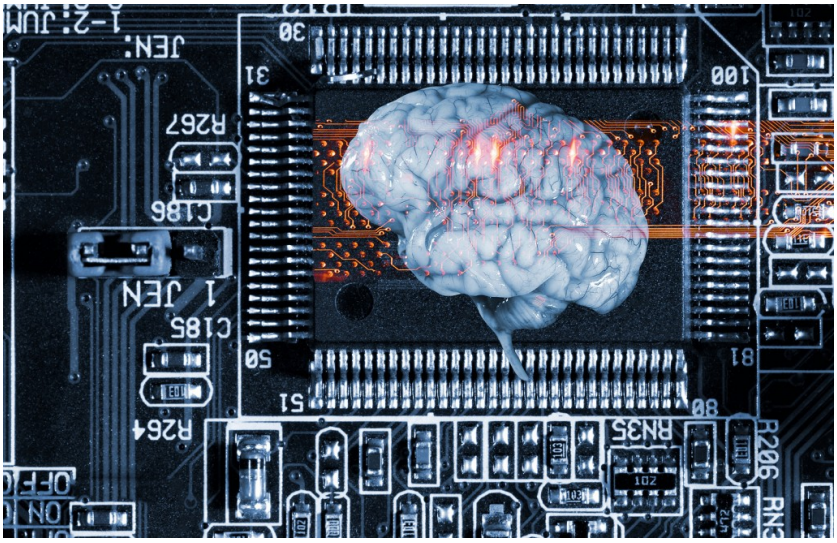
Science & Religion



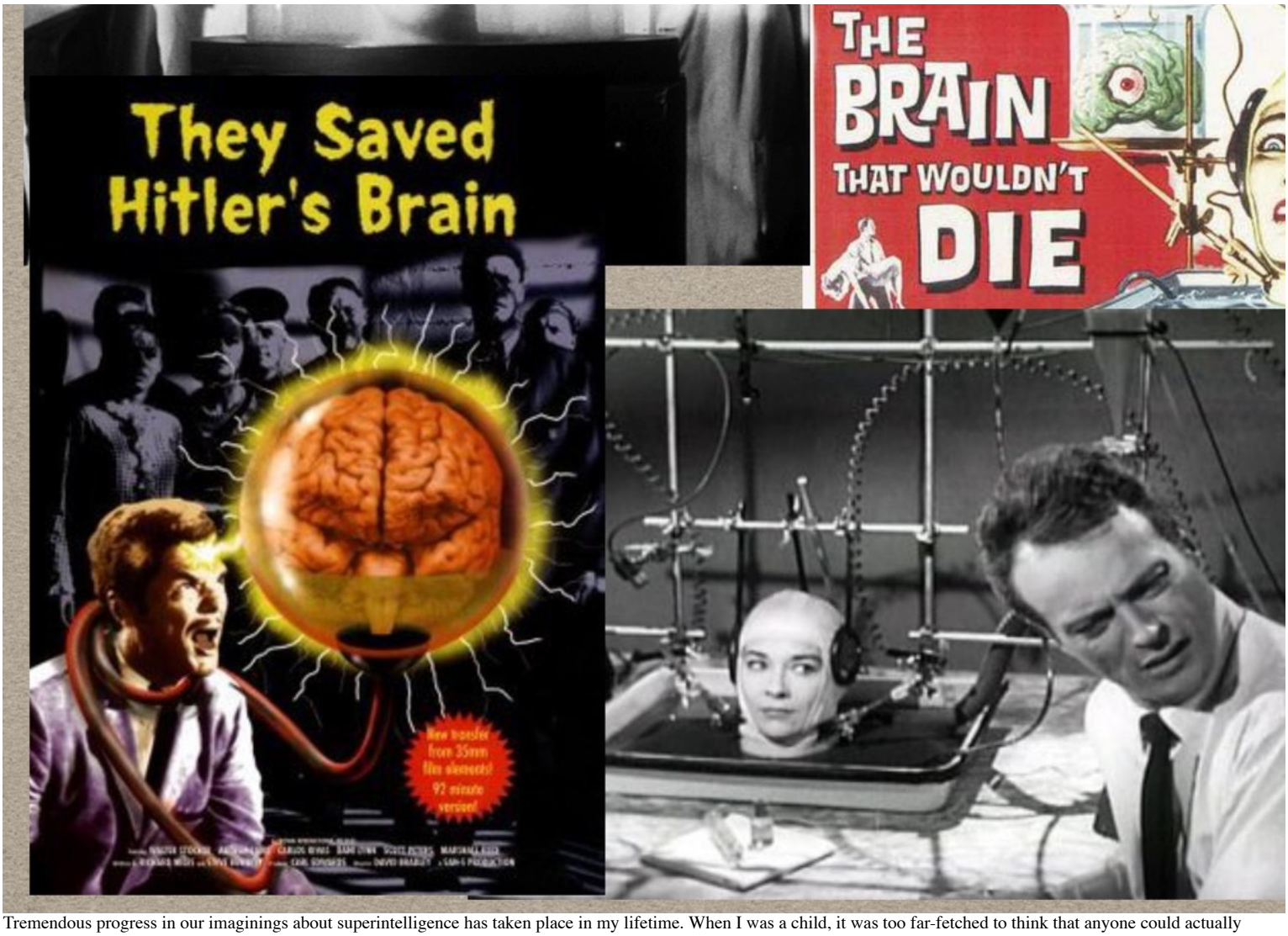
Superintelligence Through Artificial Computational "Brains"

By Jeffrey M. Bradshaw · June 20, 2016

To sign up for Meridian's Free Newsletter, please CLICK HERE.

