On Babies and Bathwater

A Cautionary Tale

Patrick J. Hayes, Kenneth M. Ford, & Neil Agnew

■ One should not throw out the baby with the bathwater, according to an old aphorism. Some popular recent positions in AI thinking have done just this, we suggest, by rejecting the useful idea of mental representations in their overenthusiastic zeal to correct some simplifications and naïveties in the way traditional AI ideas have sometimes been understood. These "situated" perspectives correctly emphasize that agents live in a social world, using their environments to help guide their actions without needing to always plan their futures in detail; but they incorrectly conclude that the very idea of mental representation is mistaken. This perspective has its intellectual roots in parts of recent sociological thinking which reject the entire fabric of western science. We discuss these ideas and disputes in the form of an illustrated fable concerning nannies and babies.

A Cast of Characters

Once upon a time there were two happy and healthy babies. We will call them Representation Baby (closely related to Mind Baby and Person Baby) and Science Baby (closely related to Reality Baby).

These babies were so charming and inspirational that for a long time their nannies cared for them very well indeed. During this period it was generally the case that ignorance was pushed back and human dignity increased. Nannies used honest, traditional methods of baby care which had evolved during the years. Like many wise old folk, they were not always able to articulate good justifications for their methods, but they worked, and the healthy, happy babies were growing well and having lots of fun.

Unfortunately, some newer nannies haven't been so careful, and the babies are in danger from their zealous ways. We will focus on two nannies who seem to be close friends and often can be seen together—Situated Nanny (called SitNanny for short) and Radical Social Constructivist Nanny (known to her friends as RadNanny).[1]

SitNanny Meets RepBaby

SitNanny is fanatical about a certain kind of firmness. She just hates for things not to be firmly attached to the world. To be fair, this obsession may be a natural consequence of her having spent so much time in California.

She believes that traditional ways of bringing up baby are far, far too slack in this regard. Babies shouldn't think and plan so much, she believes, but instead should be kept firmly attached to the ground.[2]

SitNanny declares (in fact, she *preaches,* loudly and regularly) that all this old-fashioned cognitive stuff is just dirty bathwater, and she is going to throw it out. She believes that it holds babies back, stunting their development, and that a nursery based on such nonsense is an unhealthy place. She is so fanatical about her views, however, that she is quite willing to throw out the baby with the bathwater.

Not long ago, Representation Baby (also called RepBaby) was reflecting on a heretofore happy childhood when rather suddenly he found himself threatened by SitNanny, who adopts just this position.

SitNanny argues passionately with those who suggest that keeping the baby might be a wise idea, going so far as to deny that there is a baby there at all. She argues further that her

Situated Nanny Holding Things Down

task is simply to get the nursery clean in preparation for other babies.

We think that SitNanny is at heart a good person, and many of her suggested reforms are valuable. The intellectual nursery did need some tidying up. But she has been reading political tracts (written by such people as Dreyfus and Searle—the Berkeley Brothers) and has become something of a fanatic.

Why would SitNanny want to threaten RepBaby, and what does it mean to be "situated"? There are several possible answers. One emphasizes the fact that agents' knowledge of their world is incomplete, partial and usually incorrect, but that they must nevertheless act in that world, often quite promptly. This view leads to a vision that rejects the idea of planning in order to act. Planning, it is said, makes the unjustified assumption that the planner has access to all relevant facts, and it is a time-consuming activity that ignores the ticking of the real clock (Agre & Chapman, 1987). It would be hard to play baseball, for example, if one had to be always planning one's actions. It seems clear that we often, perhaps usually, simply start on a course of action in the blithe confidence that we will be able to handle any difficulties which may arise along the way. We just *do* it. And we often rely explicitly on the external world to guide our actions. If you are driving in downtown San Francisco and want to get to the airport, the best method is to find Highway 101 south, get in the left-hand lane, and *stay in that lane* through all subsequent junctions and turns. It will eventually become the fast lane on the road to the airport. There is no need to plan a route: that left-hand lane will take you there.

