



Existenz

Volume 17, No 1, Spring 2022

ISSN 1932-1066

On Reasoning, Commonsense Knowledge, and Consciousness

Ulrich Furbach

University of Koblenz, Germany

uli@uni-koblenz.de

Abstract: This essay addresses the application of formal logic to commonsense reasoning. To this end is depicted the integration of an automated reasoner based on predicate calculus with a method for statistical reasoning that is derived from word embeddings. This combination is motivated by the results from an experiment in the field of cognitive science involving human reasoning. Additionally, I briefly address its links to the Global Workspace Theory, which is currently considered to be one of the most prominent theories of consciousness.

Keywords: Jaspers, Karl; automated reasoning; human reasoning; Choice of Plausible Alternative challenge; Wason Selection Task; knowledge graph.

Research regarding artificial intelligence (AI) has been a widely discussed topic since its inception at the Dartmouth Summer Research Project on Artificial Intelligence in 1956 that was funded by the Rockefeller Foundation. At this conference, the name AI was coined in order to define a discipline concerned with figuring out how to get machines to use varieties of natural language and subsequently finding solutions to problems hitherto reserved for being solved by humans. Of course, in the past there have been scientists with similar goals, for example, Gottfried Wilhelm Leibniz who worked on a logical calculus and a universal formal language in order to solve arguments or disputes. Leibniz writes:

Having done that when thus differences of opinion arise, there is no more need for discussion and dispute between two philosophers than there is such need between two mathematicians. For it suffices to take pens into the hand, to sit at the abacus and to tell each other (at once if it pleases the friend): *Let us calculate.*¹

Formal logic has been the primary method for constructing symbolic problem-solving systems in artificial intelligence applications. Furthermore, artificial neural networks have also been utilized in order to design learning systems. This sub-symbolic approach has recently gained prominence due to the successful implementation of natural language processing systems such as the Chat Generative Pre-Trained Transformer (ChatGPT).

In what follows I focus on symbolic AI specifically its subfield automated reasoning based on formal logic. I explore the application of automated reasoning systems to commonsense problems, which humans excel at solving when knowledge and context are involved. However, incorporating large amounts of background knowledge into a reasoning task presents challenges for an automated system as it needs to

Schriften, Sechste Reihe, Vierter Band 1677-Juni 1690, Teil A, eds. Heinrich Schepers, Martin Schneider, Gerhard Biller, Ursula Franke, and Herma Kliege-Biller, Berlin, DE: Akademie Verlag 1999, pp. 909-915, here p. 913, transl. *Existenz* editors.

¹ Gottfried Wilhelm Leibniz, "De Arte Characteristica ad Perficiendas Scientias Ratione Nitentes," in *Philosophische*

determine which aspects of available data are relevant to the problem at hand. Hence, I propose a solution to this problem by using word embeddings, which is a method for integrating the meaning of symbols that are to be used in logical problem descriptions.

Automated Reasoning

Automated reasoning is one of the most traditional sub-disciplines of artificial intelligence and has been considered as being an important sub-discipline since artificial intelligence's very beginning. The methods for developing powerful and efficient inference systems are based on calculi of formal logic. John Alan Robinson's resolution calculus certainly represented an important starting point for the development of modern efficient proof procedures; subsequently many other calculi for various logics have been developed and applied. To further this end, sophisticated web repositories offer a library of problems for automated theorem proving systems. These depositories distinguish between and explain different problem classes and different logics, provide a comprehensive overview of current logical systems, and offer online access to state-of-the-art reasoning systems.

There are many applications and success stories with respect to automated reasoning. For instance, mathematical theorems can be proven automatically or with the help of deduction systems (such as the Kepler conjecture or Boolean Robbins algebra) and the verification of soft- and hardware systems has become a very relevant commercial domain.

An interesting application domain is to use first-order logic reasoning in order to solve commonsense problems that are usually formulated in natural language and can be easily solved by humans. In order to solve these problems in an artificial environment, one must translate it into a logical formula and feed it, along with background knowledge about the world, into the device which is now dubbed "the reasoner." This process is discussed in greater detail in the following sections.

