Given the Web, What is Intelligence, Really?*

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

April 9, 2012

Abstract

We argue that existing systems on the Web cannot approach human-level intelligence, as envisioned by Descartes, without being able to achieve genuine problem solving on unseen problems. We argue how this entails committing to a strong intensional logic. In addition to revising extant arguments in favor of intensional systems, we present a novel mathematical argument to show why extensional systems can never hope to capture the inherent complexity of natural language. The argument makes its case by focusing on representing, with increasing degrees of complexity, knowledge in a first-order language. Nevertheless, the attempts at representation fail at achieving consistency, making the case for an intensional representation system for natural language clear.
Keywords: Human-level intelligence, extensional logic, intensional logic, knowledge representation, modality.

1 The Web Extended and Immediate: True Intelligence?

Let's assume that vast declarative information covering nearly all human collective knowledge, courtesy of the Semantic Web a decade hence,[1] enables a flawless version of what is known as **QA** technology.[2] And we add a second assumption: namely, that the QA cycle is mediated by direct brain-Web interfacing. Under this foreseeable-in-the-near-future assumption, if Smith is verbally asked a question q, he can internally and mentally ask it of the Semantic Web, receive back an answer a immediately to his neo-cortex, and convey a as required (e.g., by vocalizing the answer to the interlocutor before him). If Smith could do this today, surreptitiously, he would certainly cause most questioners to believe that he is rather intelligent. He would for example be able to: say which of Shakespeare's plays contain any given snippet of the Bard's immortal narratives (e.g., "What did devout Harry's former friend steal, and pay for with his life?"), answer any question about any settled part of science (e.g., "What was Frege's quirky notation for what is essentially

*The two authors are profoundly grateful for support from the John Templeton Foundation that has made possible the research reported on herein. We are also indebted to two anonymous reviewers for insightful comments and objections, and to the editors for sagacious guidance.

[1]Heretofore in the present paper, we use the terms 'Semantic Web' and 'Web' interchangeably.

[2]Undeniably, the best QA technology in the world is currently the Watson system, created by IBM to compete against humans in the long-running American television game of *Jeopardy!*, which is essentially itself a QA game. For a description of the system, see (Ferrucci, Brown, Chu-Carroll, Fan, Gondek, Kalyanpur, Lally, Murdock, Nyberg, Prager, Schlaefer & Welty 2010). Given Watson's prowess, the future we envision could very well soon arrive.

modern first-order logic?"), produce the notes in sequence for Bach's Organ Mass (and throw in a word-for-word verbalization of Luther's Catechism), and so on. If no one knew about Smith's hidden, wireless, brain-to-Web link, again, he certainly would be regarded intelligent — probably even positively brilliant, at least by many.

2 Cartesian Skepticism

By many, but not by all. What would a truly discriminating judge say? We're afraid that on accounts of *real* intelligence of the sort that Descartes had in mind, your secret Web link would be insufficient. Why? Because Descartes would know that a mere mechanical machine could in principle do just what you *qua* QA master are doing. Thus, by considering whether the Web, given current trends, will fundamentally alter the very concept of human intelligence, we find ourselves carried back to the longstanding debate about whether human (or human-*level*) intelligence can be captured in mechanical form. Descartes answered this question in the negative. Long before Turing, he claimed that only the human has *domain-independent* (and conversational) intelligence, and that therefore certain tests would be exceedingly difficult, if not outright impossible, for machines to pass. He specifically suggested two tests to use in order to separate mere machines from human persons. The first of these directly anticipates the so-called "Turing Test." The second test is the one that connects directly to domain-independent intelligence. We read:

> If there were machines which bore a resemblance to our body and imitated our actions as far as it was morally possible to do so, we should always have two very certain tests by which to recognise that, for all that, they were not real men. The first is, that they could never use speech or other signs as we do when placing our thoughts on record for the benefit of others. For we can easily understand a machine's being constituted so that it can utter words, and even emit some responses to action on it of a corporeal kind, which brings about a change in its organs; for instance, if it is touched in a particular part it may ask what we wish to say to it; if in another part it may exclaim that it is being hurt, and so on. But it never happens that it arranges its speech in various ways, in order to reply appropriately to everything that may be said in its presence, as even the lowest type of man can do. And the second difference is, that although machines can perform certain things as well as or perhaps better than any of us can do, they infallibly fall short in others, by which means we may discover that they did not act from knowledge, but only for the disposition of their organs. For while reason is a universal instrument which can serve for all contingencies, these organs have need of some special adaptation for every particular action. From this it follows that it is morally impossible that there should be sufficient diversity in any machine to allow it to act in all the events of life in the same way as our reason causes us to act. (Descartes 1911, p. 116)

A test of domain-independent intelligence requires success on topics with which the agent has had no prior experience. In our thought-experiment, by definition, you look smart specifically because you are really you *plus the Web*, and the Web gives you access to mountains of *pre-established* information — but not to "knowledge" in Descartes's sense of the word here. What he means by knowledge in this context is a capacity for general-purpose, problem-solving knowledge.