This disillusionment with the traditional AI emphasis on planning combines naturally with an interest in how simple organisms get on. SitNanny has noticed that while insects, for example, have very impoverished cognitive abilities and never seem to *plan* anything in the traditional AI sense of the term, there are an awful lot of them around. Even quite simple organisms can exhibit surprisingly complex behavior when placed in certain social and physical settings—a point made by Herb Simon (one of RepBaby's grandfathers) years ago with his example of the ant traversing a beach of pebbles.

Further, it seems insightful to ask how much human behavior arises from, or might be simulated by, mechanisms that involve little or no explicit representation of the external environment. Even if one is concerned with representations and planning and so forth, the zero case is of interest. Such a question certainly leads in a different direction than the traditional AI work in planning and focuses interest on different concerns. RepBaby can live with this. It provides at the very least a certain intellectual discipline. Babies, even RepBabies, need to be set firmly on solid ground quite often and (one might argue) should learn to walk before they try to play chess.

However, the term "situated" has become identified with a much stronger and more radical collection of assumptions, in particular an attack on the basic idea of knowledge representation. We believe that these assumptions are unwarranted and often based on fundamental misunderstandings. In particular, one 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. Most of these more radical positions seem to arise from reactions against what might be called straw babies, versions of classical AI ideas that in fact are much too simplistic. SitNanny seems to think that the AI concept of "physical symbol system" amounts to a claim that programs, mental representations, and external texts are indistinguishable (e.g., Clancey, 1993) and that the AI vision of knowledge is of something rigid, unchangeable, and context-insensitive, a kind of "mental toolkit" (Lave, 1988). In our view, all these positions misunderstand the essential content of the physical symbol system hypothesis as it is most widely understood in AI. Particular AI systems (often, ironically, those that SitNanny was working on before she got her new perspective) may make such simplifying assumptions, but there is no reason why the general idea of symbolic representation must be so limited, and indeed it traditionally has not been thought of in such a limited way.

The intellectual nursery did need some tidying up.

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. Let us be frank: we think the representational hypothesis is a great idea. The reasons for being so positive are well documented, but we have two main justifications for our enthusiasm. First, it accounts for much that is otherwise completely puzzling about how cognition could happen in the

Representation Baby Is Worried about His Future physical world; second, it allows experiments and makes empirical predictions, which have so far largely been confirmed. But one needs to understand the key word "representation" in a sufficiently broad fashion.

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 or motor skills at any cognitive level. None of these are in any way at variance with the representationalist hypothesis. RepBaby can play happily with these toys.

What a representation *means* (not what "representation" means, notice) is a complex question, but a meaning might be relative to such things as goals, attitudes or purposes, and it need not presuppose that the environment is uniquely determined. In fact, available semantic accounts of representation languages—model theories—emphasize the extent to which such relativity is probably inevitable.

SitNanny speaks with many voices, but a process of reconstruction and guesswork suggests to us that SitNanny understands representation in a different sense from that in which it is used in the framework she rejects so vehemently. In fact, several of the new critics (some having noisily jumped ship) take representation to mean something like a text or a picture, an object consciously manipulated by talking to oneself or visualizing internal imagery, and identical in nature to external representations such as writings or diagrams. For example, Clancey (1993) assumes that the use of a representation is always conscious and deliberate, and if a process is "pre-linguistic" then representations are not involved in it. But RepBaby was not brought up this way. It is at variance with the usual meaning of the term as used in the representationalist position in cognitive science, which hypothesizes internally represented knowledge in a much broader sense, including the computational modeling of unconscious cognitive processes and processes of social interaction. One might believe that this broader sense is somehow inappropriate or incoherent, but such a position needs to be argued, not simply asserted. Several of the critics simply assume that calling something a representation entails that it has this external, perceived character. We will call this mistake the "textual fallacy."

