Toward a New Science of Common Sense

Ronald J. Brachman,[1] Hector J. Levesque[2]

[1] Jacobs Technion-Cornell Institute and Cornell University
[2] Dept. of Computer Science, University of Toronto
ron.brachman@cornell.edu, hector@cs.toronto.edu

Abstract

Common sense has always been of interest in AI, but has rarely taken center stage. Despite its mention in one of John McCarthy's earliest papers and years of work by dedicated researchers, arguably no AI system with a serious amount of general common sense has ever emerged. Why is that? What's missing? Examples of AI systems' failures of common sense abound, and they point to AI's frequent focus on expertise as the cause. Those attempting to break the brittleness barrier, even in the context of modern deep learning, have tended to invest their energy in large numbers of small bits of commonsense knowledge. But all the commonsense knowledge fragments in the world don't add up to a system that actually demonstrates common sense in a human-like way. We advocate examining common sense from a broader perspective than in the past. Common sense is more complex than it has been taken to be and is worthy of its own scientific exploration.

The modern-era data-intensive machine learning juggernaut continues to roll on, with a wide array of extraordinary results and significant commercial impact. But an increasing number of articles and books (for instance Pavlus 2020; Marcus and Davis 2019) point out that even the best of current AI falls short of the robust, general intelligence envisioned by its founders. Blunders made by generally powerful systems, such as shocking misidentifications of objects in visual images, have been recounted (Szegedy et al. 2014; Mitchell 2019). Surprising gaffes of otherwise remarkable systems like GPT (Vincent 2020) have been revealed as both humorous and disturbing (Marcus and Davis 2020). Self-driving cars make terrifying unexplainable mistakes (Hogan 2021). Several authors (see, for example Levesque 2017; Marcus and Davis 2019; Mitchell 2019; Toews 2020) have made the case that AI is still missing something critical to avoiding these mistakes, and they rightly identify the missing ingredient as what we would normally call *common sense*. But recent calls for a new generation of post-modern AI systems with common sense give no real prescription for getting there—or even a hint of what it would really mean for an AI system to have it. Our intention in this paper is to stimulate the field into closing this critical gap.

Expertise and the Brittleness Challenge

Despite the name of the field, Artificial Intelligence's biggest accomplishments have generally come from expertise rather than any more general kind of intelligence. Our greatest successes have been on tasks in narrow domains or circumscribed challenge problems, such as Go, facial recognition, infectious disease diagnosis, and the like. The most obvious limit of this is what some have called "brittleness"—the failure to produce reasonable outcomes (or any outcomes at all) in the face of challenges just beyond the boundaries of the expertise. This was a well-known shortcoming of the 1980s wave of expert systems, but as recent work has shown, it applies equally well to systems trained with extensive amounts of data (Szegedy et al. 2014; Marcus and Davis 2020; Page Street Labs 2020). AI systems are often fragile and show a noticeable lack of common sense. This may not be a critical problem for a system that only plays chess and whose entire world is limited to a chessboard and chess pieces. But AI's longer-term vision aspires to embed fully integrated systems in the real world, where artificial agents will need to be able to cope with a wide range of unanticipated events.

The emerging universe of self-driving cars provides a good example. We expect such cars to operate on regular real-world streets with natural phenomena occurring all around them—other drivers, signs and signals, pedestrians and dogs, unpredictable weather and road conditions, etc. But we see that, at least for now, brittleness is still rampant and current systems fail in ways that make it clear they have no common sense to fall back on when their expertise meets its limits (Hogan 2021). They make mistakes that seem counterintuitive or just plain silly. They cannot offer drivers reasons for their behavior and we cannot correct them by offering advice. We end up with fragile, inscrutable, incorrigible systems that can have serious and even fatal consequences when operating in the real world—largely because they have no common sense.

What Is Common Sense?

If we want to develop a plan for building common sense into AI systems, then the natural question to ask is, *what exactly is it*? The point of this paper is that that question has not been sufficiently answered; if common sense were fully defined and its technical challenges clearly articulated, there would be no need for a new call to action. Unfortunately much of

the recent writing on common sense in AI is not about what it is and how it can be realized, but about its absence in current systems. There are some thoughtful treatises on how much is missing (see, for example Davis and Marcus 2015), and there have been a number of technical efforts focused on isolated fragments of *commonsense knowledge* (see below), but there is currently no real clarity on how to build an AI system that consistently demonstrates common sense.