To follow Descartes in testing for real intelligence we must present you with a problem that you have never seen before, and wait to see whether you can provide a solution by means that you invent on the spot.[3] We hold that while the Web and and associated cognitive technologies promise to make possible the fake brilliance of Smith, the Web will not provide the "universal instrument" that resides within us. Hence the Web won't bring us any closer to the flexible, general intelligence that Descartes correctly claimed would continue to separate minds from mere machines.

3 The Missing Science of Human+Web Intelligence

Even those who disagree with us must admit two things: one, our position has *some* force; and two, this very fact, combined at the same time with reluctance to wholeheartedly affirm our position, implies that the

[3] A series of simple yet remarkably subtle tests of this form, for children and adolescents, are provided by Piaget (Inhelder & Piaget 1958).

Figure 1: Using the Web to Find Declarative Knowledge

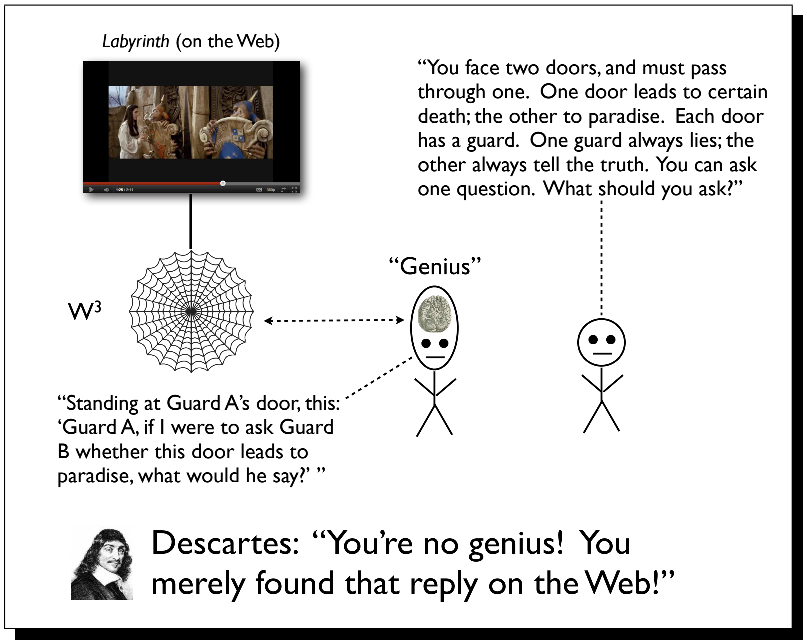
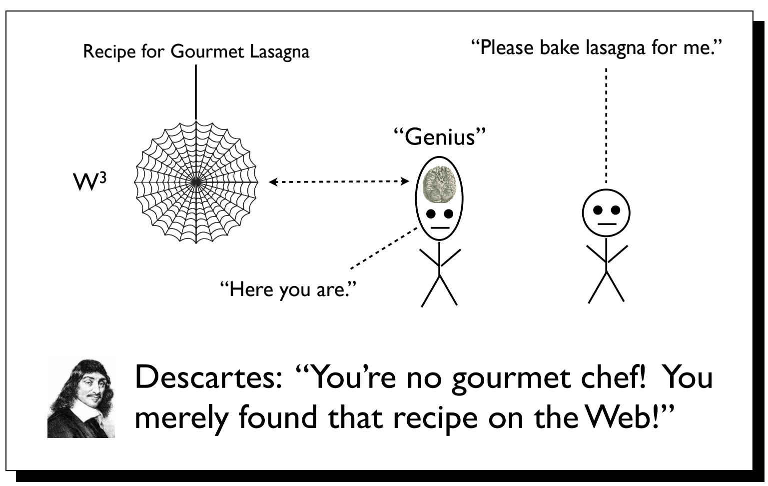


Figure 2: Using the Web to find Imperative Knowledge



matters before us are fundamentally unsettled; or to put that point another way: there simply isn't currently available a rigorous science of human-intelligence-augmented-by-the-Web (H+W). Addressing the absence of such a science is non-trivial, as we soon proceed to show. Encapsulating here: Today, all knowledge represented formally and hence amenable to automated reasoning on the Semantic Web is extensional; but human reasoning is extensional *and* intensional.[4] The problem stems from a lack of formal computational

