Toward a . . .

Serious Computational Science of Intelligence

Call for Papers for an AGI 2010 Workshop
in Lugano, Switzerland
on March 8 2010

Selmer Bringsjord *&* **Naveen Sundar G**
Department of Computer Science
Department of Cognitive Science
Rensselaer Polytechnic Institute (RPI)
Troy NY 12180 USA
`selmer@rpi.edu` • `govinn@cs.rpi.edu`

version 10.22.09

Contents

1 AI and AGI, Following the Norm, is an Ecumenical Tent

Truth be told, nearly all disciplines are vast tents under which radically different modes of operation co-exist, sometimes peacefully, sometimes not, and most frequently somewhere in between. For example, some physicists proceed via the axiomatic approach, and their pondering e.g. the possibility of time travel consists in determining whether some set of formulae are consistent with a formalization of such travel. Other physicists regard such activity to be "mere" logico-mathematics, not physics, given that by their lights physics is distinguished by *empirical* investigation. Additional examples could be effortlessly multiplied: Philosophy as a field subsumes analytic philosophy and much of formal logic, and at the same time the "fuzzy" continental tradition. Economics ranges across sweeping historical narratives considered in political economics, and also across formal, game-theoretic analyses of idealized situations. Computer science contains formal-methods folks who bemoan the prevalence of software not formally verified — yet the field is also home to those who see it as an experimental one. And so on.

AI isn't any different: it too is an ecumenical tent. This is confirmed not only by the co-existence now of both traditional AI and AGI, but by the fact that there are many much more fine-grained methodological divisions within both. Some thinkers, for example, are rather passionate about a logic-based approach to AI and cognitive science (e.g. Bringsjord 2008*b*, Bringsjord 2008*a*); others are in favor of a non-symbolic, continuous approach (e.g. Spivey 2006). And ... some people are in favor of a *serious computational science of intelligence* ($\mathcal{SCSI}$), whereas others are not. We propose a workshop devoted to explicitly considering — and, frankly, promoting — this science, which we describe in the next section.

2 What is the Serious Computational Science of Intelligence?

We confess to lacking a formal definition of the science we wish to have explicitly discussed in the prospective workshop. But we claim that five attributes, if embodied in a given research program, qualify that program as an example of $\mathcal{SCSI}$.

In presenting this quintet, we assume our audience to have command over computation, from a logico-mathematical perspective; and we specifically use 'computation' in the broad sense that covers not only Turing machine-level computation, but (suitably formalized) hypercomputation.[1] We also assume that our readers have a sufficiently firm grasp of the concept of intelligence, so that at least they can recognize it when they see it. So, with you we know that neurobiologically normal human persons are undeniably intelligent; that apes are as well, but much less so; that dogs are intelligent, but not as smart as apes; that rodents are even dimmer still; that while a case can be made for classifying viruses as intelligent, the level of intelligence they possess is exceedingly slim (but nonetheless exceedingly dangerous); and that it's easy enough to imagine beings markedly more intelligent than any of those on the continuum just sketched.

With this pair of assumptions noted, here's the key quintet.

- **Constructive**? In order for this property to be instantiated, it must be the case that the research/system in question is explicitly designed to produce determinate computational theories with considerable reach (which may or may not be implemented). It's not enough for the researcher to be merely critical, however rigorous and far-reaching the criticisms might be.

[1] In short, standard Turing machines compute (for coverage see e.g. Lewis & Papadimitriou 1981), and so do infinite-time Turing machines (which are defined in Hamkins & Lewis 2000). Please note that in order to qualify in our scorecard, one doesn't have to *embrace* hypercomputation, but a formal orientation (which as will soon be seen is a *sine qua non* for the science we have in mind) requires *familiarity* with computation, broadly understood.

- **Theorem-guided**? In order for this property to apply, the work in question must be significantly guided by theorems, secured on the strength of formalization. Note that the property here entails much more than that some formalism is used. Of course, not all the theorems in question need to be generated by the work in question: It will sometimes be helpful if the work is classified in light of established theorems.
- **General**? The research program and/or system must be intended to cover both human *and* machine intelligence. This means, for example, that computational cognitive scientists, if aiming to merely understand human intelligence, aren't sufficiently ambitious to qualify. (And as a matter of fact, such scientists generally don't pursue work that has the previous attribute — which is noted in Bringsjord 2008*a*.)
- **Continuum**? To possess this property, the research in question must include the concept of a continuum of intelligence, from dim to human-level to beyond us to *well* beyond us. (Such a continuum is broadly discussed in Bringsjord et al. 2000.) One shortcut to satisfying the **Continuum** property is to insist that candidate creatures have their intelligence explicitly tested to see just how intelligent they are. (One fascinating test-based approach to such matters, at least when is thinking about creatures with intelligence at the level of bright humans, is Floridi 2005.)
- **Cohesive**? The fifth and final requirement is that the work in question be devoted to promoting a cohesive vision or theory. Working on frameworks for capturing *pieces* of intelligence is insufficient. There must be a steadfast search for a unifying account of the numerous aspects of intelligence.

