute* a new routine based on existing ones. For example, if one's routines were the outcome of a process with a few continuously-varying parameters, and the system could reliably adjust these parameters in the light of changing context in such a way that a situated routine appropriate for dealing with the new context was the result, then perhaps this could be a solution. Or, if routines can be ranked in terms of the degree to which they are situated (roughly, the number of contexts in which they work), then perhaps a "situatedness-reducing" operation could be applied to the routine that was being used in the prior context, yielding one of slightly greater abstraction, enough to work in the new situation. Some connectionist architectures seem to implement such "perspective-dependence reducing" operations [11,17]. Of course, there should be complementary "situatedness-increasing" operations, which would allow one to maximise one's offloading onto the environment.

6. In our own (body) image?

Much discussion of EAI centres on how important embodiment is to human cognition, and this is appropriately reflected in Anderson's review. Lakoff and Johnson [33] are probably right that the nature of many, most or all of our concepts are highly constrained by our modes of perception and action, as well as the details of our neurophysiology, hormone systems, etc. Following the discussion above, this suggests either:

1. Building systems whose conceptual repertoires are constrained by their non-conceptual substrate is a novel and promising way to do AI; or
2. It is by virtue of having the non-conceptual substrate that they do that intentional systems, be they natural (with animal bodies) or artificial and symbolic, are able to demonstrate any intelligence or intentionality at all.

S0004-3702(03)00055-9/SSU AID:1985 Vol.●●●(●●●)
ELSGMLTM(ARTINT):m1a v 1.139 Prn:2/04/2003; 12:56
aij1985
P.16 (1-20)
by:ML p. 16

But some EAI discourse goes further, claiming that the dependence of our human concepts on our human embodiment demonstrates the necessity of a body, even a human one, for concepts or intelligence. Without at least one further premise, this is a *non-sequitur*; is there an argument to support this stronger claim concerning the necessity of a robust form of embodiment for AI?

I get the impression that some people who espouse this form of Strong EAI do so because of a (mistaken, in my view) belief that someone, somewhere has already established it as fact. If their beliefs are more specific than that, then, typically, the someone is John Searle and the somewhere is the Chinese room argument [41]. They take Searle to have shown that understanding/consciousness/intelligence is a biological, not computational, phenomenon. Never mind that he hasn't shown that; that the biological source he has in mind is only the brain, not the body; and that his argument concedes that a non-embodied system (e.g., the Chinese room itself) can pass the Turing test. Also, although pure behaviourism should be resisted, it would be the height of political incorrectness and anthropocentrism to deny full subjecthood to a system that behaves intelligently simply because that behaviour is not generated by the same biological mechanisms as it is in us.[26]

But philosophically astute EAI proponents can do better. One way of arguing for the strong EAI claim is to insist that:

(1) What counts as intelligence depends on what we can understand or recognise to be intelligent; and
(2) We can only understand or recognise another organism to be intelligent if they share our "way of being" (possibly the same as *Dasein* [28]?) or what Wittgenstein called our "form of life": the situations, goals, challenges, tasks, etc. that make up our experienced existence;
(3) Two organisms can share a way of being or form of life only insofar as they are physically (i.e., bodily) similar.[27]

A full analysis of this argument can't be given here, but I'll make three remarks:

- Premise 1 has all the strengths and weaknesses of a verificationist/anti-realist metaphysical position;

[26] Where "us", chillingly, refers to those in power. When people say they would deny ascriptions of mentality, consciousness, etc. to a intelligently-behaving robot because it is not built the same way they are, I fear a day when that philosophy is used to deny me my own subjectivity on the grounds that my brain is different from the brains of the powers that be! Lest one think this pure fantasy, consider that this form of reasoning already forced Anderson to undertake a defense of the intentional properties of the handicapped. This is a grotesque situation to be in.

[27] Cf., e.g., [20]. The similarity might have to be very close indeed, according to Wittgenstein ("If a lion could talk, we would not understand him" [48, II, xi, p. 223]) and Nagel ("What is it like to be a bat?" [38]). It's true that for Wittgenstein, mere sameness of behaviour is enough: "Only of a living human being and what resembles (*behaves like*) a living human being can one say: it has sensations; it sees; is blind; hears; is deaf; is conscious or unconscious." [48, Section 281]. But how similar can the behaviour of the Chinese room be to my behaviour, given that I have a body and it does not?