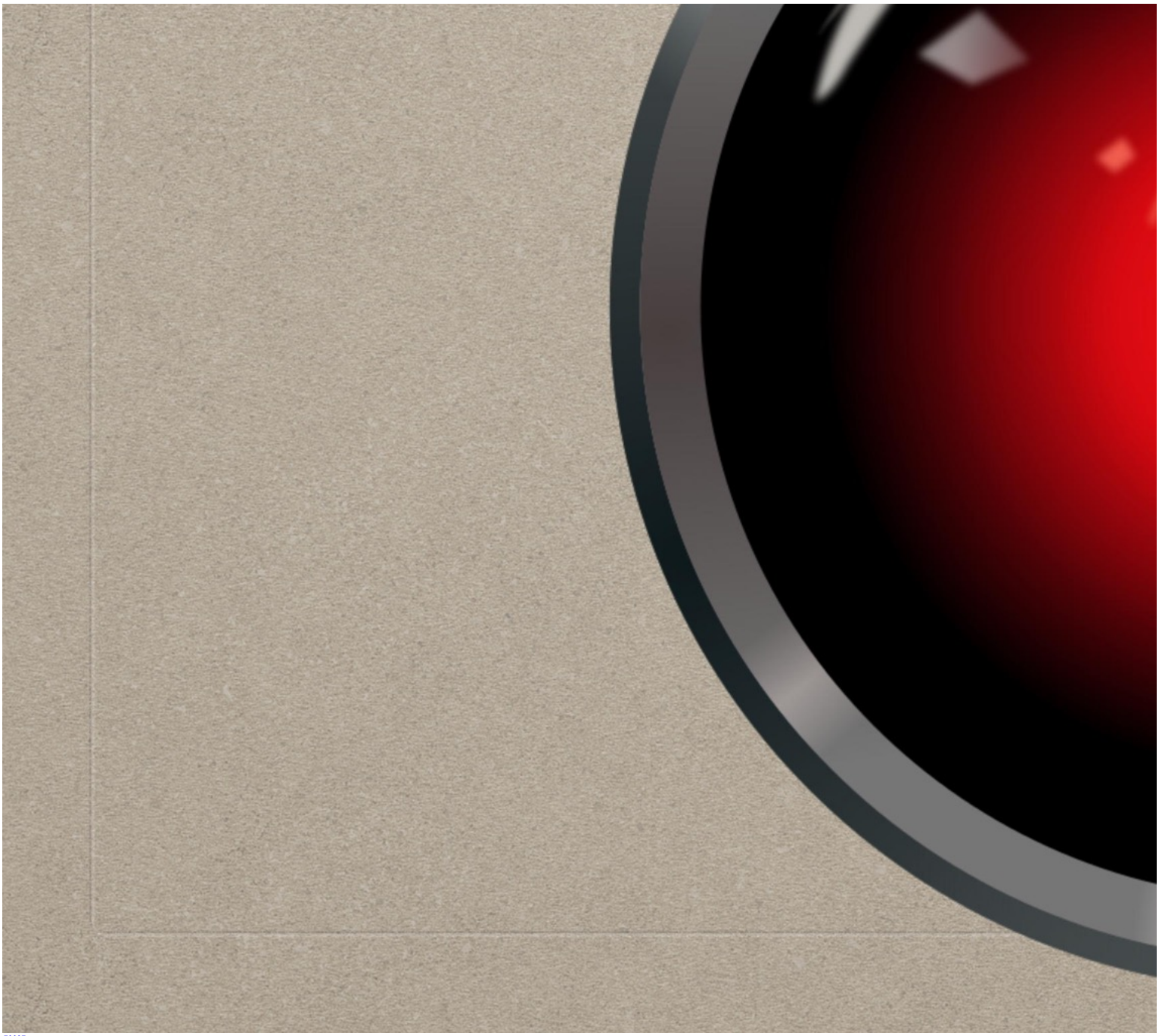
Editor's Note: The following is Part 3 in a series that expands upon a presentation given at the Second Interpreter Science and Mormonism Symposium: Body, Brain, Mind, and Spirit *at Utah Valley University in Orem, Utah, 12 March 2016. A book based on the first symposium, held in 2013, has recently been published as Bailey, David H., Jeffrey M. Bradshaw, John H. Lewis, Gregory L. Smith, and Michael L. Stark.* Science and Mormonism: Cosmos, Earth, and Man. *Orem and Salt Lake City, UT: The Interpreter Foundation and Eborn Books, 2016. For more information, including free videos of these events, see http://www.mormoninterpreter.com.*

Now I'd like to say a few words about one of the most incredible example today of the incurable optimism of researchers, namely the building of what has been termed "superintelligence."[i]



Tremendous progress in our imaginings about superintelligence has taken place in my lifetime. When I was a child, it was too far-fetched to think that anyone could actually *build* a superintelligence, so the best that science fiction could offer us was to help us imagine a *real* human brain, kept alive in a jar and tethered with wires, that was bent on either controlling or destroying the world. Thanks to the broadening of our imaginations in the computer age, we have substituted the outmoded idea of a real brain in a jar with two new and improved substitutes that have become the subject of countless blockbuster films: 1) the omniscient supercomputer, a completely artificial brain (discussed in this article, Part 3);[ii] and 2) the omniscient mind, a natural human brain that has been uploaded to a network of supercomputers (discussed in the next article in the series, Part 4). Both of these new options for superintelligence — and a few others besides — are being hotly pursued by researchers.

Figure 2



[iii]

As to the omniscient supercomputer, the current front-runner is IBM Watson, which shares conceptual genes with Arthur C. Clarke’s HAL 9000 from *2001: A Space Odyssey.* Although transposing the letters H-A-L one letter forward produces I-B-M, any deliberate connection was adamantly denied by Clarke, though later embraced by IBM.[iv]