Human Reasoning

When it comes to making simple everyday deductions, humans can do so effortlessly. However, automated reasoners often encounter computational procedure challenges when attempting to do the same. Inspired by the 1948 Caltech Hixon Symposium on "Cerebral

Mechanisms in Behavior," John McCarthy became one of the first computer scientists who pioneered the formalization of commonsense reasoning. In 1955, just one year prior to the Dartmouth conference, he had coined the term "artificial intelligence." McCarthy proposed a system program, the so-called "advice taker," which

will have available to it a fairly wide class of immediate logical consequences of anything it is told and its previous knowledge...We shall therefore say that *a program has common sense if it automatically deduces for itself a sufficiently wide class of immediate consequences of anything it is told and what it already knows.*²

In contrast to reasoning systems which were understood as theorem-proving systems in a formal logic environment, his definition of common sense makes the handling of knowledge a central issue. Inspired by this proposition, in the following decades a variety of approaches to formulate commonsense reasoning have subsequently emerged. Most of these approaches focus on logical aspects and assumed that the knowledge necessary to solve a particular problem was already formalized and made available to the reasoner. More recently, Ronald Brachmann and Hector Levesque propose to rethink the programming approach in artificial intelligence by including commonsense capabilities. They argue that

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.³

In contrast to a purely logical approach, where all necessary axioms are fed into the reasoner—as it is proposed by McCarthy—they are demanding a memory-based approach, that is apt of addressing the question

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? [NSC 3]

² John McCarthy, "Programs with Common Senses," in *Formalizing Common Sense: Papers by John McCarthy*, ed. Vladimir Lifschitz, Norwood, NJ: Ablex Publishing Corporation 1990, pp. 9-20, here p. 9.

³ Ronald J. Brachman and Hector J. Levesque, "Toward A New Science of Common Sense," *arXiv* 2112.12754v2 (6 February 2022), 1-6, here p. 3. [Henceforth cited as NSC]

One of the commonsense reasoning benchmark suites is the Choice of Plausible Alternatives (COPA) challenge.⁴ The challenge consists in being confronted with a statement together with a question regarding it and two alternative possible answers. The task is to find out which one of the two alternatives is more plausible. Drawing this conclusion requires a substantial amount of knowledge about the world and yet humans are capable to do it instantaneously.⁵

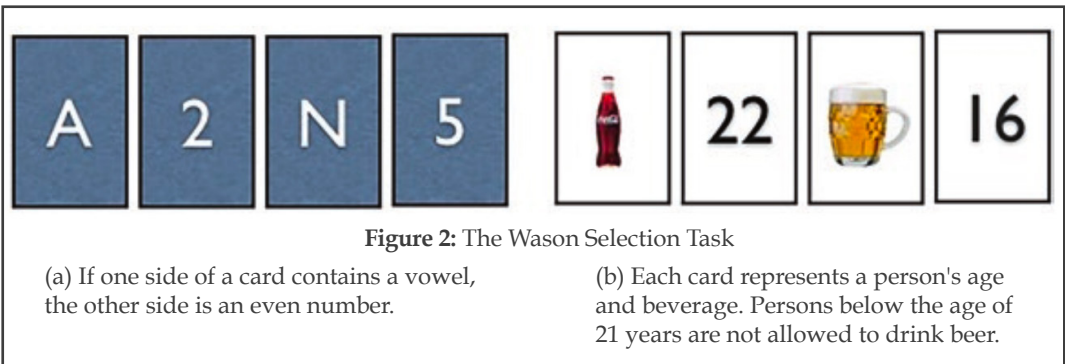
65: The family took their dog to the veterinarian. What was the CAUSE of this?
 The dog chewed on a bone.
 The dog injured his paw.

Figure 1:
 Example problem #65 from the COPA challenge.