[4]The terms "extensional" and "intensional" derive from logic. Put barbarically, extensional logics are concerned merely with the *references* of terms while intensional logics are concerned with both the *sense* and reference of a term. In concrete practice, the student of mathematical logic learns first the simplest extensional logics (propositional calculus, first-order logic, e.g.), and will only be confronted with the need to invoke intensional operators when taught *philosophical* logic. A nice overview of the latter class of logical systems is provided in (Goble 2001).

systems that can handle intensional reasoning as robustly as they can handle extensional reasoning. The formal and computational sciences are in the very early stages of understanding such systems, and here we briefly explore some of the initial moves in a research program aimed at producing a formal science of H+W.

4 A Desideratum for a Science of H+W: Modeling Knowledge

Most human reasoning is highly intensional, in that it involves belief, desire, knowledge, action, intention, perception, communication, and so on. This is in opposition to standard reasoning recorded for instance in mathematics or the natural sciences; such reasoning is extensional. Consider the following two statements:

P_1 : My infant niece does not know that the area of a circle with radius r is $\pi * r^2$

P_2 : The area of a circle with radius r is $\pi * r^2$

The first statement is intensional, the second is extensional. Why do we need to worry about statements of the first type in order to erect a rigorous science of H+W?

Consider this scenario:

Jack is married to Jill, and in preparation for their anniversary, he **communicates** his **intention** of planning a surprise dinner for Jill — through the BCI chip implanted in his brain — to his personal assistant Simon in the **hope** that Simon will take care of his **planning**. Jack also **believes** that, even though he has not **communicated** to Simon that Jill should not know of Jack's **intentions**, Simon **knows** that this should not be made **known** to Jill until the evening before the dinner.

Even this perfectly ordinary scenario demonstrates an *explosion of intensionality* that no current theory is capable of handling — despite impressive progress in AI, and in web-powered QA and web-powered personal assistants. Also, note that this task includes what could be general-purpose planning, still an unsolved problem. Further complicating the matter is the presence of indexical terms: *this*, *he*, *his*, *their*, etc.

All attempts at formalizing reasoning on the Web have so far dealt only with shallow and non-general reasoning over extensional data, reasoning woefully short of being able to handle the above scenario.[5] What is necessary to rectify this? The answer is clear: a system able to deal with intensionality, ranging across all intensional operators that appear in scenarios like the one above involving Jack and Jill. While classical mathematical logic has been applied successfully to extensional domains, the situation for the intensional case is not so bright, as we show below. This state-of-affairs imperils any serious attempt at understanding, scientifically, Web-enhanced human intelligence.

We present three possible ways of tackling just one kind of intensionality, one that takes us back to the starting parable of Smith: *declarative knowledge possessed by an agent*; and we show how these ways all lack robustness. We seek to lend credence to an approach based on intensional modal operators, but are happy to see a future unfold in which intensional operators in English are attacked in any reasonable manner.

4.1 A First Attempt

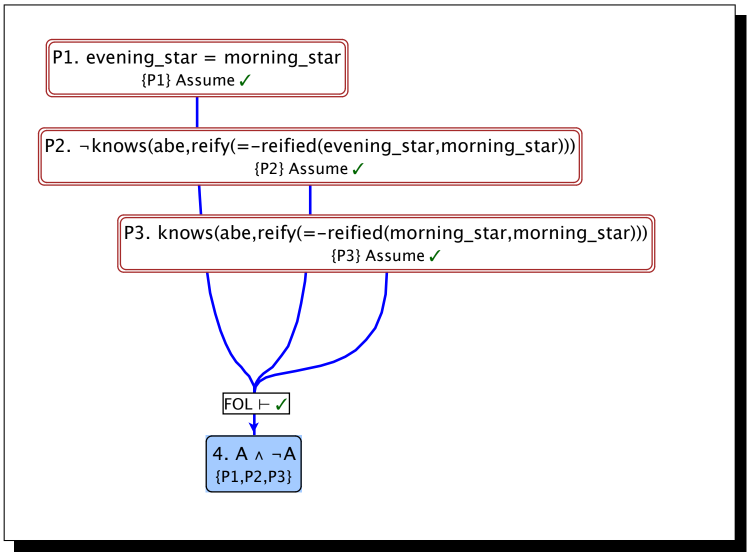
An initial attempt at modeling declarative knowledge might proceed to do so by incorporating a knowledge *predicate* directly into first-order logic. One way to go about this is by mirroring all predicate symbols with another set of new function symbols. For instance, a predicate symbol P gets mirrored by the function symbol f_P; this process is called *reification* or, more aptly given how it is often implemented (e.g., see Russell & Norvig 2002), *stringification.*