In our opinion, here is what common sense is about:

> *Common sense is the ability to make effective use of ordinary, everyday, experiential knowledge in achieving ordinary, practical goals.*

There is a lot to unpack in this characterization—words like "ability," "effective," "experiential," and "practical" are easy to gloss over but are actually nuanced and significant—but it is not our intention to do so here. (We take this up in considerable detail in (Brachman and Levesque 2022).) Rather, we want to show how the idea of *making use of knowledge* in an effective manner leads to a set of scientific questions that we believe are important for the field to consider in a systematic and unified way. Progress on this front would have a very significant effect on the ability of autonomous AI systems to operate in open-ended real life.

A Focus on Commonsense Knowledge

Of course the consideration of common sense as an aspect of intelligence is not a new phenomenon in AI. Even John McCarthy's earliest seminal paper in the field, "Programs with Common Sense" (McCarthy 1958), mentioned the goal right in the title. And a number of projects since then, including AI's longest continuously running project—Cyc—have been said to have focused on it (Lenat and Guha 1989; Matuszek et al. 2005; Metz 2016). But no robust AI system with common sense has ever emerged from this line of research. Something critical is still lacking.

In our view, the problem is that almost all the attention on common sense in AI to date has tightly focused on the commonsense knowledge that would be required, to the relative exclusion of several elements that are key to the success of natural systems with common sense. Researchers like Doug Lenat and others concentrated on the realm of missing, tacit facts (like "humans need air to breathe" or "you can't pick something up unless you're near it") that most people would know but that were never captured in formal knowledge bases. (Lenat's published example about *Romeo and Juliet* in (Lenat 2019) illustrates this). The missing facts were one reason that expert systems were stymied on edge cases, and their pursuit was a well-justified avenue for addressing the brittleness issue. But it concentrated on only one part of a bigger problem, which we will get to in a moment.

Along similar lines, researchers like Pat Hayes, Jerry Hobbs, Ernie Davis, Ray Reiter and many others developed formal logical theories of various aspects of the common sense world, where the inferences emerge as logical consequences (Hayes 1985a,b; Hobbs and Moore 1985; Davis 1990; Reiter 2001). But across the board, the effort was concentrated on foundational facts and rules and simply relied on general systems like logic (monotonic or non-monotonic) for the determination of consequences of the knowledge bits.

A different view of commonsense knowledge was embodied in the now decades-old memory-focused efforts of Marvin Minsky and Roger Schank and colleagues (Minsky 1985; Schank and Abelson 1977; Schank 1982). The emphasis there was more on the memories of past experiences than on general truths about the world. The focus was less on deriving conclusions from multiple facts, and more on recognizing patterns and drawing analogies between current circumstances and these remembered experiences as a way of solving new problems. Minsky's ideas about frames and Schank's work on scripts, plans, and other memory structures were often set in contrast with the more logical work noted above, but in the end, this line of work also spent most of its energy on knowledge and its organization. It should also be noted that even what we have called post-modern efforts in this space, like COMET (Bosselut et al. 2019), which look to build hybrid systems on top of deep learning engines, are still focused on expanding knowledge bases.

A Broader Perspective

As mentioned, no system with the kind of common sense we aspire to has ever emerged from these lines of work. It has become increasingly clear that no matter the scale of commonsense factual tidbits stored in a knowledge base, this is not enough to get over the fundamental hump of generally robust behavior outside the boundaries of expertise.

The crux of the issue is that *knowing even a vast number of commonsense facts is simply not the same as having and displaying common sense in the real world*. When a person says to another, "use a little common sense here," they are not asking only to recall some isolated bits of knowledge. Having common sense is not the same as being able to win some sort of obvious-fact trivia game ("Alex, what weighs more, a wheelbarrow or a grizzly bear?"). When we expect a person to use common sense, what we are insisting on is the use of background knowledge to influence what action to take or how to interpret an unexpected experience. Having common sense is substantially more than having commonsense knowledge. At the very least it is the appropriate and timely *application* of this knowledge that is critical.

