Logic-Based Modeling of Cognition*

Selmer Bringsjord • Michael Giancola • Naveen Sundar Govindarajulu

ver 1215211015NY

*Trenchant, sagacious feedback from four anonymous reviewers, and from Editor R. Sun, was provided, for which there is much gratitude.

1 Introduction

This chapter explains the approach to reaching the overarching scientific goal of capturing the cognition of persons in computational formal logic.[1] The cognition in question must be coherent, and the person must be at least human-level (i.e., at least a human person).[2] In what can reasonably be regarded to be a prequel to the present chapter, (Bringsjord 2008), a definition of personhood, with numerous references, was provided; for economy here, that definition is not recapitulated. This chapter shall simply take *faute de mieux* a person to be a thing that, through time, in an ongoing cycle, perceives, cognizes, and acts (Sun & Bringsjord 2009).[3] The cognizing, if the overarching goal is to be reached, must be comprised, all and only, of that which can be done in and with computational formal logics. Since it has been proved that Turing-level computation is capturable by elementary reasoning over elementary formulae in an elementary formal logic,[4] any cognition that can be modeled by standard computation is within the reach of the methodology described herein, even with only the simplest logics in the universe $\mathscr{U}$ in Figure 3, and explained below.[5] However, it is important to note a concession that stands at the heart of the logicist research program explained herein: viz. that even if this program completely succeeds, the challenge to

[1]As some readers may know, there is such a thing as *in*formal logic; but the present overview leaves aside this field, entirely. Whatever virtues informal logic may have, because it cannot be used to compute (which is true in turn simply because informal language, the basis for informal logic, cannot be a basis for computing, which by definition is formal), it is of no use to practitioners of logic-based (computational) cognitive modeling. A introduction to and overview of informal logic, which confirms its informal linguistic basis, is provided in (Groarke 1996/2017).

[2]It is of course entirely possible that there exist now or will exist in the future persons who aren't humans; this possibility, as the reader will no doubt well know, is a prominent driver of science-fiction and fantasy literature. In addition, many religions of course claim that there are non-human persons. (In the case of Christianity, e.g. The Athanasian Creed asserts that God is a person.) Even if all such religious claims are false, things clearly could have been such that some of them were true, so the concept of personhood outside of *H. sapiens* is perfectly coherent. In fact, the field of AI, which is intimately bound up with at least computational cognitive science and computational psychology, is a testament to this coherence, since, in the view of many, AI is devoted to building artificial persons (a goal e.g. explicitly set by Charniak & McDermott 1985); see (Bringsjord & Govindarajulu 2018) for a fuller discussion. Finally, it is very hard to deny that humans will increasingly modify their own brains in ways that yield "brains" far outside what physically supports the cognition of *H. sapiens*; see in this regard (Bringsjord 2014).

[3]Cf. the similar cycle given in (Pollock 1995).

[4]There are multiple proofs, in multiple routes. A direct one is a proof that the operation of a Turing machine can be captured by deduction in first-order logic = $\mathscr{L}_1$; e.g. see (Boolos, Burgess & Jeffrey 2003). An indirect route is had by way of taking note of the fact that even garden-variety logic-programming languages, e.g. Prolog, are Turing-complete.

[5]One of the advantages of capturing cognition in formal logic is that it is the primary way to understand computation *beyond* the level of standard Turing machines, something that, interestingly enough, is exactly what Turing himself explored in this dissertation under Alonzo Church, a peerless introduction to which, for those not well-versed in formal logic, is provided by Feferman (1995). For a logic-based, indeed specifically a *quantifier-based*, introduction to computation beyond what a Turing machine can muster, see (Davis, Sigal & Weyuker 1994).

cognitive science of specifying how it is that logic-based cognition emerges from, and interacts with, sub-logic-based processing in such things as neural networks will remain. Theoretically, in the artificial and alien case, where the underlying physical substrate may not be neural in nature, this challenge can be avoided, but certainly in the human case, as explained long ago by Sun (2001), it cannot: humans are ultimately brain-based cognizers, and have a "duality of mind" that spans from the sub-symbolic/neural to the symbolic/abstract.

The remainder of the chapter unfolds straightforwardly as follows. After a brief orientation to logic-based (computational) cognitive modeling (LCCM), the necessary preliminaries are conducted (e.g., it is explained what a logic is, and what it is for one to "capture" some human cognition). Next, three "microworlds" or domains are introduced; this trio is one that all readers should be comfortably familiar with (natural numbers and arithmetic; everyday vehicles, and residential schools, e.g. colleges and universities), in order to facilitate exposition in the chapter. Then the chapter introduces and briefly characterizes the ever-expanding universe $\mathscr{U}$ of formal logics, with an emphasis on three categories therein: deductive logics having no provision for directly modeling cognitive states, *non*-deductive logics suitable for modeling rational belief through time without machinery to directly model cognitive states such as *believes* and *knows*, and finally non-deductive logics that enable the kind of direct modeling of cognitive states absent from the first two types of logic. The chapter's focus then specifically is on two important aspects of human-level cognition that must be modeled in logic-based fashion: the processing of *quantification*, and *defeasible* (or *nonmonotonic*) reasoning. For coverage of the latter phenomenon, use of an illustrative parable involving a tornado is first used, and then turn the chapter turns to the Suppression Task, much studied and commented upon in cognitive science. To wrap things up, there is a brief evaluation of logic-based cognitive modeling, and offered in that connection are some comparisons with other approaches to cognitive modeling, as well as some remarks about the future of LCCM. The chapter presupposes nothing more than high-school mathematics of the standard sort on the part of the readers.

2 Preliminaries

For the goal of capturing the cognition of persons in computational formal logic to be informative to the reader, it is naturally necessary to engage in preliminary exposition to explain what a logic is, what specifically a *computational* logic is, what cognition is herein taken to be, and finally what capturing cognition via formal logic amounts to.

2.1 Anchoring Domains for Exposition: Numbers; Vehicles; Universities

In order to facilitate exposition, it will be convenient to rely upon straightforward reference to three different domains of discourse, each of which will be familiar to the reader: viz., the natural numbers and elementary arithmetic with them, which all readers presumably learned about when very young; everyday vehicles (cars, trucks, etc.); and residential schools, such as colleges and universities.

The natural numbers, customarily denoted by '$\mathbb{N}$,' is simply the set

$$\{0, 1, 2, 3, \ldots\},$$

and 'elementary arithmetic' simply refers to addition, subtraction, multiplication, and so on. Readers are assumed to know for instance that zero $\in \mathbb{N}$ multiplied by $27 \in \mathbb{N}$ is zero. (Later in the chapter, in §4.4, a rigorous, axiomatic treatment of elementary arithmetic, so-called *Peano Arithmetic* will be provided.)

As to the domain of vehicles, the reader is assumed to understand the things represented in Figure 1, which should now be viewed, taking care to read its caption. Three types of familiar vehicles are invoked; each vehicle can be either of two colors (black or grey). Each vehicle is either located at a particular position in the grid shown, or is outside and adjacent to it. The grid is oriented to the four familiar directions, of North, East, South, and West.

What about the domain of residential schools? Here nothing is assumed beyond a generic conception, according to which such institutions, for instance colleges and universities, include agents that fall into the categories of student, teacher, and staff; and include as well that the

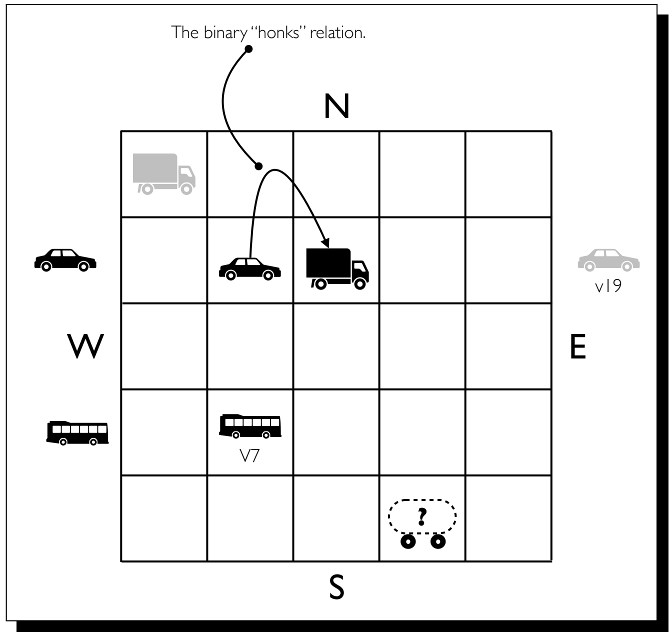

Figure 1: The Vehicular Domain. *The three types of vehicle are shown: cars, box trucks, and buses. The reader will note that there is also a diagram that indicates the existence (and perhaps location) of a "mystery" vehicle; such a vehicle is either a car or a box truck or a bus — but which it is is not conveyed via visual information. Each vehicle is either colored black or grey (there is one grey vehicle in the grid (a box truck), and one such vehicle outside the grid (a car). Notice that vehicles can be denoted by names (or constants). Finally, we have the standard four directions.*

standard buildings are in place in accordance with the standard protocols. For example, residential universities have dormitories, classrooms, and libraries. It is specifically assumed that all readers have common knowledge of the invariants seen in such schools, for instance that they commonly have classes in session, during which time students in the relevant class perceive the teacher, hold beliefs about this instructor, and so on.

2.2 What is a formal logic?

It suffices here to provide two necessary conditions for something's being a formal logic.[6]

The first of these two necessary conditions is that one can't have a formal logic unless one has a

[6] As to an an *in*formal logic, it is not known how to formally define such a thing, and at any rate doing so in anything like a scientific manner is likely conceptually impossible. On the other hand, please note that everything said in the present section is perfectly consistent with conceptions of a formal *inductive* logic, which is distinguished by reasoning that is non-deductive. For a nice, non-technical introduction to inductive logic see (Johnson 2016). For a sustained rigorous introduction to formal inductive logic of the model-theoretic variety, which subsumes probability theory, see (Paris & Vencovská 2015).

formal specification of what counts as a *formula*, and in the vast majority of cases this specification will be achieved by way of the definition of a formal language **L** composed minimally of an alphabet A and a grammar G.[7] Without this, one simply doesn't have a formal logic; with this, one has the ability to determine whether or not a given formal logic is expressive enough to represent some declarative information. Importantly, it is often the case that some natural-language content to be expressed as a formula in some (formal) logic $\mathscr{L}$ cannot be intuitively and quickly expressed correctly by a simple formula in the formal language for $\mathscr{L}$, so that the formula can then be used (for example by a computer program) instead of natural language. For example, the (declarative) natural-language sentence (1_n) "Every car is north of some bus that's south of every truck," which is true in Vehicular Scenario #1 shown in Figure 2, can't be represented in any dialect of the propositional calculus $= \mathscr{L}_{pc}$, since no object variables are permitted in this logic.[8] But this natural-language sentence is easily expressed in first-order logic $= \mathscr{L}_1$ by the following formula in its formal language:

$$(1_l)\ \ \forall x[C(x) \rightarrow \exists y(By \wedge N(x,y) \wedge \forall z(T(z) \rightarrow S(y,z)))].$$

Here x and y are object variables, C is a unary relation symbol used to express being a car, B denotes the property of being a bus, and N is a binary relation symbol that represents the property of being north-of. In addition, we have in $\mathscr{L}_1$ the two standard and ubiquitous quantifiers: Where φ is any object variable, $\exists\varphi$ says that there exists an object φ, and $\forall\varphi$ says that for every φ. The formal grammar of $\mathscr{L}_1$ is not given here, since the level of detail required for doing so is incompatible with the fact that the present chapter is first and foremost an overview of cognitive modeling via logic, not a technical overview of logics themselves. The reader should take care to verify, now, that the formula (1_l) does in fact hold of the scenario shown in Figure 2.