Unfortunately, this approach is doomed from the start. Consider a small and commonly known illustration in Figure 3, implemented in the Slate proof-engineering environment (Bringsjord, Taylor, Shilliday, Clark & Arkoudas 2008). Our hypothetical person, Abe, is quite ignorant of modern astronomy and doesn't know that the morning star is the very same object as the evening star. They are indeed the same object: the

[5] In fact, though we don't discuss it, the Semantic Web is currently quite firmly tied to not only extensional logics, but low-expressivity ones, e.g., decription logics. Such logics are covered in (Baader, Calvanese & McGuinness 2007).

planet Venus. This is a fact which was unknown to many of our ancestors; they saw the planet at dawn and dusk and believed they were different objects. Any honest modeling formalism must respect this state-of-affairs, but alas, we find that as a soon as we enter the relevant propositions into Slate we encounter a contradiction. We still have a problem if it was unknown to us that Abe does not know that the morning star is in fact the evening star: our inference becomes unsound.

Figure 3: An Attempt at Modeling Knowledge in a First-order Language



4.2 A Second Attempt

One way to escape from the contradiction and unsoundness pointed to by the above proof is to devise a knowledge predicate **K** that applies to the Gödel numbers of formulae, which blocks certain unwanted inferences, namely substitution of co-denoting terms (e.g *morning_star* with *evening_star*).

This discussion follows that laid out by Anderson (Anderson 1983). The Knower Paradox is an epistemic paradox which stems from assuming certain obvious premises about knowability. A semi-automated version of the paradox is shown in Figure 4. The paradox is arrived at as follows: Consider the well-known system Q of Robinson Arithmetic (see (Boolos, Burgess & Jeffrey 2003)) along with a predicate **K** the extension of which is supposed to be the Gödel numbers of all the sentences of Q that are known to some knower. We are also given another predicate, **I**, the extension of which is set of the ordered pairs of Gödel numbers $(\ulcorner\phi\urcorner, \ulcorner\psi\urcorner)$, such that ψ is derivable assuming ϕ in Q.[6] Now, one could reasonably suppose the following axioms for the new predicates:

Ax_1 $\mathbf{K}(\overline{\phi}) \rightarrow \phi$

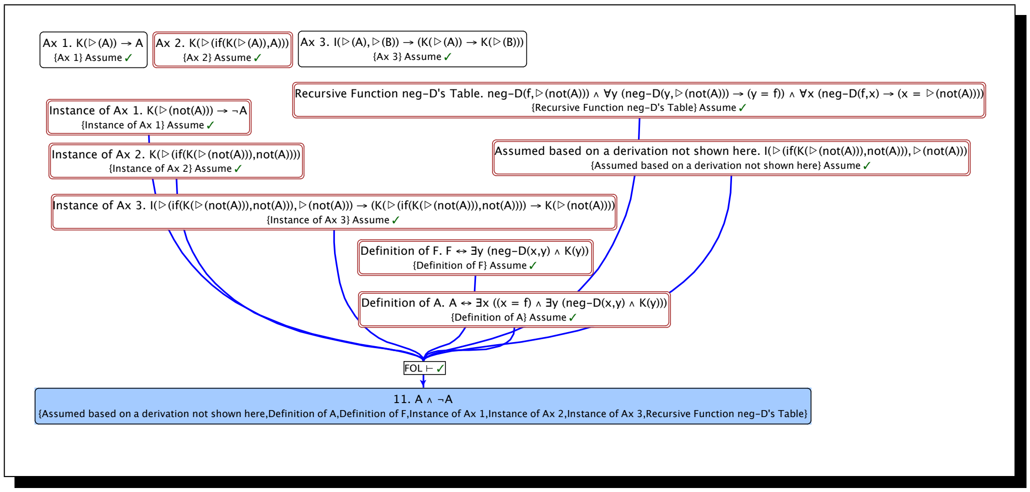
Ax_2 $\mathbf{K}(\overline{\mathbf{K}(\overline{\phi}) \rightarrow \phi})$

Ax_3 $\mathbf{I}(\overline{\phi}, \bar{\psi}) \rightarrow (\mathbf{K}(\overline{\phi}) \rightarrow \mathbf{K}(\overline{\psi}))$

Unfortunately, contradiction still rears its head, as shown in the semi-automated Slate proof in Figure 4. Note that in this proof, Gödel numerals are represented with a $\triangleright(\phi)$ instead of the conventional $\overline{\phi}$.