One gets the sense from many presentations of commonsense knowledge in action in AI papers that we are creating the equivalent of locally-scoped "fact calculators," which can be fed a number of piecemeal facts and (if all goes well) will spit out reasonable inferences. This is clearest in the case of systems based directly on logic (*à la* McCarthy/Cyc): you give the inference engine some axioms, push a button, and it can come to some interesting and even unexpected conclusions. But it's working in isolation of the contextual situation, goals, and prior history of the agent; those are all in the hands of the user pressing the buttons. The Romeo and Juliet example mentioned above is just like this.

The more memory-based research (*à la* Minsky/Schank) comes closer to advocating the use of knowledge in context. The suggestions in a frame for what to look at in a scene, or what other knowledge structures to look at to explain a

phenomenon, can direct reasoning in a more contextually-relevant way. But these are only first steps to more broadly intelligent behavior. What is missing is a global architecture that invokes the incremental reasoning steps of a frame system at the right times with focus on the right issues and controls its application to solve the problem at hand. How does an agent decide which chunks of knowledge to look at next, how they should combine with other chunks of knowledge, when to go back for another try, and even when to give up and try something different?

Even efforts that purport to focus on commonsense reasoning rather than purely on commonsense knowledge seem to have this kind of "isolated steps" feel. Take efforts in qualitative reasoning, for example. The inference mechanics in such efforts expressly attempt to avoid getting bogged down in mathematical detail, thereby reflecting what seems to be a common human trait of quick, qualitative analysis. This is no doubt important and will play a role in future AI systems with common sense. But what is the bigger picture here? How can this kind of stepwise computation be used at the right times and in the right ways by an agent being hit with an unexpected event in the world? How might a qualitative inference capability be integrated with a more logical one, to assure an integrated agent's survival and practical success in the world? Being able to reason quickly and qualitatively is crucial, but it doesn't in itself give you common sense (even with a huge knowledge base of tiny obvious facts).

Related to this, and generally missing in AI systems, is the circumstances and methods of invocation of common sense. In humans, *common sense is not always active*. Much of what humans do every day is routine: our days are dominated by mindless, rote behavior. We follow set patterns that we've learned over time—brushing our teeth, walking the dog, even driving to work on an uneventful day. What causes us to break out of a routine and think more deliberately about what we are doing? In our view, when something unusual happens during the execution of a mindless routine, the first resort is common sense—using past experience to quickly and plausibly explain the unusual event or to guide the next action to take. This seems to be one of the primary things common sense is for. If common sense fails to provide a plausible next action or if its suggestion fails, a more thoughtful, deeper analysis can then take over. The speed and facility of common sense spares cognitive workload and often provides adequate solutions to problems and guidance for actions. But when and how it is invoked, and how it decides what background knowledge to use, are questions that have not been addressed by prior AI efforts.

Time to Consider a New Science

What is needed here in our opinion is a new start on an old problem. Common sense has generally not been treated in AI as a first-class problem, studied from start to finish as a whole. As a field we need to step back from building larger training sets or capturing millions of explicit commonsense tidbits, and analyze common sense in and of itself as a significant facet of intelligence. Then we can approach its implementation in AI systems as an integrated whole. What we need is a new science of common sense.

What would such a new science aspire to cover?