Note that without having on hand a precise definition of the formal language $\mathcal{L}$ that is the basis for a given formal logic $\mathscr{L}$, there is simply no way to rigorously judge the expressive power of

[7]Please note that this pair $\langle A, G\rangle$ needn't be purely symbolic/linguistic. The pair might e.g. include purely visual or "homomorphic" elements. See the logic Vivid as a robust, specified example (Arkoudas & Bringsjord 2009). This issue is returned to at the conclusion of the chapter.

[8]Starting here and continuing through to the end of the chapter, a subscript of $_n$ simply indicates that the proposition so labeled is in **n**atural language, whereas a subscript of $_l$ conveys that the formula so labeled is in some **l**ogic.

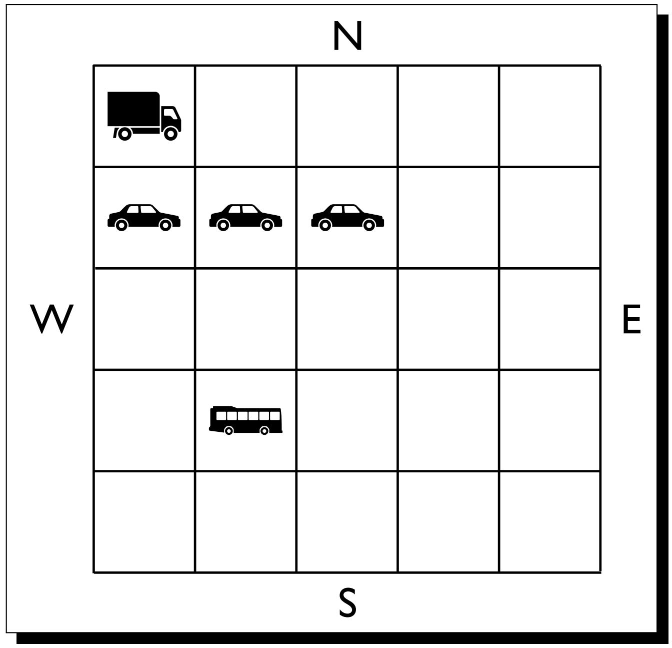

Figure 2: Vehicular Scenario #1

some $\mathscr{L}$ that is being referred to, and hence no way to judge whether $\mathscr{L}$ (or for that matter some theory in cognitive science that purports to subsume $\mathscr{L}$) is up to the task of modeling, say, some proposition that some humans apparently understand and make use of.

Now, what is the second necessary condition for $\mathscr{L}$'s being a formal logic, over and above the one saying that $\mathscr{L}$ must include some formal language? This second condition is disjunctive (*inclusive* disjunction used: i.e. either disjunct, or both, must hold) in nature, and can be stated informally thus:

> Any *bona fide* logic must have a fully specified system for checkable inference (chains of which are expressed as proofs or arguments, where each link in the chain conforms to an inference schema), and/or[9] a fully specified system for checkable assignments of semantic values (e.g., TRUE, FALSE, PROBABLE, PROBABLE AT VALUE (some number) k, INDETERMINATE, etc.) to formulae and sets thereof.

Note that above use was made of truth and falsity in connection with first-order logic = $\mathscr{L}_1$,

[9]Again, this is inclusive disjunction. The two disjuncts represent the two major, sometimes-competing schools in logic, namely proof-theory and model-theory. Proponents of the first school avoid traditional semantic notions. The reason why the disjunction is inclusive is that some logicians would desire to see *both* disjuncts satisfied. In particular, model theorists emphasize semantics, but take proofs to be witnesses of validity of formulas.

since it was said that the formula (1_l) in this formal logic is true on Vehicular Scenario #1. Note as well that the semantic categories for a given logic can often exceed the standard values of TRUE and FALSE. To make this concrete and better understand, take a look back at Figure 1 now, and consider the natural-language statement (2_n) "Car v19 is east of every truck." Expressing this declarative sentence in $\mathscr{L}_1$ as a formula yields

(2_l) $\forall x[T(x) \rightarrow (E(vr19, x) \wedge C(v19))]$,

and what is the semantic value of this formula on the scenario shown in Figure 1? There is simply no way to know, because while we know that vehicle v19 is a car, it's not in the grid. We thus can add the semantic value INDETERMINATE to what we have available for modeling; and this is the value of (2_l) on the scenario in question. For excellent treatment of a trivalent form of $\mathscr{L}_1$, in connection as well with a grid-based microworld, see (Barwise & Etchemendy 1994).

For those in favor of couching formal theories of meaning for natural language (and of cognition relating to the use of natural language) in terms of proof, (2_l) is indeterminate specifically because it can't be proved from the information given in Figure 1, nor can the negation of this formula be proved from this information. However, notice something interesting about the scenario in this figure: Suppose that we knew what kind of vehicle the mystery vehicle in Figure 1 is; specifically, suppose that that vehicle is a bus. In addition, assume that vehicle v19 is located in some square in not the eastmost column, but the column one column to the west of the eastmost column. Given this additional information, we can easily prove (2_l) from the information we have under these suppositions. For some, for instance Francez (2015) (and such thinkers are aligned with the purely inferential understanding of what a logic is within the disjunction given in the second necessary condition above), the meaning of the natural-language sentence (2_n) for an agent consists in its being inferable from what is known by that agent. We spare the reader the formal chain of inference in $\mathscr{L}_1$ that constitutes a formal proof of (2_l). Such a proof is by cases, clearly. The proof starts with noting that v19 will be in one of four different locations in the column in question, and then proceeds to consider each of the only two trucks in the scenario; both of them are west of each of these four locations.

2.3 What is a *computational* formal logic?

Since the topic at hand is cognitive modeling via logic, and cognitive modeling is by definition a computational affair, it is necessary to understand what a computational logic is. All readers will have come to this chapter with at least an intuitive conception of what a logic is (and now, given the foregoing, they will have deeper understanding), but no doubt some will be quite puzzled by the reference to a "computational" logic. This is easy to address: a computational logic is just a logic that can be used to compute, where computing is cast as inference of some sort. Since computing in any form can be conceived of as a process taking inputs to outputs by way of some function that is mechanized in some manner, in the logicist approach to cognition, the mechanization consists in taking inputs to outputs by way of reasoning from these inputs (and perhaps other available content). This is as a matter of fact exactly how logicist programming languages, for instance Prolog, work. Often the inputs are queries, and the outputs are answers, sometimes accompanied by justificatory proofs or arguments. When Newell and Simon presented their system LogicTheorist at the dawn of AI in 1956, at the Dartmouth College, this is exactly what the system did. The logic in question was the *propositional calculus*, the inputs to LogicTheorist were queries as to whether or not certain strings were theorems in this logic, and the outputs were answers with associated proofs. For more details, see the seminal paper of Simon's (1956), for a recent overview of the history to which we refer, in the context of contemporary AI, see (Russell & Norvig 2020, Bringsjord & Govindarajulu 2018).

2.4 What is cognition?

Now to the next preliminary to be addressed, which is to answer: What is cognition? And what is it to cognize? Put another way, this pair of questions distill to this question: What is the target for logicist cognitive modeling?

Fortunately, an efficient answer is available: Cognition can be taken to consist in instantiation of the familiar cognitive verbs: *communicating*, *deciding*, *reasoning*, *believing*, *knowing*, *fearing*, *perceiving*, and so on, on through all the so-called *propositional attitudes* (Nelson 2015). In other, shorter words, whatever cognitive verb is targeted in human-level cognitive psychology, for instance in any major, longstanding textbook for this sub-field of cognitive science (e.g. see Ashcraft & Radvansky 2013), must, if the overall goal of logicist modeling is to be achieved, be captured by what can be done in and with computational formal logics.

2.5 What is it to capture cognition in formal logic?

But how is it known when logicist cognitive modeling of human-level cognition succeeds? Such modeling succeeds when selected aspects of human-level cognition are *captured.* But what is it to "capture" part or all of human-level cognition in computational formal logic? After all, isn't 'capture' operating as a metaphor here, and an imprecise one at that? Actually, the concept of formal logic managing to capture some phenomena is *not* a metaphor; it's a technical concept, one easily and crucially conveyed here without going into its ins and outs: Some phenomena P is captured by some formal content C_P, expressed in a (formal) logic $\mathscr{L}$ if and only if all the elements p in P are such that from C_P one can provably infer in $\mathscr{L}$ the formal counterpart C_p that expresses p. To illustrate with a simple example, suppose that the phenomena in question is the appearance of English declarative sentences (in response, say, to some queries) about elementary arithmetic. So an element here could be (3_n) "Twelve is greater than two plus two," or (4_n)"Seven times one is seven," or (5_n) "Any (natural) number times 1 is that number," and so on. It is known that the particular, familiar formal logic *first-order logic* $= \mathscr{L}_1$ can express such sentences rather easily. For instance, if $\dot{n}$ is a constant in this logic's language to denote the natural number n, and $\times$ is a function symbol in this language for multiplication, the latter two sentences are expressed in $\mathscr{L}_1$ by two formulae (4_l) and (5_l), respectively, like this:

- $(4_l) \coloneqq \dot{7} \times \dot{1} = \dot{7}$
- $(5_l) \coloneqq \forall \dot{n}(\dot{n} \times \dot{1} = \dot{n})$

And now, what of capturing? There is a rather famous body of content, composed of a set of formulae in first-order logic, known as *Peano Arithmetic*, or just **PA**; it captures all of elementary arithmetic.[10] Given what we said above, this means that every relevant sentence s about elementary

[10]Nice coverage is provided in (Ebbinghaus, Flum & Thomas 1994).

arithmetic not only can be expressed by some corresponding formula ϕ_s in $\mathscr{L}_1$, but that every such sentence that's true can be proved from **PA**. This is in fact true of (4_l) and (5_l). Elementary arithmetic has been captured,[11] as has content in other fields outside mathematics.[12] For now, this will do in order to provide the reader with some understanding of the ambition, seen in action below, to capture the defeasible reasoning of human persons. More specifically and concretely, for this ambition to be reached, it must be shown that there is some logic such that, whenever such a person defeasibly reasons to some declarative sentence s, there is some content in that logic from which a formula ϕ_s expressing s can be defeasibly inferred. In the present chapter, this is shown in connection with a reasoning task that has been much studied in cognitive science: namely, the fascinating *suppression task*, introduced by Byrne (1989). This coming discussion will take advantage of the fact some scholars who have worked hard to model and computationally simulate human reasoning and logic, have specifically tried their hand at the suppression task, which appears to clearly call specifically for defeasible reasoning, not just purely deductive reasoning. But before discussing this task and its treatment, some preparatory work must be carried out.