S0004-3702(03)00055-9/SSU AID:1985 Vol.●●●(●●●)
ELSGMLTM(ARTINT):m1a v 1.139 Prn:2/04/2003; 12:56
aij1985
P.17 (1-20)
by:ML p. 17

- One could accept premise 3, yet insist that physical similarity can be measured in more abstract terms—functional similarity, say, which need not require sameness of embodiment;
- It would behave the EAI community to develop and refine the above argument, exploring the many possible intermediate positions between implementation-independence and chauvinism, lapsing into neural reductionism concerning mental states only as a last resort.

7. Radical embodiment

If Anderson's review is anything to go by, then I think EAI proponents understate the potential impact that embodiment may have on artificial intelligence and computation in general. Even if EAI is an entirely new way of doing AI, I think there are two ways in which EAI can be more radical that adding new methodologies to the toolbox.

7.1. Conceptual, not ontic, novelty

One of these ways has been a recurring theme in the preceding discussion. Specifically, the promise of EAI is not just the promise of a kind of AI that is an alternative to GOFAI, but the promise of providing the conceptual tools necessary for explaining much or most of the successes of traditional, symbolic computation as well. On this view, what makes even much of traditional, arch-symbolic AI *work* (when it does!) is best understood in terms of the way such systems are situated in the physical and social world, their modes of activity, and their physical manifestation. Restricting ourselves to looking only at the internal, abstract properties of computational systems, although a useful focus in some situations, is in general a blinding hindrance to full explanation and design. On this view, embodied (and situated) computational systems aren't so much *ontically* novel as *conceptually* novel: they've been around for ages, but we just didn't know it.

Seeing the contribution of EAI in this way allows one to answer the questions that I sometimes get when talking about EAI or situated robotics with some AI researchers. "Isn't all AI work embodied in some sense?" they ask. "Aren't all robots situated?". Yes, and yes; but only an AI theory which has been enriched by substantive concepts of embodiment and situatedness goes beyond this fact by *acknowledging* that these aspects of the system are (often) essential to their explanation and design, and using said concepts to those ends. Any ontic ubiquity of embodiment and situatedness in the area of intelligent systems should make us suspect that they should be made conceptually ubiquitous as well.

7.2. Interactivist empiricism

An even more radical role for EAI impacts on the development of our own concepts and understanding of computation and the mind. Much of recent cognitive science has emphasised the role of action, perception and experience, as opposed to disembodied inference and reasoning, in human cognition. But since cognitive scientists and AI

S0004-3702(03)00055-9/SSU AID:1985 Vol.●●●(●●●)
ELSGMLTM(ARTINT):m1a v 1.139 Prn:2/04/2003; 12:56
aij1985
P.18 (1-20)
by:ML p. 18

researchers are themselves humans,[28] it stands to reason that concept learning and development, as well as theory change, in cognitive science and AI should exploit the experiential aspect of cognition when possible. First, one should acknowledge that the goal of cognitive science is an explanation *for* experiencing agents, not (primarily) the appearance of a set of marks on paper in some journal. Second, one can ask what is required for such understanding; it appears that science in general has been overly preoccupied with theories. Theories doubtlessly play a crucial role, but there may be modes of understanding which only alternative forms of explanation, such as (design of and interaction with) real-world AI artefacts and implemented virtual machines, can provide. Even if all forms of understanding can, in some sense, and in principle, be written down (a concession I am not actually prepared to make), it still seems that writing them down is not always an adequate means of transmitting that understanding.

Thus, it may be helpful, or even necessary, for discovering the concepts which we will need for an understanding of intelligence (and therefore which we will benefit from or need for *designing* intelligent systems), that we interact with or personally take part in the creation of actual, working AI systems. To come up with, perhaps even to grasp, these concepts may require various forms of experience with a working system, might require getting one's hands dirty, rather than simply reading about the system (or even about experiences with the system, were they ever published) in a journal article. If this is so, then factors which are often thought to be of marginal interest to AI (or of interest for some other reason) become central. For example, the runtime details of the model/simulation, its interface/graphical display, etc. Two AI systems that are identical in their behaviour, but differ in, e.g., the GUI which allows one to perturb and observe the states of those systems, may have drastically different impacts on the understanding and conceptual development that they provide.[29] Perhaps here (finally) we have a reason why AI must be made in our own image (cf. Section 6): Because that is the only way that we will be able to grasp and refine the concepts necessary for AI development. If this is right, giving our AI systems a robust form of embodiment may have as much to do with developing our own mental abilities as it does with developing theirs.