- **Commonsense knowledge**: Years of working on large commonsense fact bases like Cyc will not be wasted, although in our view, much more focus needs to be placed on the *experiential* basis of common sense. How are experiences remembered, generalized, and organized so that they can be called to mind when needed? Thought should be given to mechanisms for representing baseline ontological information, general rules of thumb, exceptions, and a host of other items that distinguish commonsense knowledge from other forms of knowledge. The Minsky/Schank lines of thinking should be reexamined and their complementarity to the more logic-based McCarthy/Cyc line should be investigated.
- **Commonsense reasoning**: This needs a careful analysis, definition, and prescription for implementation. For any chain of commonsense inference, we need to be clear on just what the inputs and the expected outputs are going to be. Are all logical consequences going to be computed? If so, what is the plan to ensure this can be done quickly enough? If not, what exactly is going to be left out? Rapid, plausibly sound inference seems to characterize common sense but has been underdeveloped in AI.
- **Cognitive architecture**: Critical to the overall phenomenon of common sense is when and how it is invoked and how it fits with the rest of cognition, perception, and action, and how metareasoning comes into play. How it interrelates with goals and drives and overall priorities will be important. Key questions related to focus of attention will need to be addressed, including how attention is focused on relevant items of background knowledge, and how it moves away from one thing to a more promising one. We'll also need to sort out mechanisms that smoothly allow the agent to give up on commonsense reasoning and move to a more analytical, heavier-workload reasoning effort as needed.
- **Learning**: It is generally agreed that the bulk of the basis for common sense in humans is learned from experience. Machine learning is the dominant technical thread in AI right now, but it has focused heavily on classification of inputs and predictive technologies like transformers. What would a learning machine look like if it were targeted to learning general knowledge of the sort one sees in Cyc? How would a machine go about learning how to use any knowledge it may have already learned? Can some of the architectural considerations mentioned above be learned or must they be innate in the underlying framework of an AI system? Finally, can common sense itself be *taught* after the fact? There are some self-help books out there that seem to imply that it can; if so, what are the implications for AI systems?
- **Explanation and advice-taking**: Finally, we believe that no system that purports to be autonomous should be deployed without common sense—how common sense relates to autonomy, explanation, and advice-taking will need to be part of this new endeavor. Autonomous systems need to be responsible for their actions and need to be open to taking advice as necessary from others.

Building on Prior Work

There is no doubt that the commonsense knowledge that has been studied over the years will be of value in this new undertaking. Minsky's frame ideas tantalizingly hinted at how the knowledge structures should be used, for things like "differential diagnosis" and reconceptualizing. And there are other sources of insight that can be drawn upon.

While the psychology literature is surprisingly short on analyses of common sense in humans, the work of Sternberg and colleagues on what he calls "Practical Intelligence" is clearly relevant (Sternberg et al. 2000). (Sternberg cites four modes of intelligence, and equates one of these—practical intelligence—exactly with common sense.) From an AI perspective there are limitations in this work's perspective, but it is worth integrating into the big picture of synthetic common sense. Along a different dimension, psychologists have posited a "cognitive continuum," in which it is postulated that common sense fits in its own position between "intuition" and "analysis" (see (Hammond et al. 1987); Hammond calls common sense "quasi-rationality"). This kind of account may inspire how to build an integrated AI system that allows common sense to be used at the right time and to show its value in frequent avoidance of heavy cognitive burdens. Along related lines, the psychologist Daniel Kahneman postulates a distinction between rapid intuitive processing in what he calls "System 1" and more thoughtful, methodical reasoning in what he calls "System 2" (Kahneman 2011). It is not clear at this stage how well Kahneman's classification aligns with the common sense/expertise distinction we believe is central to AI. At the very least, it appears that common sense as we see it does not fall neatly within System 1 or within System 2; it instead shares some characteristics with each, and has some critical features not accounted for in either.

Given its focus on simple reasoning using models rather than general rules and abstractions, the work of psychologist Philip Johnson-Laird is worth taking into account (Johnson-Laird 1983). We would also need to account for the difference between common sense and the broader notion of *rationality*, and explore and build on relationships to bounded and minimal rationality (Simon 1990; Cherniak 1986). The work of Gary Klein on intuitive decision-making is also of potential use (Klein 2007). And prior psychology work on prototypes and exemplars is worthy of incorporation (Rosch and Lloyd 1978; Smith and Medin 1981). From an AI perspective, a number of prior efforts on cognitive architecture (Kotseruba and Tsotsos 2020) will be relevant as well.