3 Who's in the Game? A Scorecard

We have found it to be fascinating to consider who is and isn't in the game. It turns out that shockingly few people are pursuing the serious computational science of intelligence. (Of course, we could be unaware of some qualifying researchers — and, needless to say, stand ready to be educated.) In fact, we haven't located anyone who obviously satisfies all five requirements, though it seems to us that Hutter (2005) comes closest. Running a near second is the work of the late John Pollock (1995, 2001), specifically work related to his Oscar system. In addition, Schmidhuber's (2009, 2006) work on so-called "Gödel machines" would appear to be worthy of serious consideration. Note that the work of Hutter and Schmidhuber is marked by an emphasis on learning (in fact, the latter explicitly affirms the former's non-knowledge-based learning-based account of universal AI in Hutter 2005), and the conspicuous absence of declarative knowledge, while in Pollock's case the opposite is true (though Pollock explicitly expects others to work on sub-knowledge-based learning). At least so far, our analysis of $\mathcal{SCSI}$ is steadfastly *meta*-analysis, and we offer no opinions at this stage as to whether the detailed content of R&D in the game is good or bad (save for the self-criticism articulated in the next paragraph).

Note that at least heretofore the work of Bringsjord fails to earn five affirmatives, and we don't even include him in the scorecard. Bringsjord's deductive arguments against "strong" AI are critical (e.g., Bringsjord & Zenzen 1997), and his constructive work (e.g. in the area of computational creativity; see Bringsjord & Ferrucci 2000) isn't intended to provide a unifying account of all of intelligence. The same holds for systems, for example Slate (Bringsjord et al. 2008), designed to reflect all of human *reasoning* — since after all there is presumably more to intelligence than reasoning. Furthermore, while he has articulated a fairly general view of the human mind (Bringsjord & Zenzen 2003), this view doesn't in its current form reach the standards required by $\mathcal{SCSI}$.

The scorecard given immediately below contains an upper and a lower section. We have reserved the latter section for those in the computational cognitive modeling (CCM) game. This appears to be a *sub*-game, for reasons indicated by relevant occurrences of 'No' and 'Maybe' in this section of the scorecard itself. (However, note the grades earned by NARS.) But the issue of the relationship between (CCM) and $\mathcal{SCSI}$ is inevitably one that will need to be addressed, and is clearly vital, so we essentially build it into the discussion. We fully expect that the scorecard will be filled out with grades for other seminal figures in CCM beyond those we mention. These figures have their own cognitive architectures. Needless to say, we are not entirely confident that the grades we have assigned are correct in all cases; they are defeasible.

Researcher/System	**Constructive**?	**Theorem-guided**?	**General**?	**Continuum**?	**Cohesive**?
Hutter (*UAI*)	Yes	Yes	Yes	Yes	Yes
Pollock (1995, 2001) (Oscar)	Yes	Yes	Yes	No	Yes
Russell (1997)	Maybe	Yes	Maybe	Yes	No
Schmidhuber (2009, 2006)	Yes	Yes	Yes	Maybe	Maybe
Wang (2006) (NARS)	Yes	Yes	Yes	Maybe	Yes
Cassimatis et al. forthcoming (Polyscheme)	Yes	No	Yes	Maybe	Yes
Sun (2001) (Clarion)	Yes	No	Maybe	Maybe	Yes
Newell & Simon (1997) Laird (1993) (Soar)	Yes	No	Yes	No	Maybe
Anderson et al. (2003) (ACT-R)	Yes	No	No	No	Maybe
Langley et al. (2006, 2005, 2004)	Yes	No	No	No	Maybe

4 Questions Addressed by Papers for the Workshop

Papers (1–2 pages for a position statement; 6–12 for a full paper; agi-10 format must be used) should be submitted to either Bringsjord or Sundar G via email by the deadline. Papers must address one or more of the following question groups. It is anticipated that arising from the workshop there will be a special issue of the *Journal of Artificial General Intelligence* devoted to the serious computational science of intelligence. The special issue would draw from a pool of papers composed of not only those submitted for the workshop, but also of papers written outside workshop activity.

Q1 What theories and/or research projects or programs should be developed (or have been developed) for a sound and complete $\mathcal{SCSI}$?

Q2 Should the quintet of conditions be expanded to include **Implementation**? Are other modifications to the requirements desired?

Q3 Why is the serious computational science of intelligence important? Why doesn't more of AI, AGI, and Cog Sci (and, for that matter, economics, decision theory, etc.) R&D fall under this science?

Q4 How good is the work of those currently clearly in the $\mathcal{SCSI}$ game? (This question, of course, given the foregoing, cashes out as evaluation, first, of Hutter and Pollock.)

Q5 What is the future of $\mathcal{SCSI}$? Is it bright or dark, or somewhere in between? (Authors may want to at least consider to some degree the economic side of this question, in light of the brute fact that formal-methods-based computer science arguably hasn't gone swimmingly of late because of money issues.)

⋮

References

Anderson, J. & Lebiere, C. (2003), 'The newell test for a theory of cognition', *Behavioral and Brain Sciences* **26**, 587–640.