3 The Universe of Logics & This Chapter Located Therein

Please at this point consult Figure 3. This picture is intended to situate the present chapter within the context of the universe of logics that are available for modeling of cognition. There will be no concern here with any logics that permit expressions that are themselves infinitely long; therefore we are working outside the "Infinitary" oval on the left side of the all-encompassing oval shown in Figure 3. (This omission will be returned to in the final section of the chapter.) Hence discussion herein is within the "Finitary" oval shown. Notice that within that oval there are shown two sub-categories: "Intensional" versus "Extensional." Roughly speaking, the first of these categories, which subsumes what are known as *modal logics*, is marked by logics that are tailored to represent

[11] For a technical presentation of the concept of capture, including the arithmetic case just drawn from, see (Smith 2013).

[12] E.g., formal logic has successfully captured major parts of mathematical physics; specifically, e.g., classical mechanics (McKinsey, Sugar & Suppes 1953) and — much more recently — special relativity (Andréka, Madarász, Németi & Székely 2011). In addition, Pat Hayes captured significant parts of everyday, naïve physics in $\mathscr{L}_1$: (Hayes 1978, Hayes 1985).

such cognitive verbs as we cited above: for example, *believing*, *knowing*, *intending*, and also verbs that are "emotion-laden," such as *hoping*, *desiring*, *fearing*, and so on. The logics that are up to the task of representing content that is infused with such — to use again the phrase that has been popular in philosophy — *propositional attitudes* (Nelson 2015) must be sensitive to a key fact arising from the cognition involved: viz., that when an agent has such an attitude toward a proposition, it's not possible to compute compositionally what the semantic value of the overall attitude is from such values assigned to the target propositions. A simple example illustrates this phenomenon:

Consider the proposition p_1 that Umberto believes that Terry believes that Umberto is brilliant. Now suppose that p_2 it's true that Umberto is brilliant. Does it follow from the fact that p_1 is true that p_2 is as well? Clearly not. Umberto may well believe that Terry thinks that he (Umberto) is quite dim. In stark contrast, every logic in the category "Extensional" is such that the semantic values of molecular propositions built on top of "atomic" propositions are fully determined by the semantic values of the atomic propositions. In the very earliest grades of the study of mathematics, this determination is taught to students, because such students, across the globe, are first taught the rudiments of the propositional calculus (shown as $\mathscr{L}_{\text{PropCalc}}$ in Figure 3). In this logic, once one knows the value of sub-formulae within a composite formula, one can directly compute the value of the composite formula. For instance, in $\mathscr{L}_{\text{PropCalc}}$, if p is false and q is false, we know immediately that the value of the composite material conditional $p \rightarrow q$ true.

4 Quantification and Cognition

From the perspective of those searching to capture human-level cognition via logic, there can be little doubt that quantification is a key, indeed perhaps *the* key, factor upon which to focus. Some quantification at work has already been seen above, in connection with both the vehicular domain and elementary arithmetic. Hence the reader is now well aware of the fact that 'quantification' in the sense of that word operative in LCCM has nothing to do with conventional construals of such phrases as "quantitative reasoning." Such phrases usually refer to quantities or magnitudes in some numerical sense. Instead, in formal logic, and in LCCM, quantification refers specifically

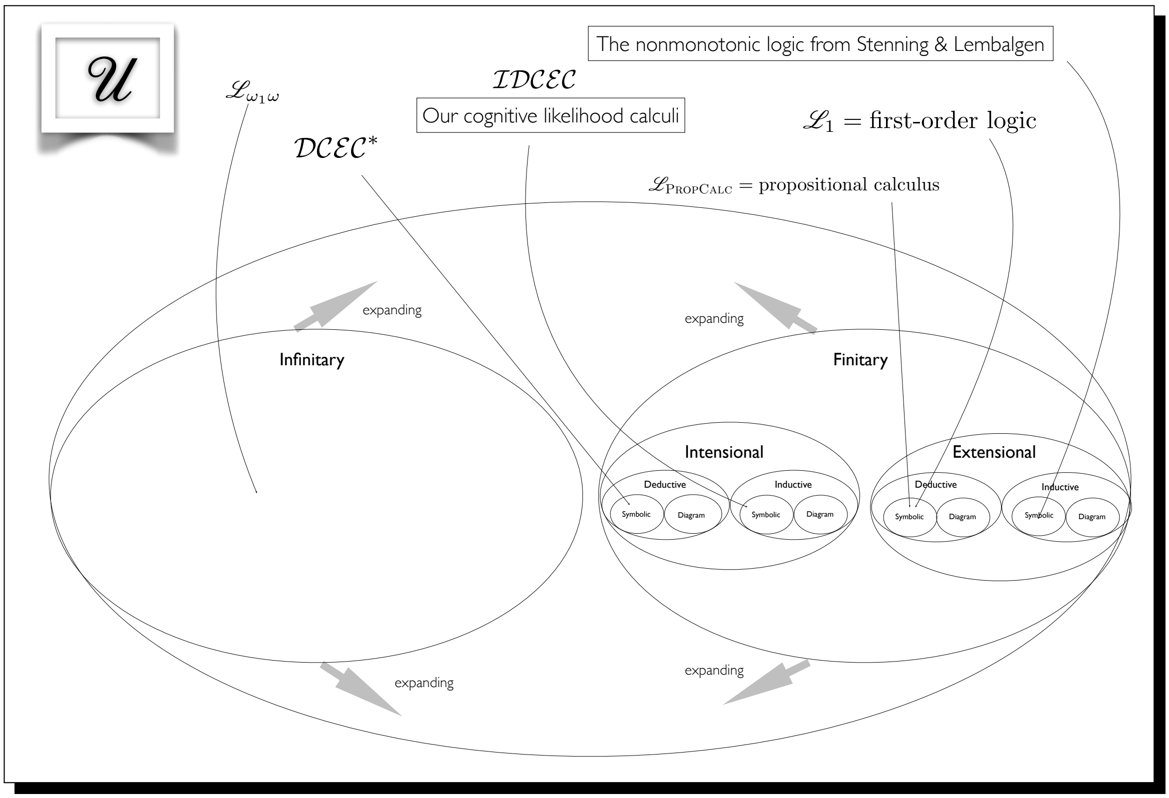

Figure 3: The Ever-Expanding Universe of Logics. *The universe of formal logics can be first divided into those that allow expressions which are infinitely long, and those that don't. Among those that don't, the propositional calculus and first-order logic have been much employed in CogSci and AI. The boxed logics are the ones key to the upcoming analysis and discussion. Note that in the previous section there was crucial use of* $\mathscr{L}_1$.

to the use of of quantifiers such as 'all,' 'some,' 'many,' 'a few,' 'most,' 'exactly three,' and so on. In particular, this chapter has placed and will continue to place emphasis upon the two quantifiers that are used most in at least deductive formal logics, the two quantifiers that (accompanied by some additional machinery) form the basis for most of the formal sciences, including mathematics and theoretical computer science. These two quantifiers are exactly the ones we have already seen in action above: $\forall$ (read as 'for every' or 'for all') and $\exists$ (read as 'there is at least one' or 'there exists at least one'). Again, when these two quantifiers are employed, almost invariably they are immediately followed by an object variable, so that the key constructions are

$$\forall \varphi \ldots$$

and

$$\exists \varphi \ldots,$$

where, as above φ is some object variable, for example x, y, or z. These constructions, as the reader will recall, are read, respectively, as "For every thing $\varphi \ldots$" and "There exists at least one thing φ such that .... The ellipses here are stand-ins for formulae in the relevant formal language.

In our experience, not only students, but also even accomplished researchers outside the formal sciences, are often initially incredulous that something so unassuming as these two constructions could be at the very heart of the formal sciences, and at the very heart of cognition. The chapter now proceeds to explain why such incredulity is mistaken.

4.1 Quantification in the Study of the Mind

As a matter of empirical fact, a focus on quantification in the study of the mind, at least when such study targets human/human-level cognition, has long been established, and is still being very actively pursued. For example, since Aristotle, there has been a sustained attempt to discover and set out a logic-based theory that could account for the cognition of those who, by the production of theorems and the proofs that confirm them, make crucial and deep use of quantification (Glymour 1992). The first substantial exemplar of such cognition known to us in the 21st century remains the remarkable Euclid, whose reasoning Aristotle strove (but failed) to formalize in *Organon* (McKeon 1941), and some of whose core results in geometry are still taught in all technologized societies the world over. In fact, it is likely that most readers will at least vaguely remember that they were asked to learn some of Euclid's axioms, and to prove at least simple theorems from them. If this request met with success, the cognition involved included understanding of quantification (over such things as points and lines, reducible therefore to quantification over real numbers).

What about *contemporary* study of human-level-or-above cognition by way of quantification? Given space restrictions, it is not possible to survey here all the particular research in question; only a few particular examples can be mentioned, before the reader is taken into a deeper understanding of quantification, and from there through a series of aspects of quantification that are important to LCCM.

As to the examples of sample quantification-centric research, Kemp (2009), under the umbrella conception that there is a human "language of thought," advances the general idea that this language is that of a logic, one that appears to correspond to a kind of merging of first- and second-order logic (i.e. $\mathscr{L}_1$ and $\mathscr{L}_2$). He advances as well the specific claim that first-order quantification is easier for the mind to handle than the second-order case. Below, the distinction between first- and second-order quantification is explained, in connection with our vehicular microworld.

As one might expect given how large a role quantification plays in all human natural languages (such as English) as a brute empirical fact (the comma that immediately follows the present parenthetical ends a phrase that has one universal quantifier and one existential one), the connection between linguistic cognition at the human-level and quantification is a deep one. In fact, Partee (2013) argues that quantifiers should be the main pivot around which cognitive linguistics from a formal point of view is pursued. In a particular foray in just this direction, more recently *Understanding Quantifiers in Language* (2009) has explored a connection between different kinds of quantifiers and computational complexity, based in part upon experiments that involve vehicular scenarios of their own (and which in part inspired the somewhat more versatile ones used herein).

It is now time to convey to the reader a deeper understanding of quantification, and the nexus between it and cognition at a number of points, starting with *higher-order* quantification.