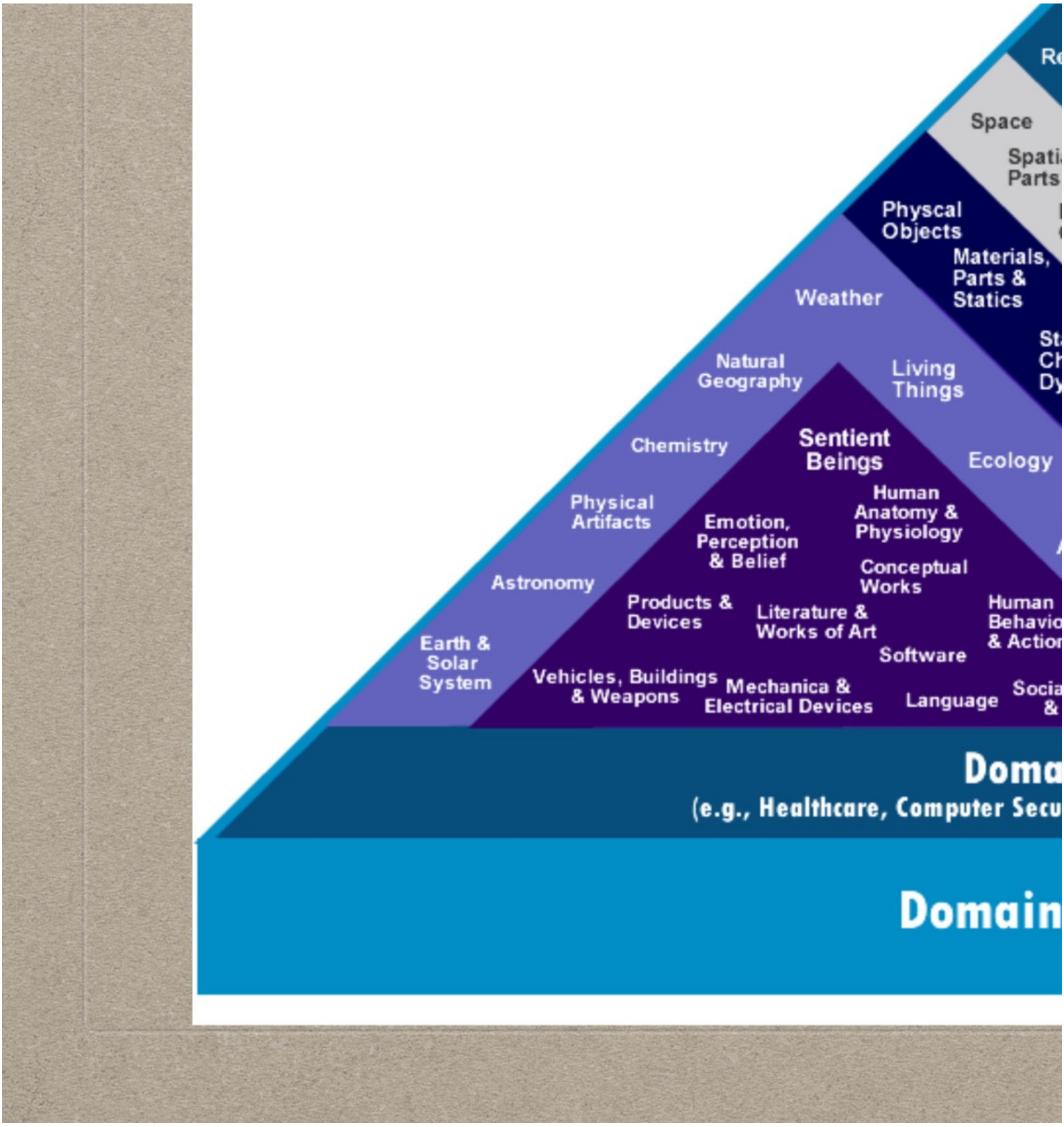
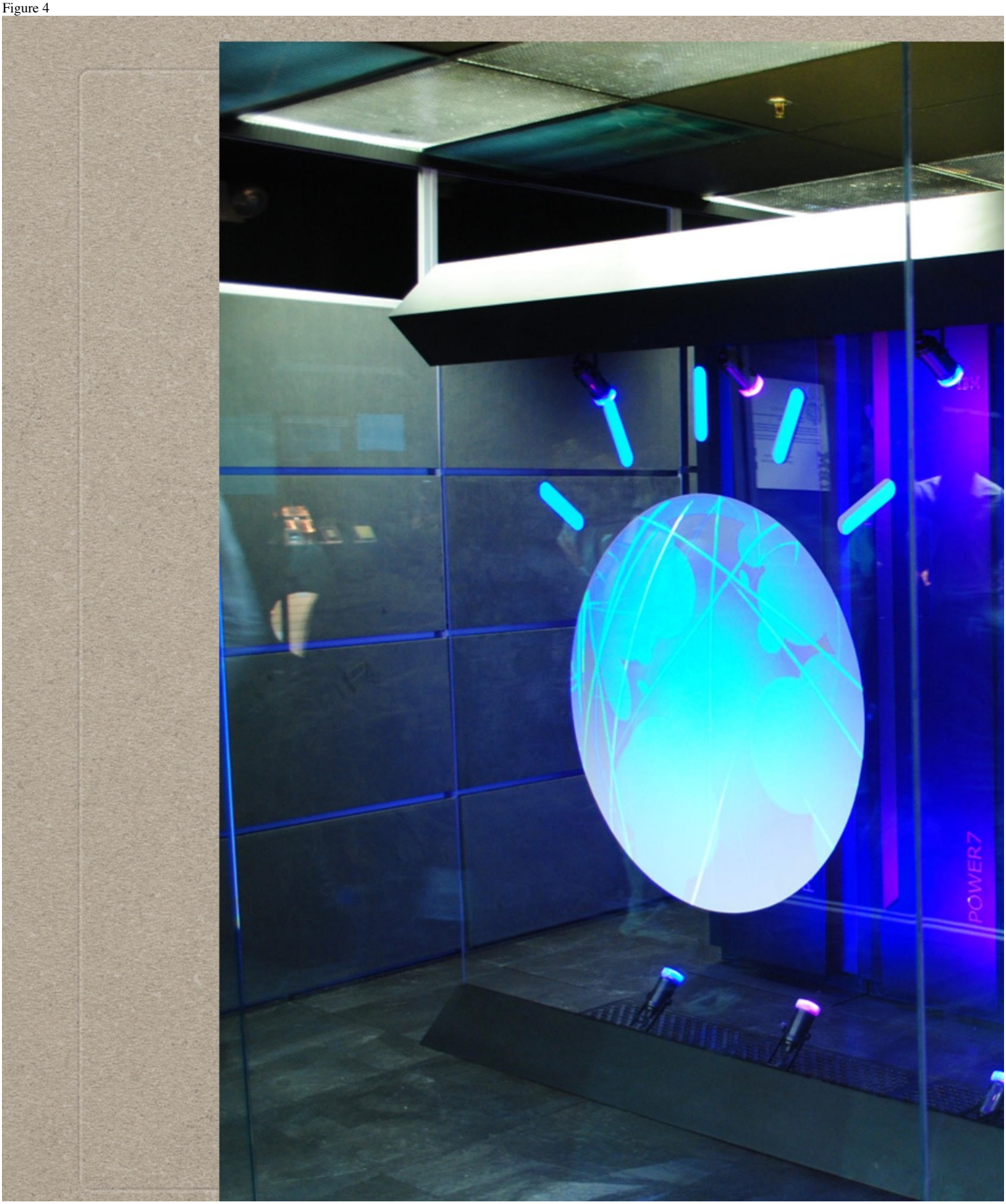


Figure 3 [v]

From a research perspective, Watson also shares conceptual genes with Doug Lenat's Cyc (as in en*cyc*lopedia), an ambitious multi-decade project to build a general purpose AI that has failed to yield the fruits its originators have always dreamed of.[vi] However, unlike the current version of Cyc, for which bits of knowledge usually have been crafted by hand, Watson has the advantage of being able to ingest large swaths of the Internet and build a knowledge base largely on its own.[vii]

Figure 4



[viii]

To the disappointment of IBM, Watson has not taken off with the speed and glory that they initially hoped since its public debut on *Jeopardy!* in 2011. Indeed, *Jeopardy!* has been the *only* application for Watson that has made much of a splash with the public. IBM's website currently lists only few dozen small company application partners, and a February 2016 article touts with unabashed optimism "future potential" much more than it parades details of its current successes over the last five years.[ix] As a super-smart search engine, a capability for complex classification or diagnosis problems, or a natural-language-based analytic assistant, it has great potential. As a superintelligence that matches the ambitions of HAL, I predict it will continue to fall short for the foreseeable future. The top researchers at IBM must already know this; though the sales and marketing folks still seem to be in denial.

Among the challenges for any machine aspiring to HAL 9000 capabilities — in addition to whatever competence it may have been designed to demonstrate in its particular area of expertise — is commonsense reasoning (e.g., "Who is taller, Prince William or his baby son Prince George? Can you make a salad out of a polyester shirt? If you stick a pin into a carrot, does it make a hole in the carrot or in the pin?"). Commonsense reasoning is a remarkable but unheralded capability that people rely on in nearly every situation but that has been recognized by AI researchers as "one critical area where progress has been extremely slow."[x]

Another underappreciated garden-variety human capability is to be able to sense and understand changes in the world or in a given situation that require adapting or abandoning the current course of action in order to respond to something more important or urgent (e.g., the problem the machine is currently working on has just been resolved by another means (or has not become irrelevant or unachievable for some reason), an earthquake has occurred, a fight has broken out, or an elephant has entered the room). Today's machines are not typically made to continually sense and understand the wide range of global and local phenomena to which people attend, let alone to make the kinds of appropriate adaptations to context that come almost naturally to humans. Moreover, since today's machines are not "aware" of the fact that the world itself is distinct from their limited and contextually impoverished model of the world — their particular "map" of the features of the world that are relevant to what they are designed to do —they must rely on humans to keep them and their models in tune with the changing dynamics of the real world in which they operate.[xi]

Pragmatics — a word that researchers usually associate with the study of language but which is just as important for studying every other kind of action — is another challenging problem for any machine that needs to be understood by people or other machines.[xii] The theoretical study of natural languages is usually divided into three areas: syntax, semantics, and pragmatics. Syntax is the study of the words of a given language and the rules that dictate how these words combine to form legal expressions (i.e., its grammar). Within "speech act theory" — often used as the basis for communication among intelligent machines — semantics and pragmatics combine to account for what the expressions of the language *mean*. For example, a sentence like "It is cold in this room" has both a syntactic analysis and a literal, semantic meaning which is constant across all of its possible uses: namely, that the temperature in the room is cold relative to the speaker.

