

# Constructing Common Sense

CATHERINE SHERRON

*Thomas More College Catherine.Sherron@thomasmore.edu*

**Abstract:** In this paper I explore a connection between artificial intelligence (AI) generally and a particular AI project called Cyc. I sketch out ties between epistemological issues about kinds of knowledge and concerns about how minority voices (with respect to class, race, gender or similar factors) might be marginalized. I take the goal of analyzing AI projects for systematic underrepresentation or exclusion of minority views to be at the heart of feminist theorizing. Importantly, the failure to adequately consider such minority views within AI projects is ultimately detrimental to the projects themselves. There are means, though, to defend this project against some of the criticisms

## 1. ARTIFICIAL INTELLIGENCE

There are several ways to define artificial intelligence. Russell and Norvig's recent introductory AI text defines artificial intelligence as the study and creation of systems that think or behave rationally or ones that think or behave like humans [12, p.4-5]. Lest one dismiss this definition as idiosyncratic, note that the authors explicitly mention that they have surveyed eight other introductory textbooks in the field in developing their definition [p.4-5]. When artificial intelligence is defined as the study of building intelligent or rational agents, the terms of the research implicitly favor mental aspects: artificial *intelligence*. Embodied knowledge, *i.e.*, knowledge from the body or skilled knowledge, on the one hand is opposed to propositional knowledge, *i.e.*, knowledge that can be formalized by writing it down for analysis, on the other. Consider what is revealed in this quote: "Humankind has given itself the scientific name **homo sapiens**

[sic]—man the wise—because our mental capacities are so important to our everyday lives and our sense of self [p.3].” The propositional knowledge versus embodied knowledge dichotomy is thus established right away in the beginning of the text. Our *mental* capacities are important to our *everyday* lives. Embodied knowledge, however, is suspiciously absent as something necessary even to daily activities. This approach assumes that these two kinds of knowledge are separable, that we can study mental capacities without regard to embodiment. Doug Lenat’s Cyc, a knowledge-engineering project located in Austin, TX, USA, at Cycorp, is an example of a particular AI project that assumes this split. See [www.cycorp.com](http://www.cycorp.com) for more information.

## 2. THE CYC PROJECT

Expert systems (ES) such as a medical diagnosing program encode the knowledge of a human expert. Unlike humans, however, expert systems are often restricted in their scope, able to work only in narrow domains. A narrow scope makes the system *brittle*, but is necessary because of the sheer amount of information involved in even the most specialized area. A program trying to cover too much ground would be extremely unwieldy and too difficult to code. Moreover, an ES say, for chemical analysis, cannot offer assistance in medical diagnosing because no ES can reason about things outside of its range of knowledge, even relatively simple things that persons regularly take for granted. Brittleness in a system is not just an annoyance; it can lead to disaster. Lenat and Guha use a humorous example to make this point about the real problem that brittleness poses. When a skin disease diagnosing system is given the description of a rusty old car, it concludes that the car has measles [10, p.2]. Attempting to overcome brittleness by formalizing *common sense* is the goal of Cyc. Common sense is viewed as some kind of substratum that serves to facilitate more complex reasoning by providing the basic or obvious information needed to make bridges and connections from one program or domain to another. The rationale behind formalizing common sense is that the information contained in the substratum can help find or make connections between information contained in expert systems and other kinds of programs, for instance word processors, spreadsheets, and Internet providers. The bridge makes more information more easily accessible to the user and allows the computer to do more work such as more sophisticated searching and automatic updating. Overcoming brittleness would likely ease the application of expertise to a wider range of problems, since systems could then be combined to solve complex problems.

The background beliefs and assumptions behind this AI project are not limited to *Cyc*, although I will limit my remarks here to *Cyc* only. Such background beliefs are found throughout AI generally and are particularly harmful because of their transparency or invisibility to those who construct, those who use, and those that are viewed stereotypically by the system. In the same way that language can be harmful when, for example, racist assumptions are embedded within it, these AI assumptions ‘infect’ the programs with which they are connected. Adam describes it this way: “*Cyc*’s models of the world are hegemonic models—unconsciously reflecting the views of those in powerful, privileged positions”[1, p.86]. An example of how invisible assumptions can colour our views follows.