Acknowledgements

I'd like to thank David Brooks, Dave Gurnell, Dean Petters and Aaron Sloman, all members of the CogAff research grouping of the School of Computer Science at the University of Birmingham, for their helpful comments. This work was supported by a Research Fellowship funded by grant F/94/BW from the Leverhulme Trust.

[28] Usually.

[29] Many of Dourish's [19] considerations, and thus Section 6 of Anderson's field guide, become very germane here, though perhaps not in a way either anticipated.

S0004-3702(03)00055-9/SSU AID:1985 Vol.●●●(●●●)
ELSGMLTM(ARTINT):m1a v 1.139 Prn:2/04/2003; 12:56
aij1985
P.19 (1-20)
by:ML p. 19

References

[1] P. Agre, Computation and Human Experience, Cambridge University Press, Cambridge, 1997.
[2] P. Agre, D. Chapman, Pengi: An implementation of a theory of activity, in: Proceedings of the Sixth National Conference on Artificial Intelligence (AAAI-87), Seattle, WA, AAAI Press, 1987, pp. 268–272.
[3] A.P. Ambler, H.G. Barrow, C.M. Brown, R.M. Burstall, R.J. Popplestone, A versatile computer-controlled assembly system, in: The Proceedings of the Third International Joint Conference on Artificial Intelligence, (IJCAI-73), Stanford, CA, 1973, pp. 298–307.
[4] M. Anderson, Embodied cognition: A field guide, Artificial Intelligence (2003), This issue.
[5] P. Armer, Attitudes towards intelligent machines, in: R.L. Chrisley (Ed.), Artificial Intelligence: Critical Concepts, Vol. IV, Routledge, London, 2000, pp. 325–342, Originally published in 1962.
[6] R. Beer, A dynamical systems perspective on agent-environment interaction, in: R.L. Chrisley (Ed.), Artificial Intelligence: Critical Concepts, Vol. III, Routledge, London, 2000, pp. 210–255, Originally published in 1995.
[7] C. Breazeal, B. Scassellati, Infant-like social interactions between a robot and a human caretaker, Adaptive Behavior 8 (2000).
[8] R. Brooks, Intelligence without representation, in: D. Kirsh (Ed.), Foundations of Artificial Intelligence, MIT Press, Cambridge, MA, 1992, pp. 139–159.
[9] L. Carroll, What the tortoise said to Achilles, Mind 4 (1895) 278–280.
[10] C. Cherniak, Minimal Rationality, MIT Press, Cambridge, MA, 1986.
[11] R. Chrisley, A. Holland, Connectionist synthetic epistemology: Requirements for the development of objectivity, in: L. Niklasson, M. Bodén (Eds.), Current Trends in Connectionism: Proceedings of the 1995 Swedish Conference on Connectionism, Lawrence Erlbaum, Hillsdale, NJ, 1995, pp. 283–309.
[12] R.L. Chrisley, Taking embodiment seriously: Non-conceptual content and robotics, in: Android Epistemology, MIT Press, Cambridge, MA, 1995, pp. 141–166.
[13] R.L. Chrisley (Ed.), Artificial Intelligence: Critical Concepts, Routledge, London, 2000.
[14] R.L. Chrisley, T. Ziemke, Embodiment, in: L. Nadel (Ed.), Encyclopedia of Cognitive Science, Macmillan, London, 2002.
[15] W.J. Clancey, Situated action: A neuropsychological interpretation response to Vera and Simon, Cognitive Sci. 17 (1993) 87–116.
[16] A. Clark, Being There, MIT Press, Cambridge, MA, 1996.
[17] A. Cussins, The connectionist construction of concepts, in: M. Boden (Ed.), The Philosophy of Artificial Intelligence, Oxford University Press, Oxford, 1990, pp. 368–440.
[18] D. Dennett, A cure for the common code?, in: D. Dennett (Ed.), Brainstorms, Harvester Press, Brighton, 1978, pp. 90–108.
[19] P. Dourish, Where the Action Is: The Foundations of Embodied Interaction, MIT Press, Cambridge, MA, 2001.
[20] H. Dreyfus, Introduction to the MIT Press edition of "What Computers Can't Do", in: R.L. Chrisley (Ed.), Artificial Intelligence: Critical Concepts, Vol. III, Routledge, London, 2000, pp. 459–493, Originally published in 1992.
[21] J. Fodor, Psychosemantics: The Problem of Meaning in the Philosophy of Mind, MIT Press, Cambridge, MA, 1987.
[22] N. Goodman, Fact, Fiction, and Forecast, Harvard University Press, Cambridge, MA, 1954.
[23] Y. Gunther (Ed.), Essays on Nonconceptual Content, MIT Press, Cambridge, MA, 2003.
[24] S. Harnad, The symbol grounding problem, Phys. D 42 (1990) 335–346.
[25] I. Harvey, Untimed and misrepresented: Connectionism and the computer metaphor, Cognitive Science Research Paper CSRP 245, School of Cognitive and Computing Sciences, University of Sussex, Brighton, November 1992.
[26] J. Haugeland, Artificial Intelligence: The Very Idea, MIT Press, Cambridge, MA, 1985.
[27] P. Hayes, K. Ford, N. Agnew, On babies and bathwater: A cautionary tale, AI Magazine 14 (1994) 15–26.
[28] M. Heidegger (Ed.), Being and Time, Harper and Row, New York, 1927.
[29] I. Horswill, Characterizing adaptation by constraint, in: Toward a Practice of Autonomous Systems: Proceedings of the First European Conference on Artificial Life, MIT Press, Cambridge, MA, 1992, pp. 58–64.