The handling of large amounts of knowledge turns out to be a major problem for reasoning systems when processing common sense operations. The situation is different in experiments that involve humans. Some researchers argue that, in contrast to artificial reasoners, human reasoning is more error-prone when problems are being presented in abstract terms; yet as soon as there is an applied context, for example, derived from social interactions or has to do with matters of self-precaution, human performance is clearly better than machine performance.

A well-studied example regarding the challenge of drawing an inference is the Wason Selection Task.⁶ In the standard version of the task, a subject is presented

with four different cards. The subject is told that each card contains a letter on one side and a number on the opposite side. The instructions include a statement such as "If there is a vowel on one side, the opposite side contains an even number." The subject is asked to verify or disprove this statement by turning over a minimum number of cards. In this abstract task, less than 25% of the subjects were able to find the solution; this result has since been confirmed with a wide variety of subjects. Even students attending logic lectures at universities get similar poor results. It should be obvious (at least to a logician) that the statement, "If there is a vowel on one side, the opposite side contains an even number" is formulated as a material implication of the form "If P, then Q." And hence one needs to flip the card with A—in this instance P is true, so it would be necessary to check whether Q holds. Similarly, one needs to turn the card with the number 5, for if there is a vowel on the other side, the implication would be false.⁷



Turning the card with number 2 is not necessary, as the statement would be true, regardless of what is depicted on the other side. Numerous experiments have shown that people have problems correctly executing this abstract, but quite simple, inference. The situation changes drastically when context is added to the problem. The right part of Figure 2 shows the context of a social contract: One side of the cards shows a beverage, namely beer or soda, and the other side shows the age of the person drinking that beverage. Thus, each card represents a person drinking a

⁴ Melissa Roemmele, Cosmin Adrian Bejan, Andrew S. Gordon "Choice of Plausible Alternatives: An Evaluation of Commonsense Causal Reasoning," in *Papers from the 2011 AAAI Spring Symposium, No. 6: Logical Formalizations of Commonsense Reasoning*, Palo Alto, CA: AAAI Press 2011, pp. 90-5.

⁵ Andrew S. Gordon, *Choice of Plausible Alternatives (COPA)*, <https://people.ict.usc.edu/~gordon/copa.html>, the set of 500 questions is posted at <https://people.ict.usc.edu/~gordon/downloads/COPA-questions-dev.txt>.

⁶ Peter C. Wason, "Reasoning About A Rule," *The Quarterly Journal of Experimental Psychology* 20/3 (August 1968), 273–281.

⁷ Ulrike Barthelmeß and Ulrich Furbach, *A Different Look at Artificial Intelligence: On Tour with Bergson, Proust and Nabokov*, Wiesbaden, DE: Springer Nature 2023, pp. 84-6, the image of Figure 2 is taken from p. 85.

beverage. The rule now is: "If a person is under 21 years old, she is not allowed to drink beer." The task is again to check if the implication, this time the social rule, is fulfilled. In this case, 75% of the subjects usually find the correct solution quickly and effortlessly, irrespective of the fact that it is still a matter of recognizing the implication "If P, then Q." Researchers tend to believe that this difference in arriving at the correct inference is due to cognitive differences in processing abstract versus concrete challenges. Some researchers suggest that for tasks involving social reasoning different areas of the limbic system of the human brain are used in contrast to the ones used for other forms of reasoning.⁸

Numerous cognitive science experiments demonstrate that humans take the meaning of symbols into account during reasoning, which means that not only the symbols themselves but also their underlying meanings and associations are being considered. This ability to understand and use the meaning of symbols is an important part of cognition and is essential for effective reasoning. Hence, reasoning with an automated system would have to master this challenge.