4.2 Quantification in Higher-Order Logic

One of the interesting, apparently undeniable, and powerful aspects of human-level cognition is that it centrally involves not only use of relations such as 'is a bus' or 'is a car' (which are of course represented, respectively, by the relation symbols B and C in the vehicular setup), but also relations that can be applied to relations. A body of cognitive-science work indicates this capacity to be present in, and indeed routinely used by, humans (Hummel 2010, Hummel & Holyoak 2003, Markman & Gentner 2001). Using resources of LCCM, specifically a logic from $\mathscr{U}$ well-known to practitioners of logic-based modeling, this aspect of human-level cognition is quite easy to express in rigorous terms. More specifically, LCCM has available to it higher-order logics. First-order logic = $\mathscr{L}_1$, as has been seen above, permits only *object* variables, so named because they refer to objects, not relations (or properties or attributes); the logic $\mathscr{L}_1$ doesn't have *relation* variables. To make this concrete, consider Vehicular Scenario #2 for a minute; this scenario is given in Figure 4. Note, upon studying this scenario, that the immediately following declarative sentence holds in it.

> (6_n) There is at least one relation that holds of every vehicle north of every bus.

Confidence that the reader apprehends the truth of (6_n) in Vehicular Scenario #2 rests on the strength of the cognitive-science work cited above, in the present section. But this natural-language sentence cannot be represented in $\mathscr{L}_1$, since this logic has no provision for expressing "There is a relation that" in this sentence. Second-order logic = $\mathscr{L}_2$ comes to the rescue, because it includes provision for quantification over relation (property) variables. To thus model what the reader apprehends in accordance with LCCM, a formula in second-order logic that expresses (6_n) is needed — and here it is:

> (6_l) $\forall x[(\forall y(B(y) \rightarrow N(x,y))) \rightarrow \exists X X(x)]$

Notice that, following longstanding tradition in formal logic, we use majuscule Roman letters X, Y, Z etc. for variables that can be instantiated with particular relations. Another look at Figure 4 and the vehicular scenario it holds will reveal to the reader that there are particular relations/properties that can serve as particular instances of X in (6_l). For example, one such relation/property is the color grey, which is indeed the color of every vehicle north of every bus.

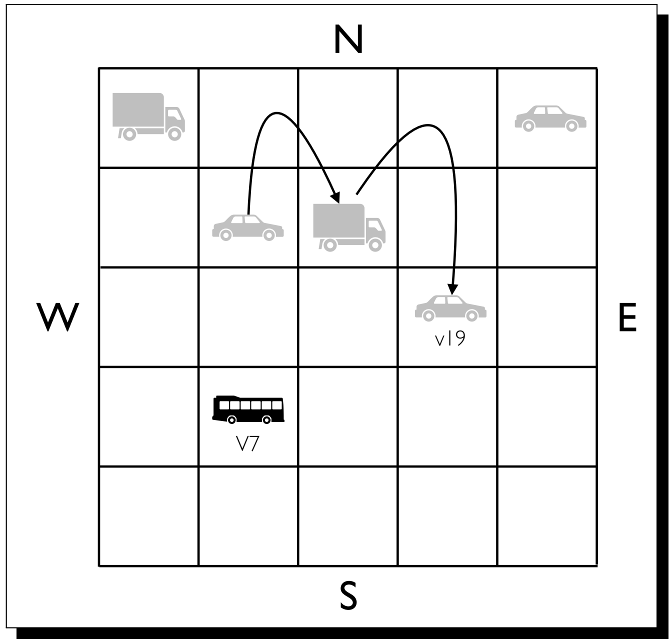

Figure 4: Vehicular Scenario #2. *Observe that in this scenario there is a relation (property) X which every vehicle north of a bus has. E.g., a witness for such an X could in this scenario be the relation 'Grey.'*

The reader may wonder whether there is a level higher than second-order logic = $\mathscr{L}_2$. There is.[13] The next step up, perhaps unsurprisingly, is *third*-order logic = $\mathscr{L}_3$. There are strong reasons to suspect that human-level cognition makes routine use of third-order propositions — though of course it is not known how such propositions are specifically encoded, in the human case, in human brains (but see the use made of Clarion for third-order formulae in Bringsjord, Licato & Bringsjord 2016). The distinguishing new feature of $\mathscr{L}_3$ is that it permits, and renders precise, the ascription of relations/properties to relations/properties; this is not permitted in $\mathscr{L}_2$. This feature can be rendered concrete with help from Vehicular Scenario #2, quickly, as follows. First, simply note that grey is a color; hence we can sensibly write

$$C(G)$$

to represent that fact. Next, to express

[13]That there is, and that plenty of humans have little trouble understanding these higher levels, suggests that the first-versus-second level focus in the aforecited (Kemp 2009) cannot be the centerpoint of the language of thought.

(7_n) There is at least one color property (relation) that holds of every vehicle north of every bus.

the following formulae of $\mathscr{L}_3$ does the trick:

(7_l) $\forall x[(\forall y(B(y) \rightarrow N(x,y))) \rightarrow \exists X(X(x) \wedge C(X))]$

4.3 Quantification and the Infinite

As is well-known, human-level cognition routinely involves infinite objects, structures, and systems. This is perhaps most clearly seen when such cognition is engaged in the learning and practice of mathematics, and formal logic itself. All readers will for example recall that even basic high-school geometry invokes at its very outset infinite sets and structures. As to such sets, we have $\mathbb{N}$ and $\mathbb{R}$, both introduced above, these being two specimens that every high-school graduate needs to demonstrate considerable understanding of. And as to structures based upon these two infinite sets, readers will remember as well that for instance two-dimensional Euclidean geometry is based upon the set of all pairs of real numbers. Within this context, it turns out that cognition associated with even some elementary quantification in $\mathscr{L}_1$ instantly and surprisingly provides an opportunity to zero in on cognition that is compelled to range over infinite scenarios; and an excellent way to acquire deeper understanding of LCCM and its resources is to reflect upon why such scenarios are forced to enter the scene. Notice that so far vehicular scenarios have been decidedly finite in size.

In order to reveal the quantification in question, consider the following three straightforward natural-language sentences pertaining to vehicles:[14]

(8_n) No vehicle honks at itself.

(9_n) If x honks at y and y honks at z, then x honks at z.

(10_n) For every vehicle x, there's a vehicle y x honks at.

This trio is quickly represented, respectively, by the following three extremely simple formulae in $\mathscr{L}_1$:

(8_l) $\forall x \neg H(x,x)$

[14] The discussion here is guided and inspired by a clever example given by Kleene (1967) (p. 292).

(9$_l$) $\forall x \forall y \forall z[(H(x,y) \wedge H(y,z)) \rightarrow H(x,z)]$

(10$_l$) $\forall x \exists y H(x,y)$

Now here is a question: Can a human understand that (8_n)–(10_n), despite their syntactic simplicity, cannot possibly be rendered true by a vehicular scenario that is finite in size? The reader can answer this question, by attempting to build a scenario that does in fact do the trick. A sample try is enlightening. For example, consider the vehicular scenario shown in Figure 5; for the moment, ignore the use made there repeatedly of the ellipsis. The reader should be able to see that the scenario in fact does *not* render (8_l)–(10_l) true, and should be able to see why. In order to construct a vehicular scenario that works, the reader will need to understand that an infinite progression of vehicles will need to be used, with an infinite number of honks. It is not difficult to see that the cognition that discovers and writes down such an infinite scenario can itself be modeled using the resources of LCCM.

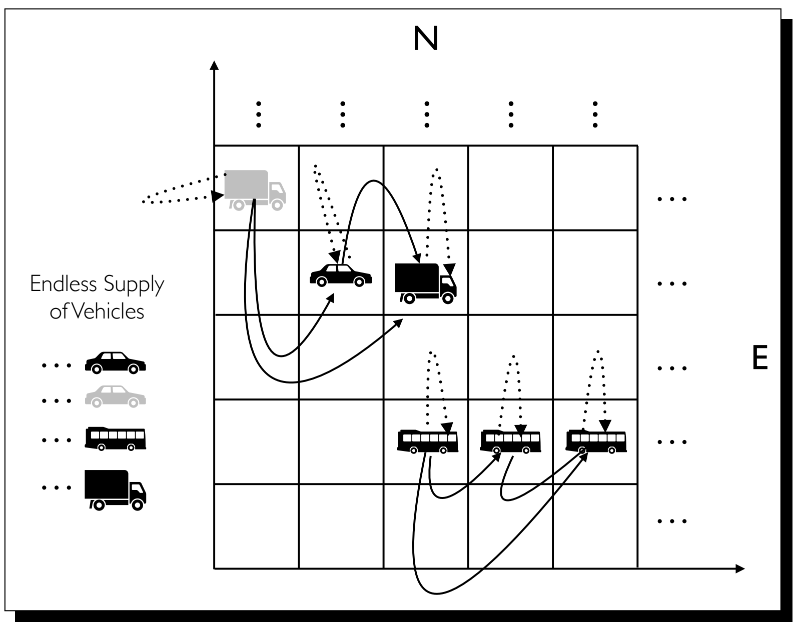

Figure 5: A "Failing" Vehicular Scenario. *The scenario here fails to model the three rather simple quantified formulas specified in the body of the present chapter. The sedulous reader should ascertain why this failure occurs.*

4.4 Quantification as the Heart of the Formal Sciences: Arithmetic and Reverse Mathematics

It is important to share herein that formal logic is the basis for all of human-known mathematics, and that given this, it seems rather likely that if mathematical cognition of the sort that produced/produces mathematics itself (as archived in the form of proved theorems passed from generation to generation) is to eventually be accurately modeled, LCCM will be the key approach to be employed. But the specific, remarkable, and relevant point to quickly make here is that it is quantification that is the bedrock of mathematics. It is the bedrock because mathematics flows from axiom systems whose power and reach are primarily determined by the modulated use of quantification.[15] To see this, we turn to arithmetic, and to the axiom system known as 'Peano Arithmetic' (**PA**), mentioned above but now to be seen in some detail. **PA** consists of the following six axioms, plus one axiom schema (which can be instantiated in an infinite number of ways). Here, the function symbol s denotes the function that, when applied to a natural number $n \in \mathbb{N}$, yields its successor (so e.g. $s(23) = 24$). Multiplication and addition are symbolized as normal.

Axiom 1 $\forall x(0 \neq s(x))$

Axiom 2 $\forall x \forall y(s(x) = s(y) \rightarrow x = y)$

Axiom 3 $\forall x(+(x, 0) = x)$

Axiom 4 $\forall x \forall y(+(x, s(y)) = s(+(x, y)))$

Axiom 5 $\forall x(\times(x, 0) = 0)$

Axiom 6 $\forall x \forall y(\times(x, s(y)) = +(\times(x, y), x))$

Induction Schema Every formula that results from a suitable instance of the following schema, produced by instantiating ϕ to a formula:

$$[\phi(0) \wedge \forall x(\phi(x) \rightarrow \phi(s(x))] \rightarrow \forall x \phi(x)$$

PA, as can be readily seen, once one understands basic quantification, is stunningly simple — so much so that some of the axioms (expressed in natural language) are even taught in elementary

[15] The exact same thing holds for computer science, since e.g. it is layered quantification that defines the hierarchical hardness of computational problems. For instance, both the Arithmetic Hierarchy of increasingly hard computational problems ranging from those a Turing machine can solve and proceeding upward from there (Davis et al. 1994), as well as the Polynomial Hierarchy that gives us the time- and space-wise complexity of Turing-solvable computational problems (Arora & Barak 2009), are based on modulated, layered quantification.

school (where e.g. schoolchildren learn that multiplying any natural number by zero returns zero: Axiom 5). Yet, as simple as it may seem, **PA** is so deep and rich that it cannot be proved consistent by standard, finitary means (this is Gödel's Second Incompleteness Theorem, essentially), and once some of the quantification in **PA** is allowed to to move to the second-order case (recall the brief tutorial above, in §4.2), one arrives at the basis for much of all of mathematics. This is something the field of *reverse mathematics* is based upon, and continues to trace out the consequence arising therefrom. Reverse mathematics is the field devoted to ascertaining what statements in extensional logics pulled from the universe $\mathscr{U}$ suffice to deduce large, particular parts of mathematics. Those wishing to know more about reverse mathematics and the starring role of quantification in this field can consult (Simpson 2010).