Bloom [3] argues that even though *National Geographic* claimed to be an unbiased source of information and pictures of peoples around the world, in fact the magazine both intentionally and unintentionally perpetuated racism. Her thesis is that ideas about science and technology work “in tandem with the trope of racial otherness to preserve a static and homogeneous notion of whiteness” [3, p.86]. For example, when ‘we’ photograph ‘them’ or when Americans bring new technology to the ‘locals’ and they, the locals, are dark-skinned, then we, the Americans, being different than them, must be light-skinned. The distinction of ‘other’ is very subtle, yet very powerful and the longer it goes unnoticed, the longer its reign of influence. Analogously, *Cyc* risks perpetuating certain views about the world, views that are hard to challenge since they are hidden within the program. The inferencing goes on in the background, so assumptions used by the programmers might never be noticed. *Cyc*’s common sense might very well ‘believe’ certain stereotypical ideas about women, gender, sexual orientation, etc., and then make inferences based on those ‘beliefs’. Without a strong challenge a homogenizing effect occurs, solidifying the original stereotype among users of the program.

My characterization could be overstating the case. As just another computer program could it have a widespread impact on society?

Technology has revolutionized the way we live in the same way that the industrial revolution did. According to the *ABC Nightly News*, Alan Greenspan believes that growth in technologies is in large part the driving force behind the strong American economy (January 14, 2000). Computer software is clearly a component of the technology industry. If Lenat has his way, *Cyc* would be installed on every new computer in much the same way that Microsoft Windows is pre-loaded on a large number of computers.

It is hoped that *Cyc* will be of sufficient value that there will be millions and eventually billions of copies of *Cyc* in use, the same way that there are millions of copies of Microsoft Word [11].

If Cyc were to gain that kind of exposure, its influence would be tremendous. Moody [11] observes that Lenat has “plans that might ultimately affect just about anyone who interacts with computers in the course of their mundane existence—which means, of course, just about everyone.”

### 3. THE (COMMON SENSE) WORLD ACCORDING TO CYC

The conceptual foundation of Cyc suffers from several interrelated problems. First, it posits the existence of a knowable, unified common sense. Second, it disguises a particular common sense as *the* universal Common Sense. By doing so it creates a homogenising effect on how such groups are viewed by virtue of perpetuating stereotypes. Third, by relying on a single common sense, Cyc is implicitly encouraging ignorance of other common senses and their potentially valuable contributions.

It is not clear that common sense can even be fully known. What is it and how would we know if we had it? Moreover, disguising a particular understanding of common sense as a universally applicable common sense is disingenuous at best and destructive at worst. It ‘disappears’ gender as well as class, race, and other potentially important factors by subsuming all of them under a single heading. Adam [1] cites Lenat to illustrate this problem:

Just as the consensual knowledge, also used by both Adam and Lenat to refer to common sense, itself is to be taken for granted, so it seems are those who possess such knowledge, “be they a professor, a waitress, a six-year old child, or even a lawyer” (p.85).

She goes on to say that “Given such a variety of type of subject, even within one culture, it would not be difficult to argue for different views of consensual knowledge for each of these subjects” (p.85). Cyc ignores the minority views, the quieter voices, and allows the majority voice to speak for everyone. This is not just putting forth a neutral unified common sense, but a *particular* view as the important one to consider. It should be of concern to feminists that these other views are being subsumed. A related problem is that the common sense of a particular group for example the Cyc programmers, might hold biased views about members of another group such as the very poor. The dissemination of such views further reinforces them in the same way that the embedded racist assumptions in the *National Geographic* example above furthered racist views.

Furthermore, the guise of unification overlooks potentially valuable contributions to the general common sense pool. For example, views of the very poor are not included since that body of common sense knowledge is

not useful to the Cyc project. The bottom line is that poor people are not going to buy the product. Embodied knowledge is not included either; presumably it is not as important as 'real' knowledge that is propositional knowledge. Perhaps these other views could be excluded if the project were billed differently: 'A-common-sense-for-materially-comfortable-business-people-who-use-often-use-computers-to-help-them-solve-problems.' Perhaps 'Cyc common sense' is just short for that long description, or perhaps the project leadership just does not consider those other views worthy.