Core Research Questions

Taking common sense seriously as its own integrated subject matter leads to a number of important research questions. The key questions of the field will need to be articulated. Here are some candidates:

- What exactly is common sense? What technical definition best suits the needs of AI?
- What are appropriate tests for the presence of common sense? How can we tell if we are getting closer to building it into our AI systems?
- How is experiential knowledge represented, accessed, and brought to bear on current situations? What is the role of analogy? How does the ability to recognize something or see something as another thing (or even as an instance of an abstract concept) develop and get used?
- How is commonsense knowledge learned as new experiences happen? How is the update different when knowledge is acquired through language?
- What ontological frameworks are critical to build into an AI system? Are there special properties of the knowledge of the physical world that need to be handled in a way that is different from its non-physical counterparts?
- What is the relationship between common sense and the broader notion of rationality (including bounded rationality, minimal rationality, etc.)?
- What overall architecture is best suited for the multiple roles of common sense? What mechanism(s) should be used to invoke common sense out of routine, rote processing, and then to sometimes go beyond it to more specialized forms of expertise?
- What role, if any, does metareasoning play?

Refocusing on Common Sense as a Phenomenon

Since the beginning of AI, McCarthy and a limited cohort of researchers have set their sights on giving computers common sense. Unfortunately, while the last sixty years has provided us impressive technology that works on narrow problems, it has failed to deliver the ability to deal with the unpredictable open world: we do not have AI systems that can use common sense to solve life's rampant mundane problems and respond reasonably and practically to unforeseen events. It is frequently said of AI that it can rival the most expert of human experts in many fields but cannot do the everyday things that a six-year-old can. Our belief is that the field's thinking about common sense has been limited and has cornered itself into a focus on commonsense knowledge and isolated islands of inference, and has never looked into what common sense as a whole may be. We need to move from systems with large amounts of independent knowledge fragments to systems that show that they can use common sense in their everyday interactions with the world. The way to do this is to step back and consider common sense in all its glory, including not just the knowledge equivalent of sound bites, but how it is based in experience and how and when it is applied. To get to true AI—systems that can be deployed and operate autonomously in the real world—we need to tackle common sense head on, as a first-class subject of study.

References

Bosselut, A.; Rashkin, H.; Sap, M.; Malaviya, C.; Celikyilmaz, A.; and Choi, Y. 2019. COMET: Commonsense transformers for automatic knowledge graph construction. Submitted June 12, 2019. arXiv:1906.05317.

Brachman, R. J.; and Levesque, H. J. 2022. *Machines like Us: Toward AI with Common Sense*. Cambridge, MA: MIT Press.

Cherniak, C. 1986. *Minimal Rationality*. Cambridge, MA: MIT Press.

Davis, E. 1990. *Representations of Commonsense Knowledge*. San Mateo, CA: Morgan Kaufmann Publishers, Inc.

Davis, E.; and Marcus, G. 2015. Commonsense reasoning and commonsense knowledge in Artificial Intelligence. *Communications of the ACM*, 58(9): 92–103.

Hammond, K. R.; Hamm, R. M.; Grassia, J.; and Pearson, T. 1987. Direct comparison of the efficacy of intuitive and analytical cognition in expert judgment. *IEEE Transactions on Systems, Man, and Cybernetics*, 17(5): 753–770.

Hayes, P. J. 1985a. The naive physics manifesto I: Ontology for liquids. In Hobbs, J. R.; and Moore, R. C., eds., *Formal Theories of the Commonsense World*, 77–101. Norwood, NJ: Ablex Publishing Corporation.

Hayes, P. J. 1985b. The second naive physics manifesto. In Hobbs, J. R.; and Moore, R. C., eds., *Formal Theories of the Commonsense World*, 1–36. Norwood, NJ: Ablex Publishing Corporation.

Hobbs, J. R.; and Moore, R. C., eds. 1985. *Formal Theories of the Commonsense World*. Norwood, NJ: Ablex Publishing Corporation.

Hogan, M. 2021. Tesla's "Full Self Driving" beta is just laughably bad and potentially dangerous. *Road and Track*, March 19, 2021.

Johnson-Laird, P. N. 1983. *Mental Models: Towards a Cognitive Science of Language, Inference, and Consciousness*. Cambridge, MA: Harvard University Press.

Kahneman, D. 2011. *Thinking, Fast and Slow*. New York: Farrar, Straus and Giroux.

Klein, G. 2007. *The Power of Intuition: How to Use Your Gut Feelings to Make Better Decisions at Work*. New York: Doubleday.

Kotseruba, I.; and Tsotsos, J. K. 2020. 40 years of cognitive architectures: Core cognitive abilities and practical applications. *Artificial Intelligence Review*, 53(1): 17–94.