Knowledge and Reasoning

In an effort of revisiting the relation between reasoning and knowledge I will briefly address an aspect of Karl Jaspers' thought regarding language. Jaspers brings up logicians of earlier centuries who differentiated between *notiones communes* and *notiones generales*, that is, between the essentialities that are communicated by way of words and abstract universal concepts that are communicated by way of using signs.⁹ Reasoning with the language of signs in the abstract case of the Wason task, can be done within a logical system, whereas reasoning with commonsense knowledge indispensably involves meaning, connections to the world, and some form of consciousness

for in it are effective awareness and freedom of a human being who makes choices and creates structures. [VW 410]

⁸ Valerie E. Stone, Leda Cosmides, John Tooby, Neal Kroll, Robert T. Knight, "Selective Impairment of Reasoning About Social Exchange in A Patient With Bilateral Limbic System Damage," *Proceedings of the National Academy of Sciences* 99/17 (August 2002), 11531– 11536.

⁹ Karl Jaspers, *Philosophische Logik, Erster Band: Von der Wahrheit*, München, DE: R. Piper & Co. Verlag 1947, pp. 408-9, my translation. [Henceforth cited as VW]

Jaspers argues that there is little chance to transfer a language of words into a language of signs; this is only possible if the topic is mathematical.

Since I want to use a logical system in conjunction with large amounts of commonsense knowledge, I would have to combine a language of signs—the language of logic—with an appropriate language of words that are derived from commonsense knowledge (current search-engine knowledge-graphs are estimated to contain more than eighteen billion data points). Subsequently, the word embeddings can then be used for capturing the meaning of words, which could be seen as being an implementation of Jaspers' claim:

The actual meaning of the words does not lie in them alone, but only in the movements of the sentences in which the words illuminate, delimit, and determine each other. [VW 409]

However, word embeddings could create their own reality that is void of existential meaning. Jaspers writes:

Through language, communication becomes possible, which does not occur in ignorant echo and involuntary imitation, but in the intention regarding topic and subject matter. [VW 412]

He cautions that:

What man once made a hard effort to understand through language remains a convenient way of speaking as words and sentences in the mouths of those who follow, who no longer understand. What used to be an expression of profundity turns into usability. People are being taken over by an enormous amount of void and distorted language: They allow themselves to be guided by such language, instead of by what is and what they are; they obtain their education in the form of being capable to speak instead in the form of factual know-how, as a cluster of speaking modes instead of as the formation of their being. [VW 429]

It should be noted that word embeddings can be seen as a statistical database based on the occurrence of words taken together with other words within a very large text corpus. Typically, the representation is a real-valued vector that encodes the meaning of the word in such a way that words that are closer in the vector space are expected to be similar in meaning.

I am expanding upon the commonsense reasoning example displayed in Figure 1 for demonstrating how logical reasoning is used together with word embeddings. Figure 3 shows the logical representation

of the first answer alternative, which is automatically constructed from the natural language sentence. This formula must be evaluated together with the formulas from the problem statement by a first-order logic theorem prover. As mentioned earlier, this is only possible provided that background knowledge about the world is added. I have described elsewhere how a logical formulation is developed in the context of a so-called Hyper Tableaus system that is being used as a reasoner.¹⁰ The activities of a reasoner are then being combined with background knowledge data that are compiled by knowledge graphs, such as, for example, ConceptNet, which is

a knowledge graph that connects words and phrases of natural language with labeled edges. Its knowledge is collected from many sources that include expert-created resources, crowd-sourcing, and games with a purpose. It is designed to represent the general knowledge involved in understanding language, improving natural language applications by allowing the application to better understand the meanings behind the words people use.¹¹

For starting the reasoning process, there is the need of a selection function that extracts the parts of the background knowledge that might help solve the problem. The figure depicts both a syntactic and a similarity selection function. The syntactic selection works with symbol names that occur in the

formulas – it supports reasoning with the language of signs (*notiones generales*), while the similarity selection is concerned with the meaning of the symbols, or the language of words (*notiones communes*) by way of analyzing and evaluating word embeddings in targeted semantic spaces. Figure 3 shows that, according to the similarity score from the word embedding, the word "manducate" is selected due to its similarity with the word "chew," which is contained in the formula at hand.¹²