Clancey (correctly) emphasizes that post hoc justifications for action are not to be identified with reasons for acting, apparently believing that this confusion—which may once have been a common one in the expert system field in which he worked—is endemic to the representationalist perspective; but it is not. John Searle makes a similar mistake when he concludes that anything represented must be equated with a conscious thought, and any proposed mechanism of unconscious activity must therefore not involve representations. Lucy Suchman (1987), a more subtle critic, also commits this mistake by drawing a sharp contrast between cognitive science's rigid, mechanical view of plans and the weak resource view endorsed by situated action theory. However these are not different views of a single idea, but two distinct ideas: plans as programs or data structures, and plans as external maps or texts.

SitNanny has been accused of this error before (Vera & Simon, 1993) and has protested that she does not mean to throw away all notions of symbols or mental representation, but here she misses the point. What makes RepBaby so useful is that he makes physical symbols provide an *explanation* of mental phenomena, and this explanatory role is what SitNanny rejects. Her new clean bathwater will be scented with all manner of things: Gibsonian ideas of direct perception; concepts from fringe neurology, sociology, ethnomethodology, and political theory; precomputational psychological theory; and God knows what else. But the central idea of physical symbol system will be purged from it (Clancey, 1993; Greeno & Moore, 1993).

SitNanny is unhappy with both parts of the term "knowledge representation."

Where's the Beef?

When listening to SitNanny tell us that representations are not in the head, we often wonder, "Where's the beef?" According to SitNanny, not only are there no mental representations, but individuals do not possess knowledge. SitNanny is unhappy with both parts of the term "knowledge representation." Not only is the idea of representation seen as misleading and false, but even the notion of knowledge itself, as we understand it, seems deeply suspect. Knowledge and meaning are seen as extra-personal and located in the community rather than in the head. We need to get our terminology clear. 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? Well, let us try to converge on the locus of plumbing knowledge. If someone were to ask a question about hot-water cylinders, it would be correct to tell him that the answer could be found in the kitchen, so it seems correct to say that the knowledge is in the kitchen.

Where in the kitchen is it, then? Perform a simple experiment by removing the plumber from the kitchen. Now nobody in the kitchen knows much about plumbing, and it would be wrong to say that that was where the answer could be found. It seems clear that the plumber's knowledge of plumbing is firmly attached to the plumber. Perhaps some kind of invisible mental leash would explain its attachment to him, but we suggest that a simpler and more plausible theory of this phenomenon might be that the plumber's knowledge of plumbing is in his head and therefore moves with him, much as his kidneys and his toenails do.

We believe that SitNanny's attack on RepBaby is unwarranted and often based on fundamental misunderstandings.

RepBaby is worth saving.

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). A memory can be damaged by banging a head, but never by stubbing a toe or digging a hole.

Notice this is not to claim that memories are veridical or that they are anything as simple as a replaying of earlier recordings. They might be the result of some complex and creative process of reconstruction with which we all constantly rewrite our past. And it might be correct to say that we are almost never doing just one thing, that our actions are always embedded in some social context; although we think this fact is less significant than SitNanny does. But whatever is going on when we think or when we are recalling something from our past, it is happening inside us, not somehow out there in the society.

There is, however, a rather different way to understand SitNanny's social-meaning claim. 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 part. Wittgenstein argued that the idea of a language spoken by a single person is incoherent: languages are systems of communication between agents. The sociological stance takes this idea rather further and places the meanings not, as it were, distributed among the heads of the members of the society, but in the society itself. Going one step beyond Jung, we have to imagine something like an invisible collective consciousness—what Harry Collins calls a "collectivity"—in which the meanings of our individual beliefs are located. 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. But notice the difference between this idea and the earlier one. This idea says that meanings of mental representations are socially determined; the other claims that the representational tokens themselves aren't in the head or that representational tokens can have only an external, social existence, or even that there isn't any representation at all. This conflict is like the difference between insisting that RepBaby wear fashionable clothes and claiming that there isn't anyone there to be dressed.