5 Defeasible/Nonmonotonic Reasoning

Deductive reasoning of the sort visited above, in connection with both arithmetic and the vehicular microworld, is *monotonic*. To put this more precisely, to say that if a formula ϕ in some logic can be deduced from some set Φ of formulae (written $\Phi \vdash_I \phi$, where the subscript I gets assigned to some particular set of inference schemata for precise deductive reasoning), then for any formula $\psi \notin \Phi$, it remains true that $\Phi \cup \{\psi\} \vdash_I \phi$. In other words, when the reasoning in question is deductive in nature, new knowledge never invalidates prior reasoning. More formally, the closure of Φ under standard deduction (i.e., the set of all formulae that can be deduced from Φ via I), denoted by $\Phi_I^{\vdash}$, is guaranteed to be a subset of $(\Phi \cup \Psi)_I^{\vdash}$, for all sets of formulas Ψ. Inductive logics within the universe $\mathscr{U}$ don't work this way, and that's a welcome fact, since much of real life doesn't conform to monotonicity, at least when it comes to the cognition of humans; this is easy to see:

Suppose — and here is the first reference herein to the domain of residential education — that at present Professor Jones knows that his house is still standing as he sits in it, preparing to teach his class a bit later at his university. If, later in the day, while away from his home and teaching at the university, the Professor learns (along with his students) by notifications pushed to smartphones that a vicious tornado is passing over the town in which his house is located, he has new information that probably leads him to reduce his confidence in the near future as to whether or not his house still stands. Or to take a different example, one much-used in AI (e.g. see the extended treatment in Genesereth & Nilsson 1987), if our Professor Jones knows that Tweety is a bird, he will probably deduce (or at least be tempted to do so) that Tweety can fly, on the strength of a general principle saying that birds can fly. But if Jones learns that Tweety is a penguin, the situation must be revised: that Tweety can fly should now not be among the propositions that Jones believes. Nonmonotonic reasoning is the form of reasoning designed to model, formally, this kind of *defeasible* inference; and some logics within $\mathscr{U}$, all of them non-deductive = inductive in nature, have been devised to specify such reasoning. In the hands of logic-based cognitive modeling, such logics, when computationally implemented and run, can then simulate the kind of human/human-level reasoning just seen in the mind of Professor Jones.

There are many different logic-based approaches that have been designed to allow such modeling and simulation, and each approach is associated with a group of logics. Such approaches include: use of default logics (Reiter 1980), circumscription (McCarthy 1980), and the approach probably most cognitively plausible: argument-based defeasible reasoning (e.g. see for an overview, and an exemplar of the approach, resp.: Pollock 1992, Prakken & Vreeswijk 2001).[16] An excellent survey, one spanning AI, philosophy, and computational cognitive science, the three fields that work in defeasible/nonmonotonic reasoning spans, is also provided in the Stanford Encyclopedia of Philosophy.[17] Because argument-based defeasible reasoning seems to most to accord best with what humans actually do as they adjust their knowledge through time (e.g., Professor Jones and his students, if queried on the spot immediately after the notification of the tornado's path as to whether Jones' house still stands, will be able to provide arguments for why their confidence that it does has just declined), this chapter emphasizes the apparent ability of argument-based defeasible reasoning to capture human/human-level defeasible reasoning. It is in fact a rather nice

[16] From a purely formal perspective, the simplest way to achieve non-monotonicity is to use the so-called *closed world assumption*, according to which, given a set Φ of initially believed declarative statements, what an agent believes after applying the closed world assumption (CWA) to the set is not only what can be deduced from Φ, but also the negation of every formula that *cannot* be deduced. It is easy to verify that it doesn't always hold that $\text{CWA}(\Phi) \subset \text{CWA}(\Phi \cup \Psi)$, for all sets Ψ. I.e., monotonicity doesn't hold. Unfortunately, while this is a rapid route to non-monotonicity, CWA isn't cognitively plausible, at all. To see this, consider the parabular Professor Jones and suppose without loss of generality that he is not a professional logician or mathematician, and hence cannot deduce, say, Gödel's famous first incompleteness theorem (= G1). By CWA, Jones should believe that G1 is false!

[17] At http://plato.stanford.edu/entries/logic-ai

thing about humans and defeasible reasoning that they are often able to explain, and sometimes show, by articulating arguments, why their beliefs have changed through time as new information is known or at least believed, where that new information leads to the defeat of reasoning that they earlier affirmed.

Now, returning to the tornado example, what is the argument that Professor Jones might give to support his belief that his house still stands, while he is in the classroom? There are many possibilities, one respectable one is what can be labeled 'Argument 1,' where the indirect indexical refers of course to Jones:

(11) I perceive that my house is still standing.

(12) If I perceive ϕ, ϕ holds.

$\therefore$ (13) My house is still standing.

The second premise is a principle that seems a bit risky, perhaps. No doubt there should be some caveats included within it: that when the perception in question occurs, Jones is not under the influence of drugs, not insane, and so on. But to ease exposition, such clauses are left aside. So, on the strength of this argument, let us assume that Jones' knowledge includes (13), at time t_1.

Later on, as has been said, the Professor finds himself in class at his university, away from home. Jones and his students quickly consult smartphone weather apps and learn that the National Weather Service reports this tornado to have touched down somewhere in the town T in which Jones' house is located, and that major damage resulted; in particular, some houses were tragically leveled. At this point (t_2, assume), if Jones were pressed to articulate his current position on (13), and his reasoning for that position, and he had sufficient time and patience to comply, he would likely offer something like this (Argument 2):

(14) A tornado has just (i.e., at some time between t_1 and t_2) touched down in T, and destroyed some houses there.

(15) My house is located in T.

(16) I have no particular evidence that my house was *not* struck to smithereens by a tornado that recently passed through the town in which my house is located.

(17) If a tornado has just destroyed some houses in (arbitrary) town T', and house h is located in T', and one has no particular evidence that h is not among the houses destroyed by the tornado, then one ought not to believe that h wasn't destroyed.

∴ (18) I ought not to believe that my house is still standing. (I.e., I ought not to believe (13).)

Assuming that Jones meets all of his "epistemic obligations" (in other words, assuming that he's rational), he will not believe (13) at t_2. (Actually, and below this is dealt with this more plausible modeling, it's more reasonable to imagine that Jones does still believe (13), but that the *strength* of his belief has declined.) Therefore, at this time, (13) will no longer be among the things he knows. (If a cognitive system s doesn't believe ϕ, it follows immediately that s doesn't know ϕ, in the sense of 'know' with which we are concerned with.) The nonmonotonicity here should be clear.

The challenge to LCCM is to devise formalisms and mechanisms that model this kind of mental activity through time. The argument-based approach to nonmonotonic reasoning does this. As to how, the main move is to allow one argument to invalidate another (and one argument to invalidate an argument that invalidates an argument, which revives the original, etc.), and to keep a running tab on which propositions should be believed at any particular time. Argument 2 above rather obviously invalidates Argument 1; this is the situation at t_2. Should Jones then learn that only two houses in town T were leveled, and that they are both located on a street other than his own, Argument 2 would be defeated by a third argument, because this third argument would overthrow (16). With Argument 2 defeated, (13) would be reinstated, and back in what Jones knows. Clearly, this ebb and flow in argument-versus-argument activity is provably impossible in straight deductive reasoning.

5.1 An Argument-Adjudication System for Defeasible Reasoning

In order to adjudicate competing arguments, such as those in the tornado example of §5, a system for quantifying the level of subjective uncertainty of declarative statements needed. To obtain this, let us invoke a system based upon *strength factors* first presented in (Govindarajulu & Bringsjord 2017). This work was in turn directly guided by a simpler and smaller system of strength-indexed belief invented over half a century ago by Chisholm (1966).[18] While recently specification of a more robust formal inductive logic ($\mathcal{IDCEC}$; note that it is located within $\mathscr{U}$, as Figure 3 indicates) for such processing, accompanied by an implementation and demonstration, had been achieved (Bringsjord, Govindarajulu & Giancola 2021), the survey nature of the present chapter means that a "higher altitude" level of detail is prudent, and in what now follows the chapter stays at that altitude. For more details, the reader can consult the lengthy technical survey provided by Prakken & Vreeswijk (2001).

The strength factors to now be employed consist of 13 values (see Figure 6) that can be used to annotate statements expressing belief or knowledge. For example, one can formalize the sentence "Jones believes it is *more likely than not* at time t_0 that his house is still standing." by the formula $\mathbf{B}^1(jones, t_0, \text{Standing}(home))$.

Note at this point that the introduction of uncertainty measures already forces a move beyond deductive reasoning into inductive reasoning and logics, as with such measures one can no longer be producing proofs, but instead, arguments. While a proof guarantees the truth of the formula it proves (as long as the axioms/premises are true), an argument only provides some level of strength that its conclusion is true. Hence, in moving from deductive reasoning to *in*ductive reasoning, such arguments are able to be expressed. The reader may at this point wish to note that in Figure 3

[18]There are formal logics that subsume probability theory, and theoretically they could be deployed to model the tornado scenario (e.g. there is *uncertain first-order logic*; see Núñez, Murthi, Premaratne, Bueno & Scheutz forthcoming). However, it doesn't seem cognitively plausible that Professor Jones (consciously) associates real numbers between 0 and 1 with the proposition that his house is still standing. One could also explore using so-called "fuzzy logic," which emerged out of fuzzy sets first introduced by Zadeh (1965). But here one must be very careful. Most of the things called "fuzzy logics" are not in fact logics at all, and are not in the universe $\mathscr{U}$. The advent of *bona fide* formal fuzzy logics, replete with formal languages, inferential machinery, and so on, came by way of the groundbreaking (Hájek 1998).