S0004-3702(03)00055-9/SSU AID:1985 Vol.●●●(●●●)
ELSGMLTM(ARTINT):m1a v 1.139 Prn:2/04/2003; 12:56
aij1985
P.20 (1-20)
by:ML p. 20

[30] D. Hume, An Enquiry Concerning Human Understanding, Oxford University Press, Oxford, 1777. Edited by Nidditch and Selby-Bigge.
[31] E. Hutchins, Cognition in the Wild, MIT Press, Cambridge, MA, 1995.
[32] F. Keijzer, Representation and Behavior, MIT Press, Cambridge, MA, 2001.
[33] G. Lakoff, M. Johnson (Eds.), Philosophy in the Flesh: The Embodied Mind and its Challenge to Western Thought, Basic Books, New York, 1999.
[34] D.B. Lenat, E.A. Feigenbaum, On the thresholds of knowledge, in: D. Kirsh (Ed.), Foundations of Artificial Intelligence, MIT Press, Cambridge, MA, 1992, pp. 185–250.
[35] H.R. Maturana, F.J. Varela, Autopoiesis and Cognition: The Realization of the Living, in: Boston Studies in the Philosophy of Science, Vol. 42, D. Reidel, Dordrecht, 1980.
[36] D. McDermott, We've been framed: Or, why AI is innocent of the frame problem, in: Z. Pylyshyn (Ed.), The Robot's Dilemma: The Frame Problem in Artificial Intelligence, Ablex, Norwood, 1987.
[37] M. Morris, Why there are no mental representations, Minds and Machines 1 (1991) 1–30.
[38] T. Nagel, What is it like to be a bat?, in: N. Block (Ed.), Readings in Philosophy of Psychology, Vol. 1, Harvard University Press, Cambridge, MA, 1980.
[39] L. Pryor, G. Collins, Cassandra: Planning with contingencies, Technical Report 41, Institute for the Learning Sciences, Northwestern University, Evanston, 1993.
[40] Z. Pylyshyn, Computation and Cognition: Towards a Foundation for Cognitive Science, MIT Press, Cambridge, MA, 1984.
[41] J. Searle, Brains, minds and programs, Behavioral and Brain Sciences 3 (1980) 417–457.
[42] R. Sheldrake, The Presence of the Past: Morphic Resonance and the Habits of Nature, Grimes, 1988.
[43] H.A. Simon, Theories of bounded rationality, in: R. Radner, C.B. Radner (Eds.), Decision and Organization, North-Holland, Amsterdam, 1972, pp. 161–176.
[44] B.C. Smith, The foundations of computing, in: M. Scheutz (Ed.), Computationalism: New Directions, MIT Press, Cambridge, MA, 2002.
[45] E. Thelen, L. Smith, A Dynamic Systems Approach to the Development of Cognition and Action, MIT Press, Cambridge, MA, 1994.
[46] A.H. Vera, H.A. Simon, Situated action: A symbolic interpretation, Cognitive Sci. 17 (1993) 7–48.
[47] B. Webb, Using robots to model animals: A cricket test, Robotics and Autonomous Systems 16 (1995) 117–134.
[48] L. Wittgenstein, Philosophical Investigations, Blackwell, Oxford, 1953. Edited by G.E.M. Anscombe and R. Rhees; Translated by G.E.M. Anscombe.
[49] M. Yazdani, B.R. Whitby, Building birds out of beer cans, Robotica 5 (1987) 89–92.