Suppose the plumber in the kitchen realizes he has forgotten his flux and knows that a soldered joint might therefore be unreliable, so he contemplates using a compression fitting in spite of its extra cost. Perhaps indeed a proper account of what these terms mean might involve a description of the plumbing community and his relationship to it. Any experience, including this one, may change and enrich his knowledge of joining pipes. But these are mental tools that he uses—or, more properly, mental mechanisms out of which he is formed—and when he leaves the kitchen, they go with him.

In summary, the central concept of mental representation is more robust than realized by SitNanny. We believe that SitNanny's attack on RepBaby is unwarranted and often based on fundamental misunderstandings. RepBaby is worth saving.

"Hey, Hey, Ho, Ho, Science Baby Must Go"

Now we come to Science Baby—an extremely robust and healthy lad. SitNanny doesn't bear Science Baby much harm. But RadNanny, one of SitNanny's mentors, is much more dangerous—RadNanny belongs to a cult. Remarkably, even this most useful baby is under attack by RadNanny and the members of her cult.

Like an anthropologist studying a primitive tribe, RadNanny (and her band of sociol-

Science Baby Doing His Sums

The Notorious RadNanny Looking for Babies ogists) has moved into scientific laboratories to peer over the shoulders of scientists. RadNanny has noticed that the behavior of real scientists typically does not follow the abstract ideal of disinterested pursuit of truth. However RadNanny draws the startling and extreme conclusion that all science is arbitrary and that reality is merely a construction of a social game.

This anti-epistemic and also anti-scientific stance is clear in statements such as "the natural world has a small or nonexistent role in the construction of scientific knowledge" (Collins 1981, p. 3) and "the fashioning of normative models of thinking from particular, 'scientific,' culturally valued, named bodies of knowledge is a cultural act" (Lave 1988, p. 172). RadNanny claims that since we have no direct access to "reality," science (like religion, politics, and literature) merely reflects the myth-making tendencies of the human tribe, with one myth being no more reality-based than another.

On this view, science should not be granted any kind of relative authority or have any particular success acknowledged. It is just one cultural tradition: Zen Buddhism, medieval Christianity, or even the semi-coherent cynicism of Beavis and Butthead are all perfectly valid alternatives. In fact, sometimes scientific rationality is seen as morally or politically *inferior* to Beavis and Butthead: "Logic, as Marx has it, is the money of the mind, and no matter how dialectical, it always expresses a reified and alienated mediation of man and reality" (Warren, 1984 p. 50, cited in Lave, 1988 p. 173). We have criticized this cultural relativism elsewhere (Agnew, Ford, & Hayes, 1994), so here we will be brief.

By any reasonable measure, Science Baby has been spectacularly successful. Perhaps no single scientist operates according to the ideal rules, as RadNanny delights in telling us, but the approximate global effect is to produce knowledge that reality (the hidden-hand editor) has had an opportunity to grip as firmly as it can be made to.

RadNanny claims that the natural world plays little or no part in the beliefs of scientists, and that logic is only a tool used by one class to enslave others; but what can be the basis for her beliefs other than empirical observation of the natural world (which, of course, includes the scientists she studies), and what logic does she use to explain her ideas? Are RadNanny's sociological friends somehow issued special glasses that allow them to see more clearly than other scientists?

SitNanny often enthusiastically endorses this radical social constructivism because it can be seen as locating knowledge in the community rather than in the head and so can be taken to support a position that denies a central role to internal representations. We sometimes wonder if SitNanny really understands the full import of RadNanny's doctrines.

It is important to notice that it is not just Science Baby that is under attack here, but the whole notion of an objective reality. RadNanny's ideas are a sort of universal baby solvent. However, we are not too worried about Reality Baby. Even RadNanny will duck if one throws a brick at her.