[6] We use $\ulcorner\phi\urcorner$ to denote the Gödel number of ϕ and the idiomatic notation $\overline{\phi}$ to denote the Gödel numeral for ϕ, instead of the more compositionally correct notation $\overline{\ulcorner\phi\urcorner}$.

Figure 4: The Knower Paradox Semi-Automated



4.3 A Third Attempt

There are many ways by which one could avoid the above paradox, but it seems that all of them are less than satisfactory. One could argue that iterated knowledge predicates don't always have truth values, or that knowledge need not be monotonic (i.e., that old knowledge can be destroyed when new knowledge is acquired). These two suggestions, while bearing the virtue of simplicity, don't accord with our pre-analytic notions of knowledge. Another more attractive route, suggested by Anderson, is to dissect the predicates $\mathbf{K}$ and $\mathbf{I}$ into a hierarchy of predicates: $\mathbf{K}_0, \mathbf{K}_1, \ldots$, for knowledge; and $\mathbf{I}_0, \mathbf{I}_1, \ldots$, for inference. We obtain the language dubbed L_ω when we augment the first-order language L of arithmetic with these hierarchies of $\mathbf{K}$ and $\mathbf{I}$ predicates. The levels of predicates are supposed to mirror an agent iteratively reflecting upon its knowledge. In this approach, the interpretation V assigns an extension to all the predicates and is also decomposed into a series of interpretations, V_0, V_1, V_2, such that V_i assigns extensions only to $\mathbf{K}_i$ and $\mathbf{I}_i$, and $V = \cup_i V_i$. V_0 extends some interpretation V_p for L to furnish extensions for $\mathbf{K}_0$ and $\mathbf{I}_0$. The following metalogical conditions are then imposed by Anderson:[7]

(i) Knowledge is preserved when reflecting deeper:

$V_i(\mathbf{K}_i) \subseteq V_{i+1}(\mathbf{K}_{i+1})$

(ii) Inferences are preserved:

$V_i(\mathbf{I}_i) \subseteq V_{i+1}(\mathbf{I}_{i+1})$

(iii) Only true things can be known:

$\ulcorner\phi\urcorner \in V_i(\mathbf{K}_i) \Rightarrow V_j(\phi) = T$ for a $j \geq i$

(iv) Only correct inferences can be made:

$(\ulcorner\phi\urcorner, \ulcorner\psi\urcorner) \in V_i(\mathbf{I}_i) \Rightarrow V_j(\phi \rightarrow \psi) = T$ for a $j \geq i$

(v) If something can be inferred from what is known then that is known:

$(\ulcorner\phi\urcorner, \ulcorner\psi\urcorner) \in V_i(\mathbf{I}_i)$ and $\ulcorner\phi\urcorner \in V_i(\mathbf{K}_i) \Rightarrow \ulcorner\psi\urcorner \in V_i(\mathbf{K}_i)$

Interpretations V that can be decomposed into a hierarchy of interpretations, $V_0, V_1, \ldots$, and satisfying the above conditions, are called *coherent*. A sentence of L_ω is called *κ-valid* iff it is true in every coherent interpretation.

[7] Note $V(\phi) = T$ is shorthand for $V \vDash \phi$ if ϕ is a sentence, and $V(P)$ denotes the extension of P if P is predicate symbol.

Now, we ask: Is κ-validity semi-decidable? Note that κ-validity is just first-order validity strengthened by conditions of coherence on the interpretations. First-order validity is of course semi-decidable: If a sentence is valid in FOL, then there is a partial-recursive function p such that $p(x) = 1$ iff x is the Gödel number of a valid first-order sentence (e.g. see (Boolos & Jeffrey 1989)). The function p can be *represented* in Q;[8] this is so because Q is strong enough to represent all the partial-recursive functions. Let us add a new predicate $\mathbf{P}$ to Q to represent first-order validity; that is, $Q \vdash \mathbf{P}(\overline{\phi})$ iff $\{\} \vdash \phi$. Since we are dealing with a provability predicate, we posit the following axioms, assuming that we are concerned about an ideal mathematician who knows all the theorems of Q at the first level of reflection $\mathbf{K}_0$.