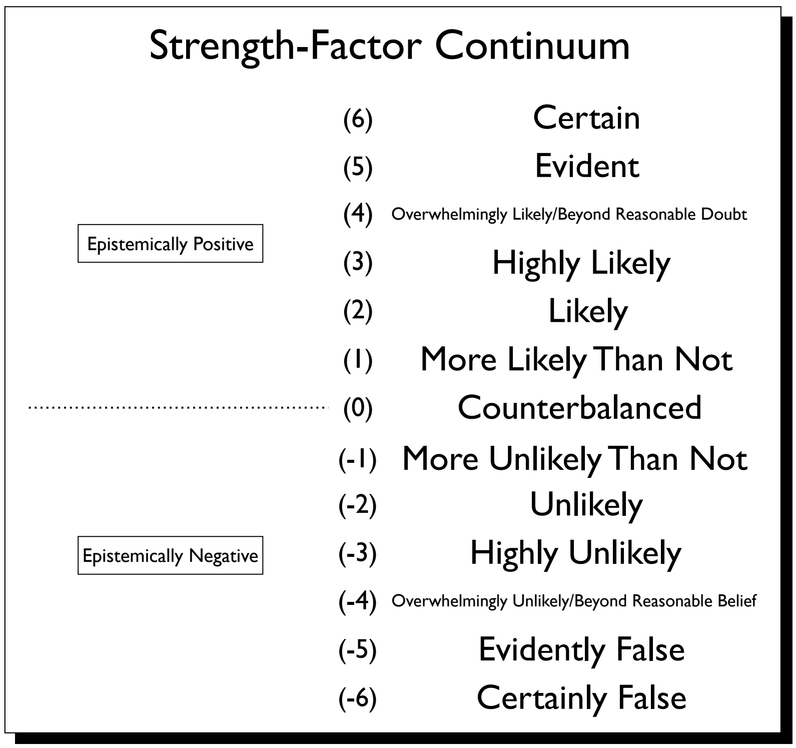

Figure 6: The Current Strength Factor Continuum. *The center value,* counterbalanced, *indicates that there is no evidence for or against belief in the subformula. Increasing positive and negative values indicate increasing and decreasing likelihood of truth in the subformula, respectively.*

inductive logics are denoted. For a recent introduction to inductive logic as an argument-based, as opposed to a proof-based, affair, the reader can consult (Johnson 2016).

Two intensional logics are will be brought to bear, both suitable for the type of modeling we need in the tornado scenario. Because the distinguishing purpose of these logics and others like them is the modeling of human-level cognitive states (such as believing and knowing a proposition at a time), and human-level reasoning, some have long referred to these logics as *cognitive calculi*, and this suit is followed here. The first cognitive calculus used here is for purely deductive reasoning; the second supports inductive reasoning. For the encapsulated formal specification of these cognitive calculi, see (Bringsjord et al. 2021). Industrious readers can find these two calculi in the universe $\mathscr{U}$ pictured in Figure 3; they are named therein as $\mathcal{DCEC}^*$ and $\mathcal{IDCEC}$; the first is a deductive intensional logic, the second an inductive intensional logic.

Note that when arguments are referred to in the present chapter, it is meant more specifically *formal* arguments. Hence, like in any respectable proof, each step must be sanctioned by the deployment of an inference schema.[19]

When one has multiple such arguments, each of which concludes with the affirmation or rejection of belief in some subformula, the adjudication process is simple: select the argument whose conclusion has the highest strength. This method will be employed in §5.2 to formalize and rigorously model the tornado example first given in §5. More complex adjudication methods for more complex sets of arguments (e.g., where the adjudication process may need to select out sub-arguments from multiple arguments in order to construct the winning argument and corresponding final conclusion) are the focus of active research outside the scope of the present chapter.

5.2 The Tornado Conquered

Consider again the following scenario, now made a bit more determinate. Professor Jones left his home (at time t_{home}) to go to his university, and while there (at time t_{work}) he learns there the disturbing news and discovers that a tornado has passed through the town (at time $t_{tornado}$) in which his house is located (*town*). Again, but in search of more precision, what should the

[19] For the relevant lists of such inference schemata, which are outside the scope of this overview chapter, the reader is directed to (Bringsjord et al. 2021).

Professor now believe with regard to whether or not his house is still standing?

This problem can be posed in the argument-adjudication framework employed here for defeasible/nonmonotonic reasoning in order to evaluate the strength of each argument and thereby allow Jones to arrive at a final belief-fixation decision. First, consider an argument Jones might plausibly use to justify his belief that his house is standing at the time that he is about to leave for work, t_{home}, an argument that is now more nuanced and plausible that what we laid about above:

	(19) $\mathbf{P}(jones, t_{home}, Standing(home))$	Jones perceived that his home was standing when he left for work.
∴	(20) $\mathbf{B}^5(jones, t_{home}, Standing(home))$	Assuming Jones was not dreaming or hallucinating, perception generates *evident* beliefs. Therefore, Jones believed it was *evident* that his home was still standing at that time.
∴	(21) $\mathbf{O}(jones, t_{home}, \mathbf{B}^5(jones, t_{home}, Standing(home)))$	Hence Jones ought to believe it is *evident* at time t_{home} that his house is still standing.

Table 1: **Argument 1:** Jones determines he ought to believe it is *evident* that his house is still standing at time t_{home}.

Here the obligation operator is of an intellectual variety; there is no reference here to anything like moral obligations and deontic operators that are at the heart of deontic logic, which is devoted to formalizing human moral reasoning. That one *ought* to believe ϕ here means that there is a rational argument compelling one to believe ϕ as a rational agent. This basic notion of intellectual obligation as part and parcel of an abstract conception of rationality is at the heart of the logic and mathematics of inductive logic (Paris & Vencovská 2015).

Next, consider another sequence of reasoning Professor Jones might go through while driving to work (at time $t_{driving}$). Since he is no longer perceiving his home, his belief cannot be at the level of *evident*. However, his previous belief can persist at the next level down, *overwhelmingly likely*, so long as Jones has not been made aware of any information to the contrary since then.

	(22)	$\neg\mathbf{P}(jones, t_{driving}, Standing(home))$	Jones no longer perceives his home.
$\therefore$	(23)	$\neg\mathbf{B}^5(jones, t_{driving}, Standing(home))$	Hence, Jones no longer believes it is *evident* that his home is still standing.
$\therefore$	(24)	$\mathbf{O}(jones, t_{driving}, \mathbf{B}^4(jones, t_{driving}, Standing(home)))$	Assuming Jones' memory is reasonably reliable, and since he has no information to the contrary, he ought to believe it is *overwhelmingly likely* at time $t_{driving}$ that his house is still standing.

Table 2: **Argument 2:** Jones retracts his previous belief that he ought to believe it is *evident* that his house is still standing at time $t_{driving}$, and replaces it with a belief at the level of *overwhelmingly likely*.

Finally, at t_{work}, Jones becomes aware of the tornado which just passed through his town. Therefore he is rationally obligated to retract his previous belief, and replace it with a weaker belief that his house is still standing:

	(25)	$\mathbf{K}(jones, t_{work}, LocatedIn(home, town))$	Jones knows his home is located in his town.
	(26)	$\mathbf{S}(news, jones, t_{work}, TornadoPassedThrough(town, t_{tornado}))$	Jones heard from the news that a tornado passed through the town where his home is located.
	(27)	$\mathbf{K}(jones, t_{work}, \forall h\ a\ t\ (TornadoPassedThrough(a, t) \wedge LocatedIn(h, a)) \rightarrow \Diamond\neg Standing(h))$	Jones knows that if a tornado passes through an area where a home is located, it is possible that that home is no longer standing.
$\therefore$	(28)	$\mathbf{K}(jones, t_{work}, \Diamond\neg Standing(home)$	Hence Jones knows it is possible that his home is no longer standing.
$\therefore$	(29)	$\neg\mathbf{B}^4(jones, t_{work}, Standing(home))$	Hence Jones no longer believes it is overwhelmingly likely that his home is still standing.
$\therefore$	(30)	$\mathbf{O}(jones, t_{work}, \mathbf{B}^2(jones, t_{work}, Standing(home)))$	However, since Jones has only evidence indicating a possibility that his home has been destroyed, he ought to believe it is *likely* at time t_{work} that his house is still standing.

Table 3: **Argument 3:** Jones determines he ought to believe it is *likely* that his house is still standing at time t_{work}.

Discussion of the tornado case study is now complete. At this point, the chapter turns from this informal, illustrative study to the suppression task, which was been explored by way of experiments reported in the cognitive-science literature.

5.3 The Suppression Task

The task in question is reported in (Byrne 1989). Three groups of subjects were asked to select which proposition from among a trio of them "follows"[20] from a set of suppositions. Each group of subjects was given a different set of suppositions. Group 1 (= G1) was given this pair of suppositions:

(s1) If she has an essay to finish, then she will study late in the library.

(s2) She has an essay to finish.

This group's options to select from were the following three:

(o1) She will study late in the library.

(o2) She will not study late in the library.

(o3) She may or may not study late in the library.

Among G1, 96% selected (o1). G2 was given suppositions consisting of (s1) and (s2), plus the following supposition:

(s3) If she has a textbook to read, then she will study late in the library.

In G2, again 96% of its members selected option (o1). G3 received (s1) and (s2), plus this supposition:

(s4) If the library stays open, then she will study late in the library.

This time things turned out quite differently: only 38% of G3 selected (o1).

From the perspective of standard zero-order logic = $\mathscr{L}_0 \in \mathscr{U}$,[21] which can accordingly be assumed here to have any standard proof theory, such as is used in early classical mathematics (e.g.

[20]Unfortunately, 'follows' is a metaphor here — but it's the term Byrne (1989) used. No firm conception of what this term means is available. From the standpoint of formal logic, what should have been said to subjects is something like 'must necessarily be deducible,' because (i) the hallmark of deduction since first systematically investigated by Aristotle has been apprehended as the fact that when deduction from givens/premises/suppositions to (a) conclusion(s) is valid, the former *necessarily* entail the latter, and because (ii) plenty of conclusions are thought by rational agents operating rationally to follow from givens/premises/suppositions that certainly don't necessitate these conclusions (e.g., consider a case in which a conclusion follows from premises by statistical syllogism). However, this being said, for now, the unfortunate use of 'follows' by Byrne (1989) must be left aside.

[21]Obtained by augmenting the formal language of the propositional calculus with provision for relation and function symbols, and the identity symbol =; but no quantifiers are allowed. Like the propositional calculus, $\mathscr{L}_0$ is Turing-decidable; not so any n-order logic $\mathscr{L}_n$ in $\mathscr{U}$, where n is a positive integer.

high-school mathematics in every technologized society/nation), this result is interesting, since, to begin, in $\mathscr{L}_0$ we might represent the declarative sentences (s1), (s2), (s3), and (s4) as follows, where a represents the female agent in question:

(s1*) $\mathit{ToFinish}(a) \rightarrow \mathit{LateLibrary}(a)$

(s2*) $\mathit{ToFinish}(a)$

(s3*) $\mathit{ToRead}(a) \rightarrow \mathit{LateLibrary}(a)$

(s4*) $\mathit{StaysOpen} \rightarrow \mathit{LateLibrary}(a)$

Next, following suit, the options would be represented thus:

(o1*) $\mathit{LateLibrary}(a)$

(o2*) $\neg \mathit{LateLibrary}(a)$

(o3*) $\neg \mathit{LateLibrary}(a) \vee \mathit{LateLibrary}(a)$

With these representations, easy-to-find proofs in $\mathscr{L}_0$ certify that

$$\{(\text{s1}^*), (\text{s2}^*), (\text{s3}^*)\} \vdash (\text{o1}^*). \qquad (+)$$

However, there is no available proof in this logic of option two from the first three suppositions; that is:

$$\{(\text{s1}^*), (\text{s2}^*), (\text{s3}^*)\} \nvdash (\text{o2}^*). \qquad (-)$$

Option (o3*) is a theorem in this logic, so it's provable from {(s1*), (s2*), (s3*)}.[22] Because we are dealing here with standard deductive reasoning, which as has been noted is non-feasible/monotonic, adding one or both of (s3*), (s4*) to {(s1*), (s2*), (s3*)} doesn't change provability/unprovability; that is, neither (+) nor (–) change. This is why group G3's behavior is odd and interesting from the point of view of $\mathscr{L}_0$, and hence from the point of view of the cognitive science of reasoning. Clearly, the formal modeling just given via $\mathscr{L}_0$ doesn't match what most of the subjects in this group were thinking when they responded.