However, pragmatics deals with the fact that the speech act that a speaker intends to perform by using this sentence depends on the context of its utterance. The sentence could be used to state a fact, request that the listener close a window, warn the listener not to enter the room, ironically state just the opposite of what the statement means (e.g., saying that the room is cold for humorous effect when the room is actually very hot), or for some other reason. In fact, natural language utterances are often used for several purposes at once — including strengthening or weakening whatever social relationships hold among those involved in the conversation (e.g., deliberately saying something in a way that will be understood one way by some of the listeners and a different way by others).

Obviously, it is much easier to design machines that are good at simply understanding the literal meaning of speech and actions than to endow them with the more human-like ability to understand the myriad implications of making a statement or performing an action in a specific situation at a given time in the presence of particular individuals. The following story illustrates both the power of that the proper exercise of knowledge about pragmatics can have in achieving desired results and the current limitations that machines have in that very respect.

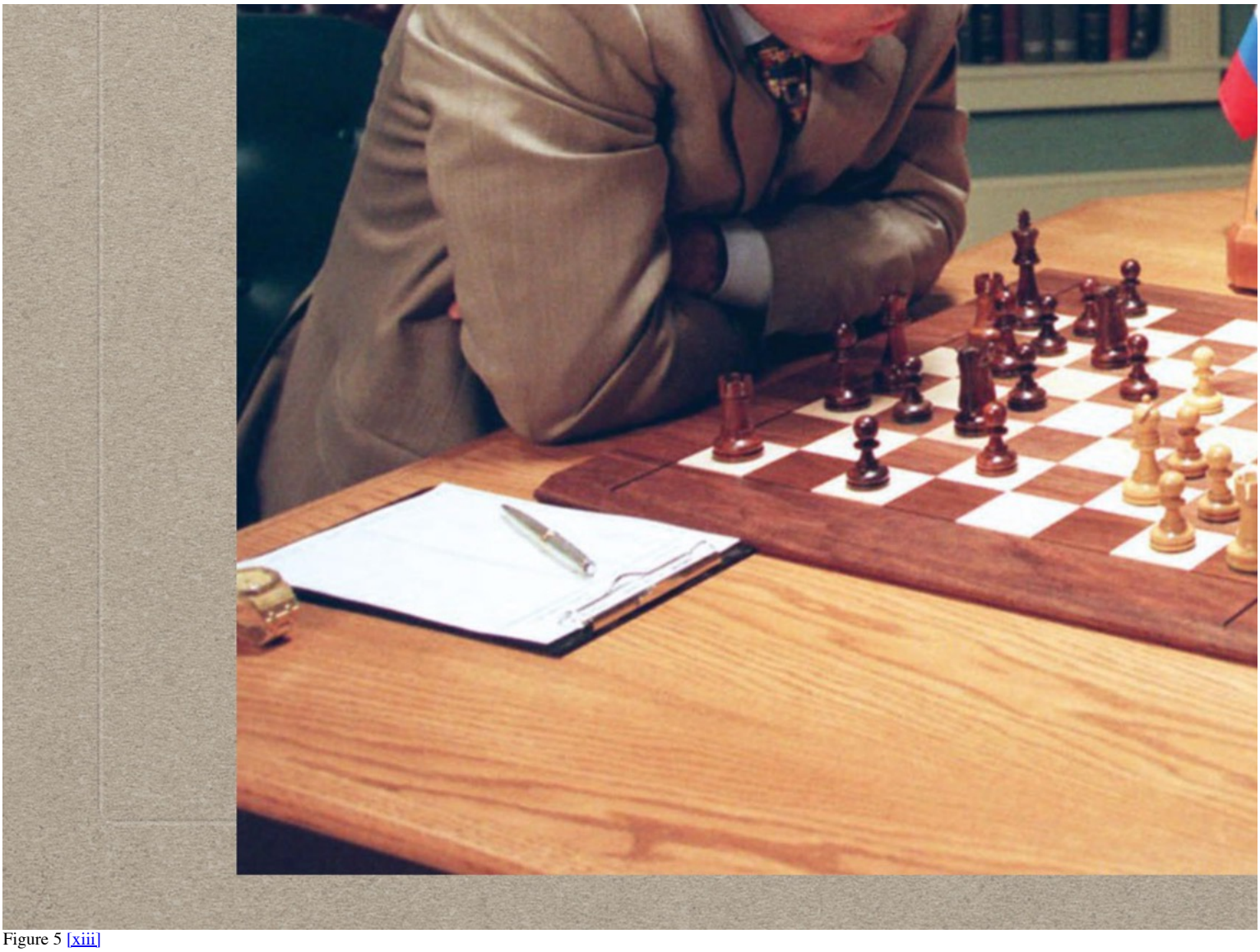


Figure 5 [xiii]

Besides Watson, you may also remember another famous game-playing computer from IBM from two decades ago named Deep Blue.[xiv] In 1996, Garry Kasparov beat Deep Blue, winning three matches and drawing two:[xv]

Figure 6[xvi]

> The next year, he played against a new and improved Deep Blue and lost the match. Once again, the psychological toll of facing off against an inscrutable opponent played a key role. Although he easily won the first game, Deep Blue dominated the second. Kasparov … was visibly perturbed — sighing and rubbing his face — before he abruptly stood and walked away, forfeiting the match.
>
> He later said he was again riled by a move the computer made that was so surprising, so un-machine-like, that he was sure the IBM team had cheated. What it may have been, in fact, was a glitch in Deep Blue's programming: [In 2014, one the designers of Deep Blue revealed what he believed happened:] Faced with too many options and no clear preference, the computer chose a move at random. … [T]he move that threw Kasparov off his game and changed the momentum of the match was not a feature, but a bug.[xvii]

What lessons could be drawn from this story? The first and most obvious lesson is that it is very difficult to anticipate and understand the size of the impact that a seemingly innocuous design rule (i.e., "choose a random move when there are too many options and no clear preference among them") may have in a specific, unforeseen situation. The second lesson is that if Deep Blue had been intelligent enough to understand the pragmatic effect of an inscrutable, random move on its opponent (something that a human expert might plausibly have anticipated), its move could have been heralded as a brilliant feature rather than derided as a bug.

CADE METZ BUSINESS 09.16.15 10

Figure 7 [xviii]

Fast-forward to 2016. Google, a company that runs off of 2 billion lines of code "in a single code repository available to all 25,000 Google engineers,"[xix] now dominates large-scale computing. "By comparison, Microsoft's Windows operating system — one of the most complex software tools ever built for a single computer, a project under development since the 1980s — is likely in the realm of 50 million lines"[xx] — only 2 ½ per cent of the size of Google's shared code base.