A_1 : If it is known that the validity of something is proven, then that is known:

$$\mathbf{K}_0(\overline{\mathbf{P}(\overline{\phi})}) \rightarrow \mathbf{K}_0(\overline{\phi})$$

A_2 : An instance of Ax_3: (Q_{conj} is the conjunction of all the axioms in Q.)

$$\mathbf{I}(\overline{Q_{conj}}, \overline{\mathbf{P}(\overline{\phi})}) \rightarrow (\mathbf{K}_0(\overline{Q_{conj}}) \rightarrow \mathbf{K}_0(\overline{\mathbf{P}(\overline{\phi})}))$$

A_3 : Q_{conj} is known:

$$\mathbf{K}_0(\overline{Q_{conj}})$$

A_4 : All κ-valid sentences ϕ can have their validity proved in Q:

$$\mathbf{I}(\overline{Q_{conj}}, \overline{\mathbf{P}(\overline{\phi})})$$

We can see that from $\{A_1, A_2, A_3, A_4\}$ we have for all κ-valid sentences ϕ that $\mathbf{K}_0(\overline{\phi})$. We also see that $V_0(\mathbf{K}_0(\overline{\phi})) = T$ for all such ϕ, since the predicate $\mathbf{K}_0$ is assigned its full extension by V_0. We have the following theorem by Anderson:

Theorem (No i-perfect logician)**.** *No coherent interpretation has an i such that for all κ-valid sentences ϕ, $V_i(\mathbf{K}_i(\overline{\phi})) = T$.*

Note that for $i = 0$, the theorem contradicts our result. Our more sophisticated attempts at avoiding contradiction seem to have failed.

Since we have assumed an ideal mathematician in this argument, one could counter that the contradiction obtained is not as strong as the ones obtained before, since — so the objection goes — ideal mathematicians have "obviously" not been observed to exist. Unfortunately, this rejoinder hardly surmounts the problem: The mere *possibility* of the existence of such a mathematician engenders a contradiction, and the mere possibility of such a contradiction is a serious problem for a serious science of knowledge in the context of H+W.

5 Some Objections; Our Responses

We now present and rebut some of the objections that might be raised against us.

5.1 Objection 1

Objection 1 is essentially a claim that human intelligence, conventionally understood, is not, if you will, "universally" intelligent, as it involves critical use of information/capability encoded in evolved genetic material. Here is one way the objection might be expressed:

> "You claim that 'A test of domain-independent intelligence requires success on topics with which the agent has had no prior experience.' Well, I assume that you understand that one could argue that this Cartesian requirement makes sense if one requires 'prior experience' to only apply to experience in an

[8] Briefly, a function $g : \mathbb{N} \mapsto \mathbb{N}$ is *representable* in a theory T in the language of arithmetic $L = \{\mathbf{0}; =, <; S, +, .\}$ iff there is a formula $G(x, y)$ such that whenever $g(m) = n$ we have $T \vdash \forall x\ G(m, x) \leftrightarrow x = n$. Consult (Boolos et al. 2003) for more background.

individual lifespan. However, one could argue that general purpose-learning is not done via scratch from humans, but via 'prior experience' encoded in selection of genetic material. So humans do not work in areas in which their species have absolutely no prior experience: Imagine a situation where a human is transported into a completely alien world. They would likely die, not adapt."

5.1.1 Our Response to Objection 1:

We counter by noting that it is exceedingly hard to see how human reasoning in abstract domains such as logic and mathematics could ever have been encoded in genetic material via the experience of one's ancestors, yet humans reason and solve problems successfully in these abstract domains.[9] It's also hard to see why human persons, if thrust into an alien world, wouldn't retain the power to solve problems there that would yield to the logico-mathematical techniques that work on Earth. Regardless, the *onus probandi* is surely on the critic to provide evidence rather than just an imaginative notion regarding ancestry-bound problem-solving. Note that the critic here says: "one could argue." Perhaps that is true enough; but where is the argument? We certainly agree that Objection 1 is indeed imaginative; we furthermore agree that in subsequent dialectic on the formal science of H+W the present exchange is worthy of sustained analysis; but these concessions leave the case we present herein quite intact.

5.2 Objection 2

The second objection is that the Semantic Web already uses, or at any rate will soon successfully use, higher-order and intensional logics. It can be expressed as follows:

> "First, the two of you assume that the Semantic Web is limited to very restrictive logics. This is not true. RDF actually has traditionally higher-order features, such as circular classes and names both referring to classes and properties. Second, first-order logic, and intensional logics, could be added to the Semantic Web in "higher-layers" as has been detailed in Tim Berners-Lee's roadmap to a 'universal logic.' Would simply adding an adequate intensional logic à la Zalta to the Semantic Web allow your thought-experimental Smith to achieve true intelligence, or is there a more properly philosophical argument over prior experience and general purpose-reasoning that prevents the 'W' in 'H+W' from *a priori* qualifying as intelligent?"

5.3 Our Response to Objection 2:

Though the formal systems used for the Semantic Web have features that resemble those in higher-order logic and intensional logics, these are but shallow surface-level syntactic commonalities entirely reducible to classical first-order logic; one such reduction is carried out by Fikes, McGuinness & Waldinger (2002). Even if the Semantic-Web community deploys *true* higher-order and intensional logics, the question of the automation of such systems poses an even graver problem.