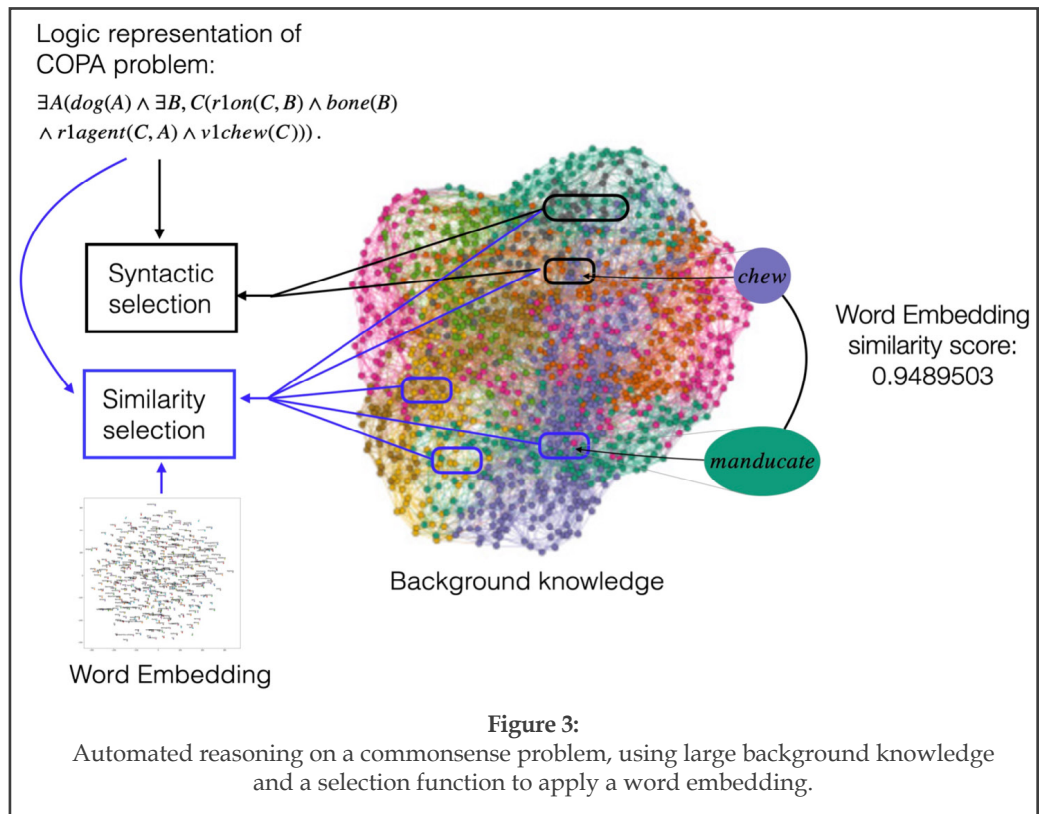


Figure 3: Automated reasoning on a commonsense problem, using large background knowledge and a selection function to apply a word embedding.

The graph shows how the similarity selection function is gauging the meaning of words when selecting relevant background knowledge to help solving a problem. The selected parts of this knowledge are added to the logic formulae which describe the problem and are being fed to the automated reasoning system, that is, to the reasoner.

¹⁰ Peter Baumgartner, Ulrich Furbach, and Björn Pelzer, "Hyper Tableaux with Equality, in *Automated Deduction – CADE-21: 21st International Conference on Automated Deduction, Bremen, Germany, July 17-20, 2007*, ed. Frank Pfenning, Berlin, DE: Springer Verlag 2007, pp. 492-507.

¹¹ Robyn Speer, Joshua Chin, and Catherine Havasi, "ConceptNet 5.5: An Open Multilingual Graph of General Knowledge," *Proceedings of the AAAI Conference on Artificial Intelligence: Thirty-First AAAI Conference on Artificial Intelligence 31/1 (2017)*, 4444-4451, here p. 4444.