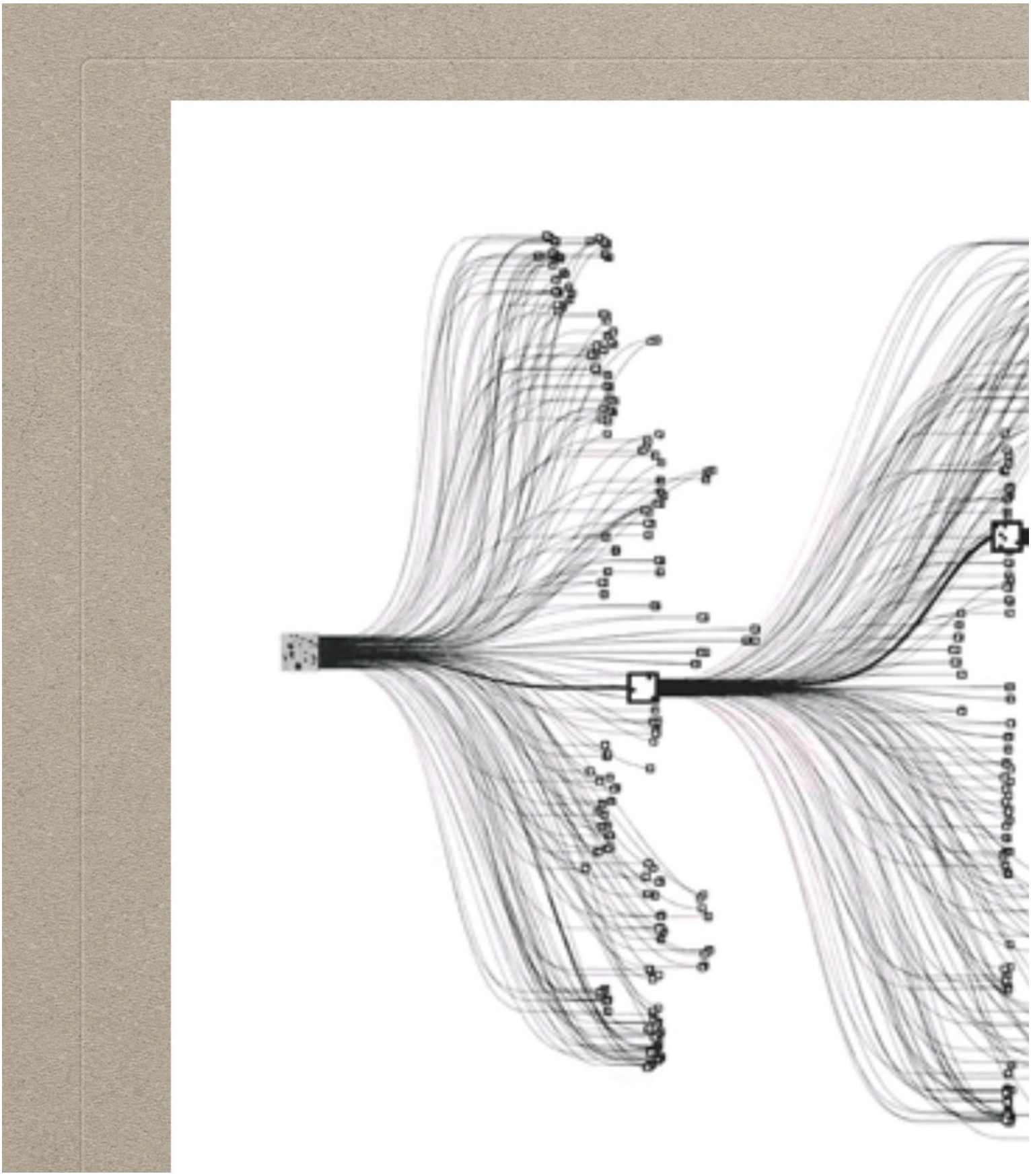
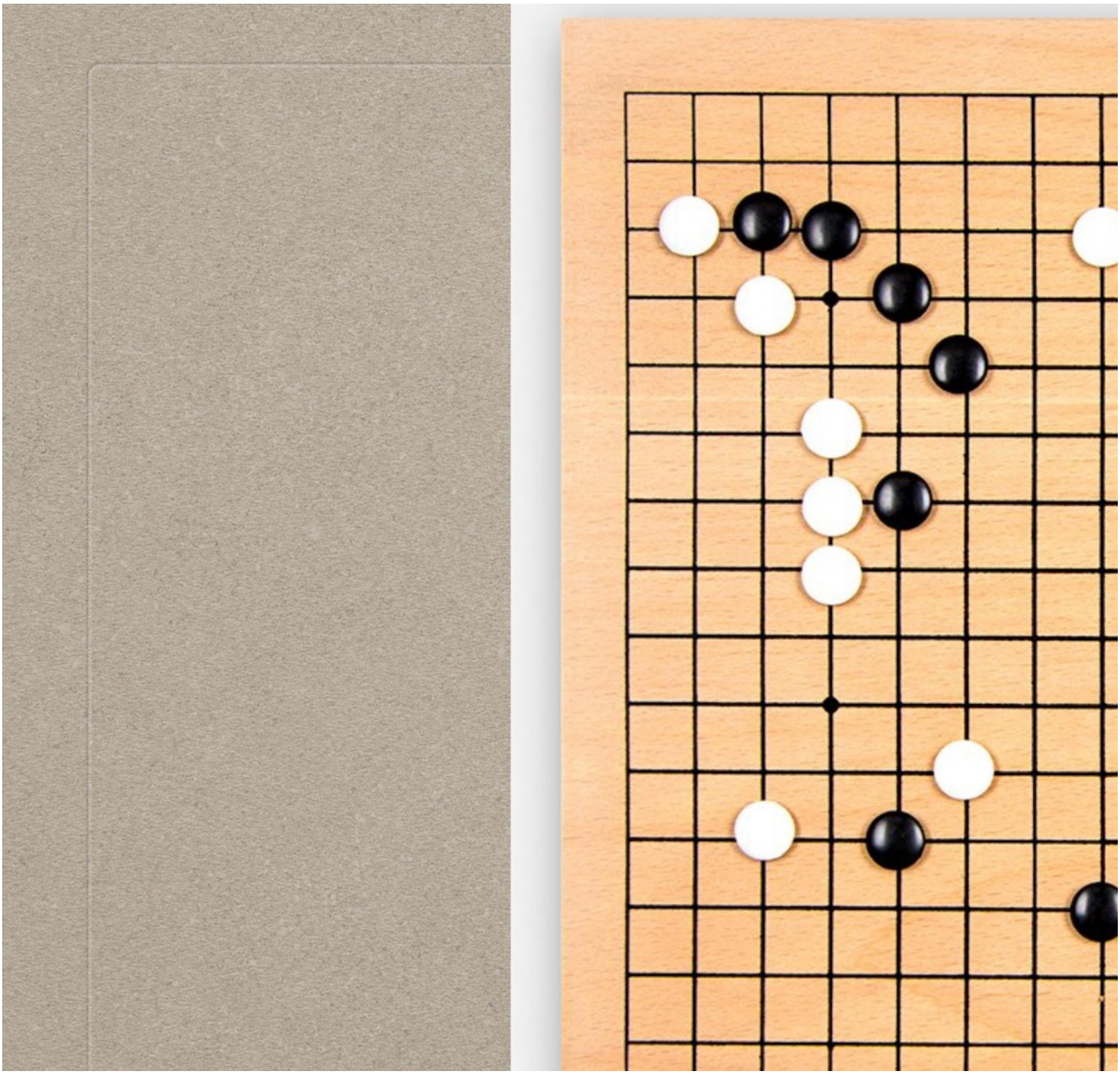


Figure 8[xxi]