There is a deeper philosophical argument underlying our reasoning. Our stance is that human persons, in reasoning in and over a some domains (at least some *abstract* domains, for example the set-theoretic universe posited in ZF), gain direct and unmediated access to the objects in these domains.[10] Machines, on the other hand, seem to be forever restricted to manipulating syntactic or symbolic structures that merely

[9] It is interesting to note that Objection 1 is similar to the justification given by Cassimatis (2006) for his fascinating *cognitive substrate hypothesis*, according to which, in nutshell form, all of our grand cognitive powers today as members of *homo sapiens sapiens* are reducible to a small core of algorithms and data structures. Presumably all trenchant objections to this hypothesis apply to Objection 1. For instance, logicians and mathematicians can grasp the truth of Goodstein's Theorem and understand its proof. (An accessible account of the theorem and its proof is given by Smith (2007).) The only proofs known of Goodstein's theorems use the ordinal numbers and properties thereof, infinitary concepts and structures which seem to be well beyond a cognitive substrate needed for such pedestrian pursuits as those that defined the lives of hunter-gatherers (Bringsjord 2001).

[10] Using Kant's terminology, one can say that humans seem to have access to "a thing-in-itself" — at least as long as the thing is purely abstract.

shadow these abstract objects.[11] Given this, we find it exceedingly hard to see how human intelligence can be genuinely augmented via interfaces with the Web. In sum, our Smith, we still maintain, is no smarter courtesy of the link from his brain to the Web than he is without it. We assume that even obdurate readers will at least concede that in light of the reasoning we have given, the rigorous science of H+W is in its infancy, and, accordingly, much work remains to be done.

6 Summing Up

Because of thought-experiments like the one given above involving Smith, the pre-scientific state of an account of the intelligence of H+W is exposed. When in response one sails out bravely in pursuit of a serious science of human-and-web intelligence (our H+W), one soon confronts the challenge of rigorously modeling intensionality in many contexts, and therefore the challenge of modeling many propositional attitudes. We have briefly demonstrated that seemingly sensible proposals for handling even just one such attitude (*knows*) are problematic. We believe the central issue is due to a category confusion: The objects of intensional attitudes such as knowing are propositions, yet all three attempts considered above ignore this brute fact. After all, our Smith may not be a genuine genius because of his brain-to-Web link, but just like us, he certainly knows many propositions. He knows, for example, that the Web houses myriad facts. In knowing this, he doesn't know a string or a number; again, he knows a *proposition.* Yet the attempts briefly canvassed above replace the target of propositional knowledge with, respectively, strings and natural numbers. As along as one fails to acknowledge explicitly that knowing-that is knowing a proposition, and resorts to a "bag-of-tricks" to squeeze the intensional into the extensional, problems will persist. If they do, not only will we fail to develop a rigorous science of H+W, but we won't any time soon even engineer a Simon-like agent capable of approaching genuine intelligence by leveraging the Web, but will instead be forced to content ourselves with only glorified digital turks capable of but fundamentally shallow feats.

References

Anderson, C. A. (1983), 'The Paradox of the Knower', *The Journal of Philosophy* **80**(6), 338–355.

Baader, F., Calvanese, D. & McGuinness, D., eds (2007), *The Description Logic Handbook: Theory, Implementation (Second Edition)*, Cambridge University Press, Cambridge, UK.

Boolos, G. S., Burgess, J. P. & Jeffrey, R. C. (2003), *Computability and Logic (Fourth Edition)*, Cambridge University Press, Cambridge, UK.

Boolos, G. S. & Jeffrey, R. C. (1989), *Computability and Logic*, Cambridge University Press, Cambridge, UK.

Bringsjord, S. (1992), *What Robots Can and Can't Be*, Kluwer, Dordrecht, The Netherlands.

Bringsjord, S. (2001), 'Are we evolved computers? A critical review of Steven Pinker's *How the Mind Works*', *Philosophical Psychology* **14**(2), 227–243. A more detailed version of this paper is available from the author, and is currently available online at `http://www.rpi.edu/~faheyj2/SB/SELPAP/PINKER/pinker.rev2.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*

[11] Alert readers will note that comparisons can be drawn between our argument and philosophical arguments against so-called "Strong" AI and machine consciousness and qualia (e.g., see those in Bringsjord 1992), but we refrain from exploring such comparisons here.

Cassimatis, N. (2006), 'Cognitive substrate for human-level intelligence', *AI Magazine* **27**(2), 71–82.