Lenat, D. B. 2019. What AI can learn from Romeo & Juliet. *Forbes*, July 3, 2019.

Lenat, D. B.; and Guha, R. V. 1989. *Building Large Knowledge-Based Systems: Representation and Inference in the Cyc Project*. Reading, MA: Addison-Wesley.

Levesque, H. J. 2017. *Common Sense, the Turing Test, and the Quest for Real AI*. Cambridge, MA: MIT Press.

Marcus, G.; and Davis, E. 2019. *Rebooting AI: Building Artificial Intelligence We Can Trust*. New York: Pantheon Books.

Marcus, G.; and Davis, E. 2020. GPT-3, bloviator: OpenAI's language generator has no idea what it's talking about. *MIT Technology Review*, August 22, 2020.

Matuszek, C.; Witbrock, M.; Kahlert, R. C.; Cabral, J.; Schneider, D.; Shall, P.; and Lenat, D. B. 2005. Searching for common sense: Populating Cyc™ from the web. In *Proceedings of the Twentieth National Conference on Artificial Intelligence (AAAI-05)*, volume 3, 1430–1435, Pittsburgh, July 2005.

McCarthy, J. 1958. Programs with common sense. In *Symposium on the Mechanization of Thought Processes*, 77–84. Teddington, UK: National Physical Laboratory, Teddington, England. Reprinted in Brachman, Ronald J. and Levesque, Hector J., editors. *Readings in Knowledge Representation*, pages 299–307. Los Altos, CA: Morgan Kaufmann Publishers, Inc., 1985.

Metz, C. 2016. One genius' lonely crusade to teach a computer common sense. *Wired*, March 24, 2016.

Minsky, M. L. 1985. A framework for representing knowledge. In Brachman, R. J.; and Levesque, H. J., eds., *Readings in Knowledge Representation*, 245–262. Los Altos, CA: Morgan Kaufmann Publishers, Inc.

Mitchell, M. 2019. *Artificial Intelligence: A Guide for Thinking Humans*. New York: Farrar, Straus and Giroux.

Page Street Labs. 2020. GPT-3 and a typology of hype. https://pagestlabs.substack.com/p/gpt-3-and-a-typology-of-hype. Accessed: 2021-12-06.

Pavlus, J. 2020. The easy questions that stump computers. *The Atlantic*, May 3, 2020.

Reiter, R. 2001. *Knowledge in Action: Logical Foundations for Specifying and Implementing Dynamical Systems*. Cambridge, MA: MIT Press.

Rosch, E.; and Lloyd, B. B., eds. 1978. *Cognition and Categorization*. Hillsdale, NJ: Erlbaum.

Schank, R. C. 1982. *Dynamic Memory: A Theory of Reminding and Learning in Computers and People*. Cambridge: Cambridge University Press.

Schank, R. C.; and Abelson, R. P. 1977. *Scripts, Plans, Goals, and Understanding: An Inquiry into Human Knowledge Structures*. Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.

Simon, H. A. 1990. Bounded rationality. In Eatwell, J.; Milgate, M.; and Newman, P., eds., *Utility and Probability*, 15–18. New York: Springer.

Smith, E. E.; and Medin, D. L. 1981. *Categories and Concepts*. Number 4 in Cognitive Science Series. Cambridge, MA: Harvard University Press.

Sternberg, R. J.; Forsythe, G. B.; Hedlund, J.; Horvath, J. A.; Wagner, R. K.; Williams, W. M.; Snook, S. A.; and Grigorenko, E. L. 2000. *Practical Intelligence in Everyday Life*. New York: Cambridge University Press.

Szegedy, C.; Zaremba, W.; Sutskever, I.; Bruna, J.; Erhan, D.; Goodfellow, I.; and Fergus, R. 2014. Intriguing properties of neural networks. In *Proceedings of the Second International Conference on Learning Representations ICLR 2014*, Banff, Canada, April 2014.

Toews, R. 2020. GPT-3 is amazing—and overhyped. *Forbes*, July 19, 2020.

Vincent, J. 2020. OpenAI's latest breakthrough is astonishingly powerful, but still fighting its flaws. *The Verge*, July 30, 2020.