What is at stake in not representing these quiet voices? Is there any real harm perpetuated? Some examples should clarify my concern. A system that only incorporates technological medical knowledge and excluded knowledge of traditional practices implies that the traditional practices and beliefs are not worth formalising. In doing this—picking out some information as worth and discarding others—the programmers are subtly shaping the identities of themselves and potential users. Some knowledge comes to qualify as truth and not just superstition, mere belief, old wives' tales, myth or folk belief [1, 7]. Another example of the power of programmers focuses on a medical diagnosing expert system. For a programmer to assume that the users who are health professionals and the subjects, the patients, are similar to the programmers themselves can create serious problems. As discussed by Boom, Chavez-Oest, and Boom [4], the inclusion of local patient data in a medical ES is important in obtaining correct diagnoses. If patient data is understood as universal character, with no local variation, there would be no reason to be concerned about including it. The programmers have the choice to include local data or to make the user aware of the importance of incorporating local data, but first they must recognise that this power lies in their hands.

To return to Cyc, note that not every belief can be included as common sense; some of our knowledge is specialised and so is not common, some is not knowledge but is trivial or information, or opinion. When such information/knowledge is formalised, someone or some group is deciding what to include under the label 'common sense.' Power is embedded in the programming process. This power is invisible, however, and that makes it hard to show its potential for harm. Common sense is ubiquitous, yet simultaneously extremely hard to define. Those facts make it difficult to show that certain knowledge is or is not definitively important to include in the Cyc project. Again, the worry is that unfairly biased views about a variety of characteristics of persons will become established as part of common sense, in the same way that racism became established within *National Geographic* or that data in an ES can reflect the programmers' own situation or characteristics.

Cyc, and, analogously, a large proportion of AI, is particularly ignorant of embodied knowledge (E-knowledge) in deference to propositional knowledge (P-knowledge). To elaborate on the definition stated at the beginning of this paper, embodied knowledge is closely related to skilled knowledge, knowledge that arises as the result of practising a skill or task until one can do it 'without thinking.' This definition makes it possible for E-knowledge and P-knowledge to be separate. One starts with propositional knowledge, *i.e.*, instructions that can be written down such that the knowledge is fully transferable to the reader. As one develops the skill one moves away from those rules. Another sense of E-knowledge is that which is the fundamental basis of all knowledge, that by which all other knowledge is gained and interpreted. This is the sense employed by Adam and others [1,2, 5, 8, 9]. Note that not all these authors use these terms in exactly the same way. For the purposes of this paper, embodied knowledge and skilled knowledge are treated together, denoted under 'embodied knowledge.' Glossing over the fact that not all knowledge is propositional in character negatively impacts the project and any programs incorporating the Cyc module. It prevents the incorporation of embodied knowledge. In turn, this makes Cyc vulnerable to gaps in its knowledge since certain information sources are neglected. Second, by excluding embodied knowledge, Cyc potentially limits the contributions of some women and some minorities, since historically women and working class people tend to develop embodied knowledge more so than propositional knowledge.

Is there a way to keep from dichotomising embodied and propositional knowledge within Cyc? Is it the case that embodied knowledge is foundational to propositional knowledge as Adam [1] and Collins [5], following the lead of Dreyfus [8], claim? If this epistemological priority did exist, Cyc would have no chance of being successful because it is simply missing a fundamental component in its knowledge construction. Is it true? Is Cyc doomed if it neglects embodied knowledge? The fact that E-knowledge *is not even considered* for inclusion in the common sense project is problematic in itself, yet I think that the criticism is overstated. Cyc is a tool, not an artificial thinker. Cyc is not creating new knowledge; it is reorganising and analysing the knowledge that is entered by persons, so it does not need to have the actual embodied experience in order to 'have' knowledge. It might find new connections or find more connections at a faster rate than a person, but that is not the same as thinking on its own.