[22] As a matter of fact it's not appropriate to represent (o3) as having the form $\phi \vee \neg\phi$, but this issue is left aside here.

5.3.1 Stenning & van Lambalgen's Extensional Treatment of the Suppression Task

Byrne, in her presentation of the suppression task (Byrne 1989), argues that the findings of her study imply that people don't strictly apply valid methods of logical deduction when reasoning. Therefore, so her diagnosis goes, logic is not sufficient for modeling human reasoning. She states that "...in order to explain how people reason, we need to explain how premises of the same apparent logical form can be interpreted in quite different ways" (Byrne 1989).

Stenning & van Lambalgen (S&V) (2008) formalize this concept of what can be called "premise interpretation."[23] They claim that humans, when presented with a set of premises and possible conclusions, first reason *toward* some rational interpretation of the premises, then *from* that interpretation to some conclusion. They formalize this process in a Horn-style[24] propositional logic, supplemented with a formalization of the Closed World Assumption (CWA). [25] Given this context, when presented with a set of assumptions and a conclusion to prove, S&V follow this three-step algorithm:

1. Reason to an interpretation.
2. Apply nonmonotonic closed-world reasoning (i.e., apply CWA) to the interpretation produced by 1.
3. Reason from the result of what step 2. produces.

Let us now consider the application of these three steps to the first experiment in Byrne's (1989) study, but first we need to have handy here again the stimuli presented to subjects. In her first experiment, subjects are given the two suppositions

[23]S&V are not the only LCCMers who have tried their hand at modeling ST: Dietz et al. previously took two distinct logic-based approaches to modeling the Suppression Task. In their first approach, they used a three-valued Łukasiewicz logic which allows the expression of a third truth-value beyond *true* and *false*: *unknown* (Dietz, Hölldobler & Ragni 2012, Dietz, Hölldobler & Wernhard 2014). More recently, they have taken an approach which aims to model the suppression task in a more cognitively-plausible way: (Saldanha & Kakas 2020). Their framework, *cognitive argumentation*, formalizes methods of reasoning used by humans (which may or may not be logically sound) as *cognitive principles*. For example, their "Maxim of Quality" expresses that we (humans) typically assume statements we are told are true if we don't have a reason to believe otherwise (e.g. that the speaker may be lying or incompetent). In the context of the suppression task, the Maxim of Quality dictates that the subjects will assume that all of the statements made by the experimenters are true (e.g. "She has an essay to finish.").

[24]Horn-style logics have formal languages permitting conditionals only of a highly restricted sort; details are left aside. The programming language Prolog, mentioned above, is for example based upon a Horn-style fragment of first-order logic $= \mathscr{L}_1$. Prolog programs are frequently called "logic programs," and as the reader will soon see, S&V call a key part of their modeling of the suppression task "logic programs."

[25]Recall that, in a word, CWA is the assumption that everything about a domain is known. Formally, as explained above, any proposition which is not known to be true (or not provable) is assumed to be false.

(s1) If she has an essay to write, she will study late in the library.

(s2) She has an essay to write.

and are then asked to choose from the following set of conclusions which one follows from the premises.[26]

(o1) She will study late in the library.

(o2) She will not study late in the library.

(o3) She may or may not study late in the library.

Now comes the application of the three-step algorithm.

5.3.1.1 The Algorithm, Applied

Step 1: Reasoning to an Interpretation The first part of this step is appending the antecedent of every conditional with "$\neg ab$," where this addition, intuitively, means "no abnormalities." The idea here is that people interpret the conditional $p \rightarrow q$ as $(p \wedge \neg ab) \rightarrow q$. That is, p implies q, *if* no external factors of which the subject is currently unaware (i.e. the abnormalities represented by ab) subvert the implication.

The last part of this step is to collect the assumptions as modified above into a set which S&V refer to as the *logic program* corresponding to the assumptions. Given the foregoing, the output of Step 1 for Experiment 1 would be the set:

$$\{\ EssayToWrite; EssayToWrite \wedge \neg ab \rightarrow StudyLateInLibrary\ \} \tag{1}$$

Step 2: Applying Nonmonotonic Closed-World Reasoning to the Interpretation This step also consists of two sub-parts. First, for all atoms q in the logic program produced in Step 1, if there is no antecedent p such that $p \rightarrow q$, the conditional $\bot \rightarrow q$ is added to the logic program. Note that in S&V's logic, the meaning of an atom p in the assumption base is really

[26]Note again that Byrne uses the informal term 'follows' and not one necessitating formal entailment like 'logically deduces.'

$\top \to p$; but for clarity, they typically just write p; the same is done here. Therefore, in the example above, the only atom for which this step applies is ab; hence the conditional $\bot \to ab$ is added to the logic program:

$$\{\ EssayToWrite; EssayToWrite \wedge \neg ab \to StudyLateInLibrary; \bot \to ab\ \} \tag{2}$$

The second part of Step 2 is what S&V refer to as *constructing the completion* of the logic program. This involves first joining all implications $\phi_i \to q$ (i.e. those implications whose consequent is q) into a single implication $\vee_i \phi_i \to q$.[27] Second, all conditionals are converted to biconditionals. Therefore the final logic program (also, the interpretation of the premises) is:

$$\{\ EssayToWrite; EssayToWrite \wedge \neg ab \leftrightarrow StudyLateInLibrary; \bot \leftrightarrow ab\ \} \tag{3}$$

Step 3: Reasoning from the Result of Step 2 The third and final step is fairly straightforward: the subject reasons from the final set of premises using the inference rules of standard propositional logic. Notice that, because $\bot \leftrightarrow ab$, we have $\top \leftrightarrow \neg ab$; hence the logic program above can be simplified to:

$$\{EssayToWrite; EssayToWrite \leftrightarrow StudyLateInLibrary\} \tag{4}$$

Finally, it is obvious that from these premises one can deduce *StudyLateInLibrary*. Note that while the conclusion was obvious in this case, this method of reasoning to and from an interpretation matches the reasoning process of the majority of people in all of Byrne's experiments. Next follows a walk-through of S&V's algorithm for a slightly more complicated (and more interesting) case, in which an additional premise is introduced.

[27] There are no instances of this in this example, but there will be in the next.

5.3.1.2 Applying the Algorithm to the Additional-Premise Case

In the second experiment, recall, Byrne gave her subjects the following set of premises:

> If she has an essay to write, she will study late in the library.
>
> If the library stays open, she will study late in the library.
>
> She has an essay to write.

This additional premise is modeled using the same form as the original two premises:

$$LibraryOpen \wedge \neg ab' \rightarrow StudyLateInLibrary \tag{5}$$

However, in this case, S&V also (naturally) add the following premise:

$$\neg LibraryOpen \rightarrow ab \tag{6}$$

This premise is intended to model the belief of those who believed that *modus ponens* applied in Experiment 1, but not in Experiment 2. (In other words, the introduction of the additional premise suppressed their belief.) More specifically, this conditional states that if the library is not open, then it would be abnormal for her to go to study late in the library. The symmetric condition $\neg EssayToWrite \rightarrow ab'$ can also be added; that is, if she does not have an essay to write, it would be abnormal for her to study late in the library.[28]

Now, performing Step 1 will produce the program:

$$\left\{ \begin{array}{c} EssayToWrite \wedge \neg ab \rightarrow StudyLateInLibrary \\ LibraryOpen \wedge \neg ab' \rightarrow StudyLateInLibrary \\ EssayToWrite \\ \neg LibraryOpen \rightarrow ab \\ \neg EssayToWrite \rightarrow ab' \end{array} \right\} \tag{7}$$

[28]This is not necessary but will allow for a simplification of the final result.

Next, applying nonmonotonic closed-world reasoning yields:

$$\left\{ \begin{array}{c} (EssayToWrite \wedge \neg ab) \vee (LibraryOpen \wedge \neg ab') \leftrightarrow StudyLateInLibrary \\ EssayToWrite \\ (\bot \vee \neg LibraryOpen) \leftrightarrow ab \\ (\bot \vee \neg EssayToWrite) \leftrightarrow ab' \end{array} \right\} \tag{8}$$

And next, using standard logical deduction for the propositional calculus, we can simplify this set to:

$$\{EssayToWrite; (EssayToWrite \wedge LibraryOpen) \leftrightarrow StudyLateInLibrary\} \tag{9}$$

Finally, the subject reasons from this interpretation of the premises. Note that the second statement says "She will study late in the library if and only if she has an essay to write and the library stays open." Since the premise set doesn't include the proposition *LibraryOpen*, one cannot deduce *StudyLateInLibrary*. This result matches the common human intuition[29] that the additional premise hinders the successful application of *modus ponens* to the original premises.

5.3.2 Modeling the Suppression with *Intensional* Logic

It is now quickly demonstrated that the human reasoning in the suppression task can be easily and efficiently modeled in a way simpler than that employed by S&V. In this alternate route, (a) take timepoints are taken seriously within the narrative that are implicit in what the subjects are given; and (b), use is made of these timepoints in connection with a simple intensional logic that includes (i) a way to represent and reason with what is *known* and what is *believed*, and (ii) includes an operator for what is *possibly* the case.[30]

This first step in carrying out these two steps is to simply announce a simple set of symbols used

[29] I.e., the intuition of the majority of the people in Byrne's study.

[30] Thus, use is made of basic constructs from *epistemic* logic (Hendricks & Symons 2006), which formalizes attitudes like *believes* and *knows*; and also basic constructs from *alethic modal logic* (Konyndyk 1986), which formalizes concepts like *possibly* and *necessarily*. Epistemic logic are intensional logics within the universe $\mathscr{U}$.

to enable the formulae that express what is presented to subjects in the suppression task. This is done by way of the following table, which simply presents the referent in each case intuitively, so that no technical specifications are needed.