Descartes, R. (1911), *The Philosophical Works of Descartes, Volume 1. Translated by Elizabeth S. Haldane and G.R.T. Ross*, Cambridge University Press, Cambridge, UK.

Ferrucci, D., Brown, E., Chu-Carroll, J., Fan, J., Gondek, D., Kalyanpur, A., Lally, A., Murdock, W., Nyberg, E., Prager, J., Schlaefer, N. & Welty, C. (2010), 'Building Watson: An Overview of the DeepQA Project', *AI Magazine* pp. 59–79.
URL: *http://www.stanford.edu/class/cs124/AIMagzine-DeepQA.pdf*

Fikes, R., McGuinness, D. & Waldinger, R. (2002), A first-order logic semantics for semantic web markup languages, Technical Report KSL-02-01, Knowledge Systems, AI Laboratory, Stanford University.

Goble, L., ed. (2001), *The Blackwell Guide to Philosophical Logic*, Blackwell Publishing, Oxford, UK.

Inhelder, B. & Piaget, J. (1958), *The Growth of Logical Thinking from Childhood to Adolescence*, Basic Books, New York, NY.

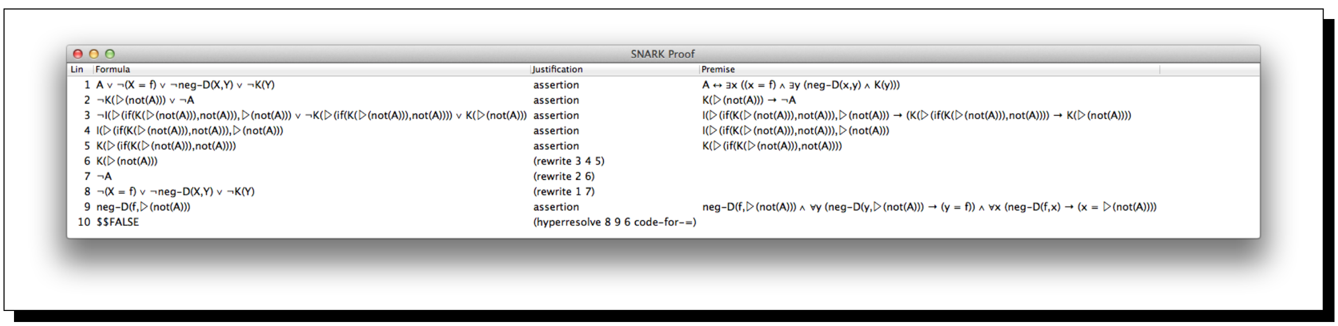
Russell, S. & Norvig, P. (2002), *Artificial Intelligence: A Modern Approach*, Prentice Hall, Upper Saddle River, NJ.

Smith, P. (2007), *An Introduction to Gödel's Theorems*, Cambridge University Press, Cambridge, UK.

Appendix

Figure 5 contains the machine proof of the Knower's Paradox proposition proved in the workspace shown in Figure 4.

Figure 5: Proof by SNARK of the Knower Paradox



SNARK Proof

Lin	Formula	Justification	Premise
1	A ∨ ¬(X = f) ∨ ¬neg-D(X,Y) ∨ ¬K(Y)	assertion	A ↔ ∃x ((x = f) ∧ ∃y (neg-D(x,y) ∧ K(y)))
2	¬K(▷(not(A))) ∨ ¬A	assertion	K(▷(not(A))) → ¬A
3	¬I(▷(if(K(▷(not(A))),not(A))),▷(not(A))) ∨ ¬K(▷(if(K(▷(not(A))),not(A)))) ∨ K(▷(not(A)))	assertion	I(▷(if(K(▷(not(A))),not(A))),▷(not(A))) → (K(▷(if(K(▷(not(A))),not(A)))) → K(▷(not(A))))
4	I(▷(if(K(▷(not(A))),not(A))),▷(not(A)))	assertion	I(▷(if(K(▷(not(A))),not(A))),▷(not(A)))
5	K(▷(if(K(▷(not(A))),not(A))))	assertion	K(▷(if(K(▷(not(A))),not(A))))
6	K(▷(not(A)))	(rewrite 3 4 5)	
7	¬A	(rewrite 2 6)	
8	¬(X = f) ∨ ¬neg-D(X,Y) ∨ ¬K(Y)	(rewrite 1 7)	
9	neg-D(f,▷(not(A)))	assertion	neg-D(f,▷(not(A))) ∧ ∀y (neg-D(y,▷(not(A))) → (y = f)) ∧ ∀x (neg-D(f,x) → (x = ▷(not(A))))
10	$$FALSE	(hyperresolve 8 9 6 code-for-=)	