By its formidable computing prowess, Google has well earned the right to challenge the human champions of Go, reputed to be the most complex game ever invented. Go is exponentially complex, with each move in the game presenting another 250 possible moves.[xxii] For this reason, mere "doublings in computing power and Monte Carlo approaches have been … inadequate."[xxiii] In January 2016, when Google's AlphaGo program beat the European Go champion 5-0, one of my research colleagues quipped: "My reaction when this happened was the same as Ken Jennings when beaten by Watson at *Jeopardy!* — 'I for one welcome our new computer overlords.' But first I want to see them win the pie-eating contest. ;)"[xxiv] Two months later, AlphaGo beat Lee Sedol, a world-renowned South Korean master of Go.[xxv]



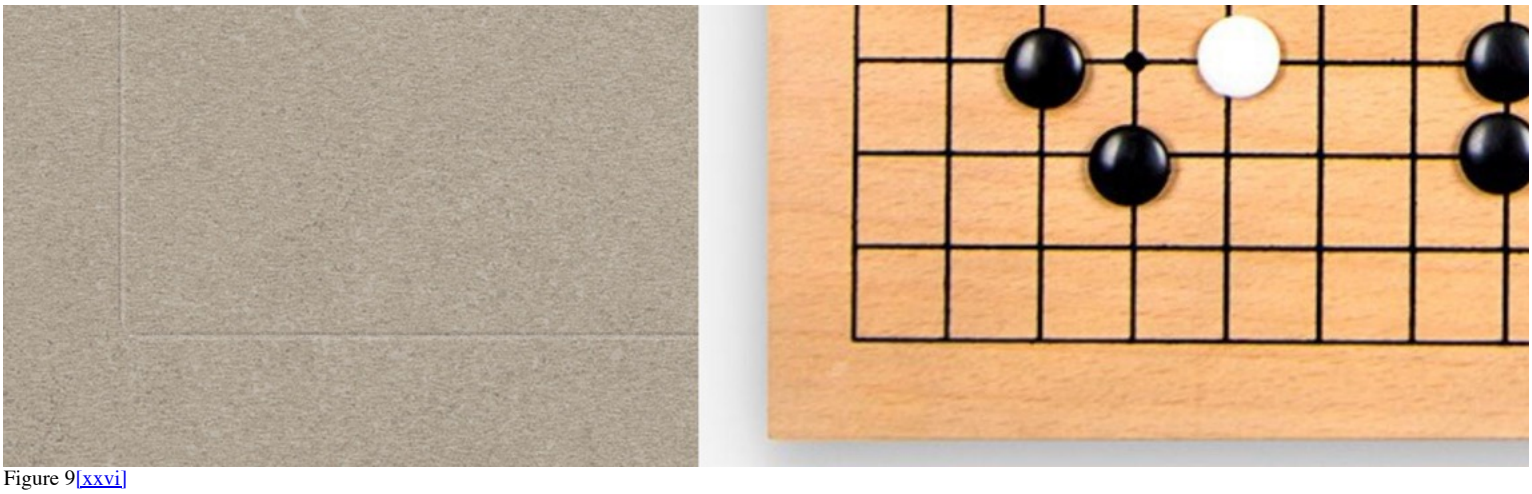


Figure 9[xxvi]

What makes all this important is not merely that AlphaGo was able to play successfully against formidable human opponents, but rather the degree to which it represented a significant shift in emphasis in *how* the game was played by the machine. My son Robert, who works for Google and knows a lot more than I do about AlphaGo's approach, outlined two reasons why an alternative was needed to typical game-playing strategies that rely largely on high-powered look-ahead searches through the space of possible future moves to see what current move would be most likely to lead to a later advantage. First, because the search space of possible moves in Go is so large, simply outpacing the human through extensive search is no longer feasible — there is no practical alternative at the current time but to use machine learning to a much greater degree. Second, because there are relatively few expert-level games available for the machine to learn from, machine-learning strategies needed to be pushed further than ever before in order to yield more results with less training data. The aspiration of the researchers is for the machine to develop something "akin to an intuition about what good positions and moves are." Although this is not the first time that people have tried to attack the problem of playing Go using machine learning, it is the first time they've figured out how to do it effectively — and the value of the novel techniques that have come out of this research do not seem to be confined to playing Go. Writes Robert:[xxvii]

> I personally find this most exciting because a lot of improvements in [machine learning] seem to have been due to being able to train on orders of magnitude more data (which, of course, is non-trivial theoretical and engineering challenge), but once you've trained on (say) all digitized bilingual text in the world there's not much further to go from there. These techniques are starting to explore the path of more effectively extracting "intelligence" out of (relatively) smaller corpora of data. It's also techniques like this that will allow it to produce results *better* than the data it trained on, which is a more fascinating proposition.

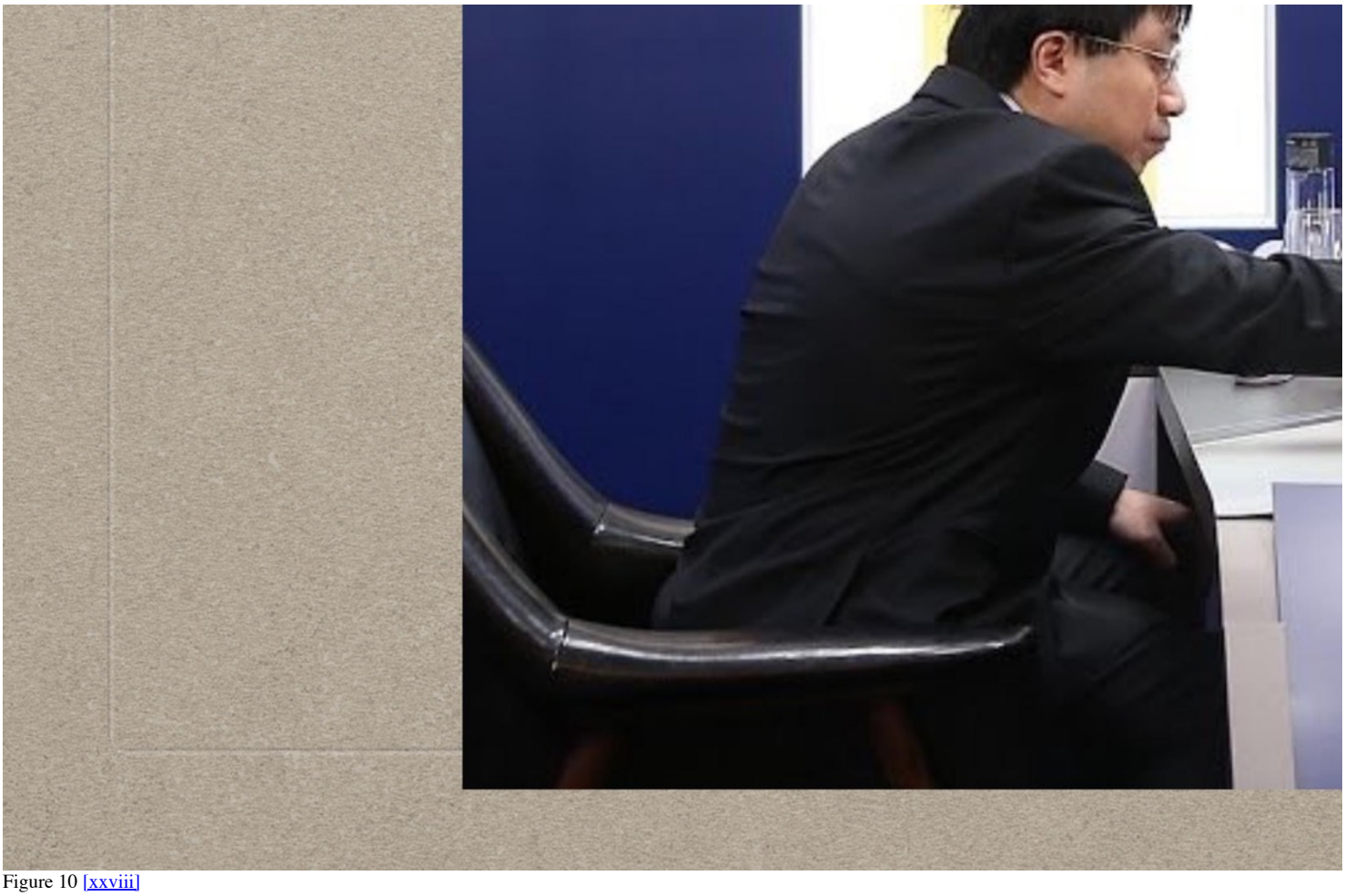


Figure 10 [xxviii]