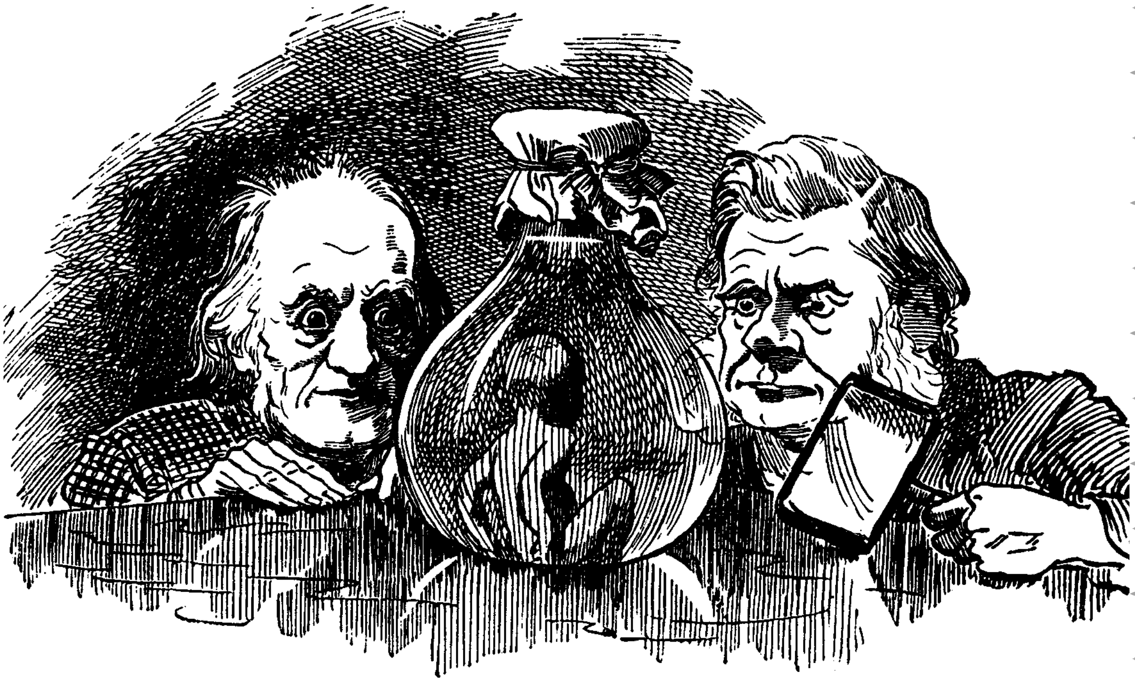
Conclusion

These new nannies have a lot of valuable reforms to make. We aren't trying to argue simply for conservatism. Sometimes the old-fashioned nannies were (and still are) using dirty bathwater. One does need to keep RepBaby under control—if you let him, he will try to get hold of everything.

SitNanny is right to emphasize that a lot can be done without involving detailed, explicit knowledge of it. Sometimes the best plan is just to follow the road. But this musn't be taken too far. It is really an observation about what kind of information *should* be represented rather than a rejection of the idea of knowledge representation itself. In any case, the question of how much planning goes on is essentially an empirical one. Something clever happens when a good baseball fielder begins to run toward a catch before the bat has made contact with the ball. Maybe this isn't planning in a very simplistic sense, but it isn't just following the left lane either.

Even RadNanny has some things of value to give us. A certain kind of naïve realism is wrong, and human scientists sometimes do all sorts of strange human things. But if the radical constructivists would just take the time to actually look at what the knowledge representation idea means, they would see that it has already passed that level of naïveté. If someone's perceptions are representations, then of course they aren't in direct touch with the world (whatever that might mean). RepBaby is a pragmatic constructivist, but he gets on quite happily with Reality Baby. Science Baby likes both of them. RadNanny, SitNanny, and others of heightened social awareness have been doing some new and exciting work that can rinse away a

Reality Baby in a Relativist Stew



Examining our Assumptions

certain kind of conceptual grime that sometimes clouds our thinking. It is healthy to be obliged to examine our own assumptions from time to time.