Bringsjord, S. (2008*a*), Declarative/logic-based cognitive modeling, *in* R. Sun, ed., 'The Handbook of Computational Psychology', Cambridge University Press, Cambridge, UK, pp. 127–169.
URL: *http://kryten.mm.rpi.edu/sb_lccm_ab-toc_031607.pdf*

Bringsjord, S. (2008*b*), 'The logicist manifesto: At long last let logic-based AI become a field unto itself', *Journal of Applied Logic* **6**(4), 502–525.
URL: *http://kryten.mm.rpi.edu/SB_LAI_Manifesto_091808.pdf*

Bringsjord, S. & Ferrucci, D. (2000), *Artificial Intelligence and Literary Creativity: Inside the Mind of Brutus, a Storytelling Machine*, Lawrence Erlbaum, Mahwah, NJ.

Bringsjord, S., Noel, R. & Caporale, C. (2000), 'Animals, zombanimals, and the total Turing test: The essence of artificial intelligence', *Journal of Logic, Language, and Information* **9**, 397–418.
URL: *http://kryten.mm.rpi.edu/zombanimals.pdf*

Bringsjord, S., Taylor, J., Shilliday, A., Clark, M. & Arkoudas, K. (2008), Slate: An Argument-Centered Intelligent Assistant to Human Reasoners, *in* F. Grasso, N. Green, R. Kibble & C. Reed, eds, 'Proceedings of the 8th International Workshop on Computational Models of Natural Argument (CMNA 8)', Patras, Greece, pp. 1–10.
URL: *http://kryten.mm.rpi.edu/Bringsjord_etal_Slate_cmna_crc_061708.pdf*

Bringsjord, S. & Zenzen, M. (1997), 'Cognition is not computation: The argument from irreversibility?', *Synthese* **113**, 285–320.

Bringsjord, S. & Zenzen, M. (2003), *Superminds: People Harness Hypercomputation, and More*, Kluwer Academic Publishers, Dordrecht, The Netherlands.

Cassimatis, N., Bignoli, P., Bugajska, M., Dugas, S., Murugesan, A. & Bello, P. (forthcoming), 'An architecture for adaptive algorithmic hybrids', *IEEE Transactions on Systems, Man, and Cybernetics (Part B)* .

Choi, D., Kaufman, M., Langley, P., Nejati, N. & Shapiro, D. (2004), An architecture for persistent reactive behavior, *in* 'Proceedings of the Third International Joint Conference on Autonomous Agents and Multiagent Systems-Volume 2', IEEE Computer Society Washington, DC, USA, pp. 988–995.

Choi, D. & Langley, P. (2005), Learning teleoreactive logic programs from problem solving, *in* 'Proceedings of the Fifteenth International Conference on Inductive Logic Programming', Springer, pp. 51–68.

Eisenstadt, S. & Simon, H. (1997), 'Logic and thought', *Minds and Machines* **7**(3), 365–385.

Floridi, L. (2005), 'Consciousness, agents and the knowledge game', *Minds and Machines* **15**(3-4), 415–444.
URL: *http://www.philosophyofinformation.net/publications/pdf/caatkg.pdf*

Hamkins, J. D. & Lewis, A. (2000), 'Infinite time Turing machines', *Journal of Symbolic Logic* **65**(2), 567–604.

Hutter, M. (2005), *Universal Artificial Intelligence: Sequential Decisions Based on Algorithmic Probability*, Springer, New York, NY.

Langley, P. (2006), 'Cognitive architectures and general intelligent systems', *AI Magazine* **27**(2), 33–44.

Lewis, H. & Papadimitriou, C. (1981), *Elements of the Theory of Computation*, Prentice Hall, Englewood Cliffs, NJ.

Pollock, J. (1995), *Cognitive Carpentry: A Blueprint for How to Build a Person*, MIT Press, Cambridge, MA.

Pollock, J. (2001), 'Defasible reasoning with variable degrees of justification', *Artificial Intelligence* **133**, 233–282.

Rosenbloom, P., Laird, J. & Newell, A., eds (1993), *The Soar Papers: Research on Integrated Intelligence*, MIT Press, Cambridge, MA.

Russell, S. (1997), 'Rationality and Intelligence', *Artificial Intelligence* **94**(1-2), 57–77.

Schmidhüber, J. (2006), Gödel machines: Fully self-referential optimal universal self-improvers, *in* B. Goertzel & C. Pennachin, eds, 'Artificial General Intelligence', Springer, New York, NY, pp. 119–226.

Schmidhüber, J. (2009), 'Ultimate cognition à la Gödel', *Cognitive Computation* **1**(2), 177–193.

Spivey, M. (2006), *The Continuity of Mind*, Oxford University Press, Oxford, UK.

Sun, R. (2001), *Duality of the Mind*, Lawrence Erlbaum Associates, Mahwah, NJ.

Wang, P. (2006), The logic of intelligence, *in* B. Goertzel & C. Pennachin, eds, 'Artificial General Intelligence', Springer, New York, NY, pp. 31–62.