¹² The figure is adopted, in part, from Ulrike Barthelmeß, Ulrich Furbach, and Claudia Schon, "Consciousness and Automated Reasoning," in *Proceedings of the 6th Workshop on Formal and Cognitive Reasoning*, eds. Christoph Beierle, Marco Ragni, Frieder Stolzenburg, and Matthias Thimm, *CEUR Workshop Proceedings 2680 (2020)*, 12-26, here p. 20, <https://ceur-ws.org/Vol-2680/paper2.pdf>.

The Theater of Consciousness

The above-described mechanism for processing very large amounts of knowledge turns out to share many features with Bernard Baars' Global Workspace Theory (GWT), which since its conception in the 1980s has gained a prominent place in the fields of psychology and philosophy as one of the theories for explaining consciousness.

The theory departs from the observation that the human brain has a very limited working memory. Humans can actively manipulate approximately seven separate things at the same time within their working memory. This is an astonishing small number of tasks considering the large amount of more than 100 billion neurons in a human brain. Another variable is that human attentiveness is limited to only one single stream of input. For example, one can listen attentively only to one speaker at any one time. This also means, for instance, that one cannot have a thoughtful conversation with a passenger while driving in heavy traffic, and so on. At the same time there are numerous processes running in parallel yet unconsciously so. Global Workspace Theory uses the metaphor of a theater to model how consciousness enables humans to handle the huge amount of knowledge, memories, and sensory input that the brain is controlling at every moment.

Proponents of the Global Workspace Theory assume a theater consisting of a stage, an attentional spotlight shining at the stage, actors which represent the contents, an audience, and some people behind the scene. In a nutshell, the functionality of these components is as follows:

- * The stage. The working memory consists of verbal and imagined items. Most parts of the working memory are inactive, hidden in the dark, yet there are a few active items that are usually contained in the short-term memory.
- * The spotlight of attention. This bright spotlight helps in guiding and navigating through the working memory. Humans can shift it at will, by imagining things or events.
- * The actors. This is the crew of the working memory; crew members are competing against each other to gain access to the spotlight of attention.
- * The audience. The audience represents the vast collection of specialized knowledge. It can be considered as a kind of long-term memory and

consists of specialized properties, which are unconscious. Navigation through this part of the knowledge is done mostly unconsciously. This part of the theater is additionally responsible for interpretation of certain contents such as objects, faces, or speech—there are some unconscious automatisms occurring in the audience.

- * Context behind the scene. Behind the scenes, the director coordinates the show while stage designers and make-up artists prepare the next scenes.

The automated reasoning system discussed in the previous section exhibits remarkable parallels with the Global Workspace Theory. The data structure representing logical formulae corresponds to the stage, while the spotlight represents the mechanism that selects portions of the formulae, the actors, for inferences. Behind the scenes, heuristics and strategies prepare the next logical inference steps, while the audience corresponds to the background knowledge.

Even though the Global Workspace Theory uses a theater metaphor, it is nonetheless very different from a model such as the Cartesian theater in which René Descartes advanced the thesis that there is a certain region within the brain, a homunculus, which accounts for the location of consciousness. In Baars' theater, the entire brain is the theater and hence there is no special location responsible for human consciousness as it is the entire integrated structure which is conscious.

While there are numerous other approaches to understanding consciousness, it is important to note that when discussing this subject matter, the brain cannot be separated from the body, which is situated in the world.

Conclusion

In this essay I show how automated reasoning is a fundamental aspect of symbolic AI by way of exploring the necessity of utilizing extensive background knowledge for solving commonsense problems and by demonstrating that borrowing results from cognitive science regarding human reasoning are helpful toward achieving this end. There is evidence that the meaning of symbols which appear in the logical description of a task can be used to navigate the vast searchable space that is given by the background knowledge. This integration of symbolic and non-symbolic statistical methods can be seen as a step toward combining logical deduction with machine learning techniques.