Despite such exciting advances, an panel sponsored by the United States White House Office of Science and Technology Policy concluded in May 2016 that "A. I. research is still far from matching the flexibility and learning capability of the human mind. … 'The A. I. community keeps climbing one mountain after another, and as it gets to the top of each mountain, it sees ahead still more mountains,'" [xxix] summarized one of the scientists. Another researcher observed that "attention-getting feats like Google's AlphaGo program … had plenty of humans behind the machine doing the work" — at least for the foreseeable future.[xxx]

(To be continued in Part 4)

References

Anonymized. E-mail message to Jeffrey M. Bradshaw, January 30, 2016.

Anonymous. "How IBM makes its partners smarter: IBM's Watson turns cognitive computing into business success." *CyberTrend 14*, no. 2 (February 2016): 10-15. http://www.cybertrend.com/article/19837/how-ibm-makes-its-partners-smarter. (accessed March 5, 2016).

Bostrom, Nick. *Superintelligence: Paths, Dangers, Strategies*. Oxford, England: Oxford University Press, 2014.

Bradshaw, Robert W. E-mail message to Jeffrey M. Bradshaw, January 30, 2016.

Clarke, Arthur C. "The future isn't what it used to be." *Engineering and Science 33*, no. 7 (1970): 4-9. http://resolver.caltech.edu/CaltechES:33.7.clarke. (accessed September 18, 2015).

———. "Foreword: The birth of HAL." In *HAL's Legacy: 2001's Computer as Dream and Reality,* edited by David G. Stork, xi-xvi. Cambridge, MA: The MIT Press, 1997.

Cyc. In *Wikipedia*. https://en.wikipedia.org/wiki/Cyc. (accessed March 15, 2016).

Davis, Ernest, and Gary Marcus. "Commonsense reasoning and commonsense knowledge in Artificial Intelligence." *Communications of the ACM* 58, no. 9 (September 2015): 92-103.

Deep Blue (chess computer). In *Wikipedia*. https://en.wikipedia.org/wiki/Deep_Blue_(chess_computer). (accessed June 8, 2016).

Gaines, Brian R. "An ounce of knowledge is worth a ton of data: Quantitative studies of the trade-off between expertise and data based on statistically well-founded empirical induction." In *Proceedings of the Sixth International Workshop on Machine Learning,* edited by Alberto Maria Segre, 156-59. San Francisco, CA: Morgan Kaufman, 1989. http://pages.cpsc.ucalgary.ca/~gaines/reports/MŁML89/ML89.pdf. (accessed June 10, 2016).

Greaves, Mark, Heather Holmback, and Jeffrey M. Bradshaw. "What is a conversation policy?" *Proceedings of the Autonomous Agents '99 Workshop on Specifying and Implementing Conversation Policies*, Seattle, WA, 1-5 May 1999, 1999, 1-9.

Guha, R.V., and Douglas B. Lenat. "Cyc: A midterm report." *AI Magazine*, Fall 1990, 33-58.

Hoffman, Robert, Paul Feltovich, Kenneth M. Ford, David D. Woods, Gary Klein, and Anne Feltovich. "A rose by any other name… would probably be given an acronym." *IEEE Intelligent Systems*, July-August 2002, 72-80.

Holmback, Heather, Mark T. Greaves, and Jeffrey M. Bradshaw. "'Agent A, can you pass the salt?': The role of pragmatics in agent communication (expanded version accessible online)." In *Proceedings of the Third Annual Conference on Autonomous Agents (Autonomous Agents '99)*, edited by Oren Etzioni, J. P. Müller and Jeffrey M. Bradshaw, 368-69. Seattle, WA: New York City, NY: ACM Press. http://www.jeffreymbradshaw.net/publications/Pragmatics.submitted.new.pdf. (accessed June 18, 2016).

Jobs, Steve. 1983. The future isn't what it used to be (Presintation to the nternational Design Conference in Aspen (IDCA), 15 June 1983). In *SoundCloud*. https://w.soundcloud.com/player/?url=http%3A%2F%2Fapi.soundcloud.com%2Ftracks%2F62010118&show_artwork=true. (accessed September 18, 2015).

Kasparov playing against Deep Blue (8 June 2016). 2016. In *Britannica Online for Kids*. http://kids.britannica.com/comptons/art-158037/Garry-Kasparov-plays-against-Deep-Blue-the-chess-playing-computer. (accessed June 8, 2016).

Latson, Jennifer. 2015. Did Deep Blue beat Kasparov because of a system glitch? (17 February 2015). In *TIME*. http://time.com/3705316/deep-blue-kasparov/. (accessed June 8, 2016).

Lenat, Douglas B., M. Prakash, and M. Shepard. "CYC: Using Common Sense Knowledge to Overcome Brittleness and Knowledge Acquisition Bottlenecks." *AI Magazine 6*, no. 4 (1986).

Lenat, Douglas B., and R.V. Guha. *Building Large Knowledge-based Systems*. Reading, MA: Addison-Wesley, 1990.

Lenat, Douglas B. "From 2001 to *2001*: Common sense and the mind of HAL." In *HAL's Legacy: 2001's Computer as Dream and Reality,* edited by David G. Stork, 193-209. Cambridge, MA: The MIT Press, 1997.

———. 2015. Computers with common sense. TEDxYouth@Austin (video published 21 May 2015). In *YouTube*. https://www.youtube.com/watch?v=2w_ekB08ohU. (accessed March 15, 2016).

Markoff, John. 2016. Artificial Intelligence is far from matching humans, panel says (25 May 2016). In *The New York Times*. Document3. (accessed June 8, 2016).

McDermott, Drew. "Artificial intelligence meets natural stupidity." *SIGART Newsletter*, no. 57 (1976): 4-9.

McGourty, Colin. 2014. Man vs Machine: A poet on Kasparov-Deep Blue (29 October 2014). In *Chess24*. https://chess24.com/en/read/news/man-vs-machine-a-poet-on-kasparov-deep-blue. (accessed June 8, 2016).

Metz, Cade. 2015. Google is 2 billion lines of code — and it's all in one place (15 September 2015). In *WIRED*. http://www.wired.com/2015/09/google-2-billion-lines-codeand-one-place. (accessed June 8, 2016).

Mok, Kimberley. 2016. Artificial 'imagination' helped Google AI master Go, the most complex game ever invented. In *The New Stack*. http://thenewstack.iøgoogle-ai-beats-human-champion-complex-game-ever-invented/. (accessed June 8, 2016).

