of mind that closely links to brain functions (e.g., O'Reilly & Munakata 2000). The two approaches are distinct in that symbols simply do not exist at a subsymbolic level. A unified ToM needs to encompass both levels of description, though each may be embodied in separate cognitive architectures. We regard the attempt to vertically stretch one level of analysis to linearly map to another as problematic. For example, we doubt that there is such a simple one-to-one mapping between ACT-R components and brain structures, as suggested in Figure 1 of the target article. It is hard to imagine (and not supported by neuroscience evidence) that the damage to the basal ganglia would completely destroy the work of mind, given the fundamental role that production rules play in ACT-R.

In summary, although we agree with Newell and A&L that a unified ToM is needed to advance cognitive science, we have different opinions regarding how to achieve such a unified theory. Our position is that, instead of hybridizing different approaches or linearly mapping them to boost the Newell Test score, we need to recognize the multilevel nature of the human mind and develop complementary theories at both psychological and connectionist levels, and cross-validate them.

NOTE

1. Hongbin Wang is the corresponding author for this commentary.

Newell's program, like Hilbert's, is dead; let's move on

Yingrui Yang[a] and Selmer Bringsjord[b]

[a]*Department of Cognitive Science;* [b]*Department of Computer Science, Rensselaer Polytechnic Institute, Troy, NY 12180.* **yangyri@rpi.edu selmer@rpi.edu http://www.rpi.edu/~brings**

Abstract: We draw an analogy between Hilbert's program (HP) for mathematics and Newell's program (NP) for cognitive modeling. The analogy reveals that NP, like HP before it, is fundamentally flawed. The only alternative is a program anchored by an admission that cognition is more than computation.

As most readers will know, Hilbert's program (HP) was devoted to building a system at or below the level of Turing machines (and their equivalents) to definitively settle all mathematical questions. Most readers will also know that in 1931, a young Viennese logician, Kurt Gödel, proved two incompleteness theorems – and Hilbert's program (HP) instantly died. Out of this death was born a much more sophisticated approach: In a word, true *meta* mathematics arose. One specific innovation was to devise and study infinitary logics not bound by Gödelian incompleteness, because, from an information-processing perspective, these logics are *beyond* Turing machines (Barwise 1980). Newell's program (NP) – the attempt to build a system at or below the level of Turing machines able to satisfy the criteria that Anderson & Lebiere (A&L) distill for us – has likewise expired. The difference is that apparently many influential cognitive scientists want to pretend the funeral never happened.

A&L, and in fact all those who see themselves as advancing NP, presuppose an exceedingly convenient sense of *universality.* According to this sense, a system is "universal" if and only if it can compute all Turing-computable functions. The construal is convenient because the vast majority of functions (in the relevant classes; e.g., functions from N to N) *aren't* Turing-computable (Boolos & Jeffrey 1989). A&L, the classical connectionists they evaluate, those they count as fans (e.g., Dennett), and so forth – all *assume* that human cognition can be nicely packaged beneath the Turing Limit. Yet, after decades of work, no system at or below the Turing Limit has the conversational power of a toddler (to pick just one criterion: 7). You would think the notion Newell (and his similarly sanguine partner, Simon) so confidently preached at the dawn of cognitive science (that thinking is computing at or below the Turing Limit, and computers operating at or below this limit with human-level intelligence will soon arrive) would, like Hilbert's dream, be a carcass at this point, but yet here is a *BBS* target article still stubbornly clinging to the religion (by cheerfully acting as if everyone is a believer). What arguments support the doctrine that cognition can be captured by standard computation? Surely no cogent argument is to be found on the basis of what has been built. LOGIC THEORIST, at the 1956 Dartmouth conference that kicked off AI, was able to prove the marvelously subtle theorem that *if p then q* implies *if not-q then not-p,* and this prompted Simon to declare that thinking machines would soon be among us. The situation is no different now: Here are A&L confidently pressing on to capture cognition in simple computation – but on the strength of what impressive artifact? Since seeing is believing, you will pardon us for not believing.

The problem isn't just criterion 7. Faced with consciousness, NP irremediably fails. Yet A&L once again cook up the convenient: They construe *consciousness* (in criterion 8) so that it simply leaves out the concept that threatens Newell's enterprise: namely, *phenomenal* consciousness. Block (1995) has recently explained the issue in this very journal. ACT-R and all forms of connectionism, and indeed every approach to sustaining NP, can't even take the first step toward expressing, in a third-person scheme, what it feels like to taste deep chocolate ice cream. ACT-R will be used to at most create what one of us (Bringsjord 2000) has called "zombanimals," that is, artificial animals with no inner lives. A robot with the behavioral repertoire of a dog, but with the inner life of a rock, might well be something NP, fueled by ACT-R, can produce.

That NP, as driven specifically by ACT-R, is dead, can be seen with help from concrete, not just philosophical, challenges. ACT-R is wholly incapable of modeling beliefs of the sort that human readers have when reading murder mysteries. For example, as detailed in Bringsjord (2000), readers have n-order beliefs about villains and detectives, and they make inferences based on these beliefs. For example, the reader of a murder mystery often believes that the villain believes that the detective believes that the villain believes the villain will never be caught. You can't represent this in ACT-R, period, because ACT-R is at best part of first-order logic devoid of doxastic operators. Of course, one could hook up a robust theory of human and machine reasoning (e.g., see Yang & Bringsjord, forthcoming) to ACT-R, but then in what sense *is* that new composite system ACT-R? If ACT-R is to be genuine science, it must be falsifiable. Yet A&L seem to describe an evolution in which serious challenges are handled by simply augmenting the system.

Just as the death of HP gave birth to infinitary logic, so should the death of NP give rise to cognitive modeling untrammeled by standard computation. Cognitive modelers need to step outside the notion that mere computation will suffice. They must face up to the fact, first, that the human mind encompasses not just the ordinary, humble computation that Newell and all his followers can't see beyond, but also *hyper*computation: information processing at a level *above* Turing machines, a level that can be formalized with help from analog chaotic neural nets, trial-and-error machines, Zeus machines, and the like (Bringsjord & Zenzen 2003; Siegelmann 1999).

In his famous "twenty questions" paper, Newell (1973) tells us that a sound science of the mind should not be steered by the willy-nilly dictates of experiment-driven empiricism. Instead, we are to do computational cognitive modeling. But such modeling, if limited by NP, fails to let cold hard reality lead the way. Instead, it lets speculative assumptions (e.g., that the mind processes information at or below the Turing Limit) prevent nature from proclaiming that we are more than ordinary machines.

https://doi.org/10.1017/S0140525X03440132 Published online by Cambridge University Press