#### **4. RESCUING CYC**

Cyc's goal is to fill in the underlying strata that humans rely on to figure out problems and increase their knowledge. This would reduce brittleness, allowing expert systems to do more work for us. Including a diverse set of views increases the scope of the knowledge base. Importantly it offers more chances to reduce brittleness by finding connections among a variety of information sources. For example, cross checking a medical problem against a diverse set of common sense perspectives about disease, treatments, causes, etc., could facilitate the discovery of unexpected solutions. In contrast, relying on a mainstream, unified common sense might overlook a non-standard solution. If we continue to think of common sense as unified then there will be no justification or any motivation to search beyond the standard views for alternative explanations. So, if the goal of Cyc is to reduce brittleness, more information not less is required. Thus, Cyc should include more voices, especially under-represented ones. The basic or obvious information needed to make bridges and connections from one program or domain to another plausibly includes the common sense of minorities. In addition, the practice of disclosing the "location" (background beliefs and characteristics) of the programmers should continue. This allows solutions to be traced back to their component beliefs and ideas for testing against the original problem specifications. This encourages responsibility on the part of the programmers and allows users to perform careful checks by hand should the need arise. In fact, Cyc is constructed in a way to handle just this method.

#### **5. MICROTHEORIES**

There is a solution to the problem of using a unified common sense and to the problem of using multiple common senses that contradict each other; isolate them within their own little worlds called microtheories. Cyc permits, and even expects, its knowledge to be structured into lattices of reified, formalized contexts. These microtheories "can represent different points of view, levels of granularity, cultural differences, age differences, time periods, corporate cultures, etc." [12]. In addition, the feminist-friendly practice of disclosing the location of programmers should continue. This allows inferences to be traced back to their component beliefs and ideas, which allows them to be tested against the original problem specifications. A microtheory would be tagged according to its developer. Also, this encourages responsibility on the part of the programmers, allows the development process to predict possible conflicts or unusual readings, and

can warn users about such conflicts, permitting users to perform checks by hand by going back to the source should the need arise. Although Cyc does not currently include multiple perspectives or embodied knowledge, if these deficiencies were straightforwardly acknowledged, they could be included at some later date without damaging the structure of the program. It is only when the deficiencies are not acknowledged and thus not addressed, that we should conclude that Cyc is detrimental to minorities.

## REFERENCES

1. Adam, A. (1998). *Artificial Knowing: Gender and the Thinking Machine*. New York: Routledge.
2. \_\_\_\_\_ (1995). Embodying Knowledge: A Feminist Critique of Artificial Intelligence. *The European Journal of Women's Studies*, 2, 355-377.
3. Bloom, L. (1994). "Constructing Whiteness: Popular Science and National Geographic in the Age of Multiculturalism." *Configurations*, 1, 15-32.
4. Boom, R. A., Chavez-Oest, J. A., and Boom, T. (1995). "The Importance of Local Data Bases in Medical Expert Systems: TICITL." In , Greens, R.A. et al, (Eds.), MEDINFO 95 Proceedings
5. Collins, H. M. (1990). *Artificial Experts: Social Knowledge and Intelligent Machines*. Cambridge, MA: The MIT Press.
6. Cycorp homepage. ([www.cyc.com](http://www.cyc.com)), (15 November 1999).
7. Dalmiya, V. and Alcoff, L. (1993). "Are 'Old Wives' Tales' Justified?" In Alcoff, L. and Potter, E. (Eds.). *Feminist Epistemologies*.
8. Dreyfus, H. (1992). *What Computers Still Can't Do: A Critique of Artificial Reason*. Cambridge, MA: The MIT Press.
9. Lakoff, G. and Johnsonm M. (1980). *Metaphors We Live By*, (2nd edition). Chicago: University of Chicago Press.
10. Lenat, D. and Guha, R. V. (1990). *Building Large Knowledge-Based Systems: Representation and Inference in the Cyc Project*. New York: Addison-Wesley.
11. Moody, S. (24 December, 1999). The Brain Behind Cyc. The Austin Chronicle. ([www.auschron.com/issues/dispatch/1999-12-23/screens\\_feature.html](http://www.auschron.com/issues/dispatch/1999-12-23/screens_feature.html)).
12. Russell, S. and Norvig, P. (1995). *Artificial Intelligence: A Modern Approach*. Upper Saddle River, NJ: Prentice-Hall.