Moody, Sidney. 1999. The brain behind Cyc: Scientist Doug Lenat discusses Artificial Intelligence. In *The Austin Chronicle*. http://www.austinchronicle.com/screens/1999-12-24/75252/. (accessed March 15, 2016).

Ormerod, David. 2016. Lee Sedol defeats AlphaGo in masterfaul comeback – Game 4. In *Go Game Guru*. https://gogameguru.com/lee-sedol-defeats-alphago-masterful-comeback-game-4/. (accessed June 8, 2016).

Sang-Hun, Choe. 2016. Google's computer program beats Lee Sedol in Go tournament. In *The New York Times (15 March 2016)*. http://www.nytimes.com/2016/03/16/world/asia/korea-alphago-vs-lee-sedol-go.html. (accessed March 31, 2016).

Smith, Ira A., Phil R. Cohen, Jeffrey M. Bradshaw, Mark Greaves, and Heather Holmback. "Designing conversation policies using joint intention theory." *Proceedings of the Third International Conference on Multi-Agent Systems (ICMAS-98)*, Paris, France, 2-8 July, 1998, 269-76.

Stoffer, Shawn. 2000. Cyc: Building HAL. In *Department of Computer Science, University of New Mexico*. http://www.cs.unm.edu/~storm/docs/Cyc.htm. (accessed March 15, 2016).

Valéry, Paul. 1937. "Our destiny and literature." In *Reflections on the World Today*. Translated by Francis Scarfe, 131-55. New York City, NY: Pantheon Books, 1948.

Washington, Glynn, Stephanie Foo, and Murray Campbell. 2014. Kasparov vs. Deep Blue (8 August 2014). In *NPR Now*. http://www.npr.org/2014/08/08/338850323/kasparov-vs-deep-blue. (accessed June 8, 2016).

End Notes

[i] N. Bostrom, Superintelligence.

[ii] Doug Lenat's Cyc can be seen as a philosophical grandfather to such efforts. In 1999, Lenat wrote: "HAL was a general artificial intelligence, and Cyc is the closest thing that exists in the world to that kind of general artificial intelligence" (S. Moody, Brain). See also D. B. Lenat, From 2001.

[iii] https://upload.wikimedia.org/wikipedia/commons/thumb/f/f6/HAL9000.svg/1024px-HAL9000.svg.png.

[iv] A. C. Clarke, Foreword, p. xi.

[v] http://www.cyc.com/wp-content/uploads/2015/04/kbase.png.

[vi] For brief summaries of some of the most common criticisms of Cyc, see Cyc; S. Stoffer, Cyc. For Lenat's own assessments of his work on Cyc, see R. V. Guha et al., Cyc: A midterm report; D. B. Lenat et al., CYC: Using Common Sense Knowledge to Overcome Brittleness and Knowledge Acquisition Bottlenecks; D. B. Lenat et al., Building Large Knowledge-based Systems; D. B. Lenat, From 2001. For a recent video perspective by Lenat, see D. B. Lenat, Computers with Common Sense.

[vii] See B. R. Gaines, Ounce of Knowledge for a preliminary study that suggests "the possibility of developing a quantitative science of knowledge in terms of the amount of data reduction that knowledge buys us when carrying out empirical induction."

[viii] https://commons.wikimedia.org/wiki/File:IBM_Watson.PNG.

[ix] Anonymous, How IBM.

[x] E. Davis et al., Commonsense Reasoning, p. 92.

[xi] R. Hoffman et al., Rose, p. 75.

[xii] See, e.g., H. Holmback et al., 'Agent A, can you pass the salt?' See also M. Greaves et al., What is a conversation policy?; I. A. Smith et al., Designing conversation policies using joint intention theory.

[xiii] Kasparov Playing.

[xiv] Deep Blue.

[xv] J. Latson, Did Deep Blue Beat Kasparov?

[xvi] C. McGourty, Man vs Machine.

[xvii] G. Washington et al., Kasparov Vs. Deep Blue.

[xviii] C. Metz, Google Is Two Billion Lines of Code.

[xix] Ibid.

[xx] Ibid.

[xxi] K. Mok, Artificial Imagination.

[xxii] Ibid.

[xxiii] Anonymized, 30 January 2016.

[xxiv] Ibid.

[xxv] C. Sang-Hun, Google's Computer Program.

[xxvi] https://gogameguru.com/i/2016/01/DeepMind-AlphaGo.jpg.

[xxvii] R. W. Bradshaw, 30 January 2016. Robert's cogent summary of what makes AlphaGo's approach interesting is informative:

Your best move is the one that maximizes your position. Your position is defined by your opponent's best move. So it's really a recursive definition. Of course for any reasonably complex game, you can't all the way to the win/lose state, so at some point you have to "judge" a position and/or move on its own (without looking forward to what could happen) based on its intrinsic properties. Both computers and humans do both, but computers tend to (need to) compensate for poor "leaf position" judgment abilities by searching much broader and deeper.

For chess, this works well enough both because the number of sensible (or at least legal) moves is often quite limited (say tens, often less) and one can come up with simple scoring "limited intelligence required" algorithms that work pretty well (e.g. assign a number of points to each live player, and add them up–even great players have trouble coming back from a significant deficit using the standard weights). These two facts, some clever tricks, and the computing power in our reach, make Deep Blue (and its successors) a reality.

On the other hand, for Go, both the number of possible moves at any point in the game is so great as to make it infeasible to search a reasonable fraction of moves deeply, and the "goodness" of a position is harder to quantify. This is where the AI comes into play — it's used to (intrinsically but opaquely) judge how "reasonable/good" a move is, and how "good" a position is, (more) similar to how a person works, drastically reducing the search space (though it still tries out *way* more possibilities a human does). This better AI is the differentiator here.

Put another way, the AI problem they're trying to solve with deep neural nets is "what's the best move here" and though they make a great leap forward (agreeing with expert players over half the time now) it still needs the to be "amplified" by search to be competitive. But the fact that its good enough that when paired with a *feasible* search it can compete with expert humans makes this a measurable milestone. (And the claim is that without *any* search it still plays at the early amateur level.) Whether that's intelligent, well, the colloquial definition of AI excludes what has been accomplished.

Of course these aren't Go experts, they're just interested in (incrementally) advancing the state of AI in general. It seems the most novel theoretical aspect of this is in "reinforcement learning." There just aren't the same vast quantities of (expert-level) Go games played to rely on volume alone (like one can with, say, machine translation) without over-fitting, so they've had to rely more heavily on these self-learning techniques which is a lot trickier to get right (similar to how we had missionary apartments that only spoke French among themselves, which helped vocabulary but was horrible in reinforcing bad accents), and this helps validate that work.

I personally find this most exciting because a lot of improvements in [machine learning] seem to have been due to being able to train on orders of magnitude more data (which, of course, is non-trivial theoretical and engineering challenge), but once you've trained on (say) all digitized bilingual text in the world there's not much further to go from there. These techniques are starting to explore the path of more effectively extracting "intelligence" out of (relatively) smaller corpora of data. It's also techniques like this that will allow it to produce results *better* than the data it trained on, which is a more fascinating proposition.

[xxviii] D. Ormerod, Lee Sedol Defeats AlphaGo.

[xxix] J. Markoff, Artificial Intelligence Is Far From Matching.

[xxx] Ibid.