But we must not overreact and throw away everything. Such absolute rejections reflect thinking that is just as rigid and dogmatic as the simplistic naïveté they purport to overcome. People sometimes forget that the only reason for having nannies at all is to look after the babies.

Notes

(1) There are those who say that these are not new nannies at all, but old nannies in new dresses. For example, it is sometimes said that SitNanny often looks like an older nanny who was fired for cause, the notorious Nanny Skinner.

(2) Like insects, in fact (Brooks, 1991).

References

Agnew, N.; Ford, K. M.; and Hayes, P. J. 1994. Expertise in Context: Personally Constructed, Socially Selected and Reality Relevant? *International Journal of Expert Systems* 7(1): 65–88.

Agre, P., and Chapman, D. 1987. PENGI: An Implementation of a Theory of Activity. In Proceedings of the Seventh National Conference on Artificial Intelligence, 268–272. Menlo Park, Calif.: American Association for Artificial Intelligence.

Brooks, R. A. 1991. Intelligence without Representation. *Artificial Intelligence* 47:139–159.

Clancey, W. 1993. Situated Action: A Neuropsychological Interpretation. *Cognitive Science* 17:87–116.

Collins, H. M. 1981. Stages in the Empirical Programme of Relativism. *Social Studies of Science* 11:3–10.

Greeno, J. G., and Moore, J. 1993. Situativity and Symbols. *Cognitive Science* 17:49–59.

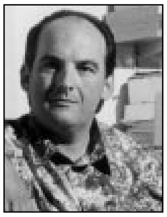
Lave, J. 1988. *Cognition in Practice: Mind, Mathematics, Culture in Everyday Life*. New York: Cambridge University Press.

Suchman, L. 1987. *Plans and Situated Actions: The Problem of Human-Computer Communication*. New York: Cambridge University Press.

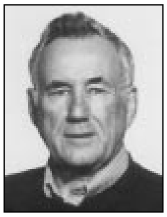
Vera, A., and Simon, A. H. 1993. Situated Action: A Symbolic Interpretation. *Cognitive Science* 17:7–48.

Warren, S. 1984. *The Emergence of Dialectical Theory.* Chicago: University of Chicago Press.

Pat Hayes is a professor of computer science and philosophy at the University of Illinois at Urbana. Hayes's research interests include knowledge representation, machine inference, and the philosophical foundations of AI. He has published several influential papers in these areas. Hayes organized one of the first interdisciplinary cognitive science programs, at the University of Rochester. He has held several academic and industrial positions in aspects of AI. Hayes has been secretary of AISB, a chairman and trustee of IJCAI, governor of the Cognitive Science Society, and President of AAAI.



Ken Ford is currently Director of the Institute for the Interdisciplinary Study of Human and Machine Cognition at the University of West Florida. His research interests include the philosophical foundations of AI, knowledge-based systems, issues of representation, and constructivist approaches to cognitive science and AI. He has written more than a hundred scientific papers, and is the author and editor of three books. He received his Ph.D. in computer science from Tulane University. Ford is Editor-in-Chief of the AAAI Press, Executive Editor of the *International Journal of Expert Systems,* Associate Editor of the *Journal of Experimental and Theoretical Artificial Intelligence,* and is a *Behavioral and Brain Sciences* (BBS) Associate. He is past president of Florida Artificial Intelligence Research Society (FLAIRS).



Neil Agnew is a professor of psychology at York University, Toronto Canada. His academic interests include decision making under uncertainty, models of individual and institutional change, and the theory of science. He has published extensively in these fields and has just published the sixth edition of his text *The Science Game.* Neil has occupied various posts in both professional and academic associations, and is a fellow of the Canadian Psychological Association.