Table 4: Symbols for Intensional Mod & Sim of Suppression Task

Symbol	**Referent**
s (object variable)	student
e (object variable)	essay
b (object variable)	book
$t, t', \ldots$ (object variables)	timepoints
t_1 (constant)	the particular, initial timepoint
ℓ (constant)	the library
$\mathbf{a}, \mathbf{b}$ (constants)	two particular agents
$>$ (2-place relation)	later than
$\mathit{ToFinish}(s, t, e)$ (3-place relation)	s at t has e to finish
$\mathit{NearFuture}(t', t)$ (2-place relation)	t' is in near future of t
$\mathit{LateLibrary}(s, t)$ (2-place relation)	s works late in the library at t
$\mathit{Open}(\ell, t)$ (2-place relation)	the library is open at t
$\mathit{ToRead}(s, t, b)$ (3-place relation)	s at t has textbook b to read
$\Diamond$ (alethic operator)	'possibly'
$\mathbf{B}_x$ (epistemic operator)	agent x believes that
ToRead(s, t, b) (3-place relation)	s has at t to read b
$\mathbf{K}_x$ (epistemic operator)	agent x knows that

Given this more expressive vocabulary, one extended into the realm of intensional logics, here is how the key propositions from above in the suppression task are expressed in the intensional approach:

$$\exists e \ \mathit{ToFinish}(s, t_1, e) \rightarrow \exists t > t_1 \ \big(\mathit{NearFuture}(t, t_1) \wedge \mathit{LateLibrary}(s, t)\big) \tag{s1}$$

$$\exists e \ \mathit{ToFinish}(s, t_1, e) \tag{s2}$$

$$\exists t > t_1(\mathit{NearFuture}(t, t_1) \wedge \mathit{LateLibrary}(s, t)) \tag{o1}$$

$$\neg\big(\exists t > t_1(\mathit{NearFuture}(t, t_1) \wedge \mathit{LateLibrary}(s, t)))\big) \tag{o2}$$

$$\Diamond(\text{o1}) \wedge (\Diamond\neg(\text{o1}) \vee \Diamond O2) \tag{o3}$$

$$\exists b \ \mathit{ToRead}(s, t_1, b) \rightarrow \exists t > t_1 \ (\mathit{NearFuture}(t, t_1) \wedge \mathit{LateLibrary}(s, t)) \tag{s3}$$

$$[\mathit{Open}(\ell, t_1) \wedge \forall t > t_1 \ (\mathit{NearFuture}((t, t_1) \rightarrow \mathit{Open}(\ell, t)) \wedge \exists e \ \mathit{ToFinish}(s, t_1, e)] \tag{s4}$$

$$\rightarrow \exists t > t_1(\mathit{NearFuture}(t, t_1) \wedge \mathit{LateLibrary}(s, t))$$

And here is an economical summation of the deductive "facts of the case" under the more expressive rubric afforded by Table 4, where $\Gamma \vdash \phi$, as above, is the ubiquitous way in formal logic, AI, and computer science of saying that ϕ can be deduced from a set Γ of formulae (and $\nvdash$ means 'not deducible'):

- $\{(\text{s1}), (\text{s2})\} \vdash (\text{o1})$
- $\{(\text{s1}), (\text{s2})\} \nvdash (\text{o2})$
- $\{(\text{s1}), (\text{s2})\} \nvdash (\text{o3})$
- $\{(\text{s1}), (\text{s2}), (\text{s3})\} \vdash (\text{o1})$
- $\{(\text{s1}), (\text{s2}), (\text{s3})\} \nvdash (\text{o2})$
- $\{(\text{s1}), (\text{s2}), (\text{s3})\} \nvdash (\text{o3})$

Very well. And now what is the intensional modeling that matches what occurs when subjects are run in the suppression task? Such modeling, as said, takes time, possibility, and epistemic attitudes (belief and knowledge) seriously. Specifically, the heart of the matter is a simple inference schema that formalizes the principle that if an agent believes some set Φ of propositions, and knows that from this set it can be deduced specifically that proposition ϕ holds, then the agent will believe ϕ as well. Here is the inference schema, $\mathcal{S}$, expressed in a manner used in the computational simulations in question:

$$\frac{\mathbf{B}_a\Phi,\ \mathbf{K}_a\Phi \vdash \phi}{\mathbf{B}_a\phi}\ \mathcal{S}$$

And now, getting down to inferential brass tacks for computational simulation, let '**a**' denote an arbitrary agent in both Group I and Group II in the suppression-task experiment recounted above. It is then assumed, at the particular timepoint t_1, that

$$\mathbf{B}_a\{(\text{s1}), (\text{s2})\};$$

and in addition that

$$\mathbf{K_a}\{(\text{s1}), (\text{s2})\} \vdash (\text{o1}).$$

Then, by way of crucial use of $\mathcal{S}$, processing automatically locates a proof corresponding to the responses of agents in Groups I and II: $\mathbf{B}_a(\text{o1})$. In a simulation using an automated theorem prover, this result (and the corresponding proof) was returned in 10^{-4} seconds.[31]

But now, what about the "peculiar" subjects in Group III? That is, what about subjects who clearly reason defeasibly/nonmonotonically, because they go from believing that (o1) "follows," to believing, after receiving new information, that this proposition no longer does? These are of course the subjects that motivated the innovation of S&V. But how is the inferential behavior of these subjects modeled and simulated in the *intensional* approach? The answer is perfectly straightforward; it is that, first, Group-III subjects obviously know that when a library is closed (= not open) at some time t, no student can work in that library at t. This underlying principle is in the modeling here expressed thus:

(u) $\forall s \forall t \ [\neg Open(\ell, t) \rightarrow \neg LateLibrary(a, t)]$

In addition, of course, subjects in Group III know from what they have been told that

($*$) $\exists s \exists e \ ToFinish(s, t_1, e)$,

and know as well that at all near-future times relative to t_1 the library is closed;[32] that is:

($\star$) $\forall t(NearFuture(t, t_1) \rightarrow \neg Open(\ell, t))$.

Given the pair of formulae ($*$) and ($\star$) it follows by elementary deduction in $\mathscr{L}_1$ that $\neg$(s1). Therefore, while it is rationally presumed that Group-III subjects — denoted by $\mathbf{b}$ — are (like their

[31] Two automated reasoners were used to generate these simulation results. The first, ShadowProver (Govindarajulu, Bringsjord & Peveler 2019), uses a novel technique to prove formulae in a modal logic. It alternates between "shadowing" modal formulae down to first-order logic and applying modal inference schemata. The second, ShadowAdjudicator (Giancola, Bringsjord, Govindarajulu & Varela 2020), builds upon ShadowProver, providing the ability to generate *arguments* (as opposed to proofs) which can be justified using *inductive* inference schemata (as opposed to purely deductive inference schemata).

[32] Actually, as alert readers will apprehend, it's necessary here to use the alethic operator $\Diamond$ that has been introduced, since what the subjects in Group III come to know by virtue of the new information given them is that *it might possibly be* that the library is closed in the near future.

counterparts in Groups I and II) such that

$$\mathbf{K_b}\{(\text{s1}), (\text{s2})\} \vdash (\text{o1}),$$

they no longer believe (s1), and hence the use of schema $\mathcal{S}$ is blocked. In addition, it is reasonably modeled that Group-III subjects do believe (s4). But also

$$\{(\text{s4}), (\text{s2})\} \nvdash (\text{o1}),$$

and these subjects presumably know this. Hence these subjects cannot possibly know that $\{(\text{s4}), (\text{s2})\} \vdash (\text{o1})$, and this too blocks any use of schema $\mathcal{S}$ to arrive at the belief that (s1) holds.[33]

There is little point in asserting that capturing the suppression task via intensional logic is superior to the extensional-logic approach taken by S&V. However, it is very important for the student and scholar of computational cognitive science to understand that any such ambition as to capture *all* of human-level-and-above reasoning and decision-making in computational formal logic must early on confront modeling-and-simulation challenges that *necessitate* use of highly expressive intensional logics from $\mathscr{U}$.

6 Evaluating Logic-based Cognitive Modeling, Briefly

Logic-based/logicist computational cognitive modeling, LCCM as it has been abbreviated, surely seems be a rather nice fit when the cognition to be modeled is explicit, rational, and intensely inference-centric. But how accurate and informative is such modeling? And how much reach does does such an approach to cognitive modeling have, in light of the fact that surely plenty of human-level cognition is neither explicit, nor rational, nor inference-centric? This is not the venue for polemical positions to be expressed in response to such questions. But it is surely worth pointing out that "accuracy" of a cognitive model is itself not exactly the clearest concept in science, and that LCCM tantalizingly offers the opportunity to itself provide the machinery to render this

[33]Simulations of these lines of reasoning found by the relevant automated-reasoning technology are strikingly fast, stopwatch reports are left aside so as not to have to delve into rather tricky simultaneous use of the alethic operator $\Diamond$ in combination with $\mathbf{K}$ (knows) and $\mathbf{B}$ (believes). Please see note 32.

concept precise. The relationship of a model M to a targeted phenomenon P to be modeled, in LCCM, should itself be a relation formalized in some logic in the universe $\mathscr{U}$. If the relation $\mathcal{A}$ stands for "accurately models," it can then be declared that what is needed is the completion of the biconditional

$$(\star) \quad \mathcal{A}(M,P) \leftrightarrow \boxed{??}.$$

With this completion accomplished, LCCM would provide the very framework that could be used to assess its own accuracy, because one would be able to prove that $\boxed{??}$ holds in the case at hand, and then reason from right to left on the biconditional in order to deduce $\mathcal{A}(M,P)$. It is certainly not easy to find any other approach to cognitive modeling that can hold out the promise of such self-containedness.

As to the reach of LCCM, some mental phenomena do seem, at least at first glance, to be fundamentally ill-suited to this approach, for instance emotions and emotional states — and yet such mental phenomena conform remarkably well to collections of formulae from relatively simple modal (i.e. intensional) logics in $\mathscr{U}$ (Adam, Herzig & Longin 2009).

One final point regarding the assessment of LCCM, a point that follows from the above definition of what it is for logicist computational cognitive modeling to *capture* some aspect or part of human-level cognition. The point is simply this: Whether or not some attempt to cognitively model (in the LCCM approach) some phenomenon succeeds or not can be settled formally, by proof/disproof. The ultimate strong suit of LCCM is indeed formal verifiability of capture. The cognitive scientist can know that some phenomenon has been captured, period, because outright proof is available. Unfortunately, carrying this out in practice in a wide way would require the formalization of $\boxed{??}$ so that $(\star)$ can be employed in the manner described above.

7 Conclusion

It should be clear to the reader that formal computational logic is plausibly up to the challenge of modeling and simulating both quantification-centric reasoning and defeasible (= nonmonotonic) reasoning at the human level and in the human case, even when this challenge is required to be substantively based upon arguments of the sort that human agents routinely form as they adjust their belief and knowledge through time. But for the overarching program of LCCM, is the ambitious long-term goal of capturing *all* rational human cognition in computational logic reasonable? And if it is, what is next to be done?

While the present chapter extends the rather narrow deduction-focused overview of LCCM given earlier (Bringsjord 2008) into the important realms of quantification and dynamic defeasible reasoning in the human sphere, certainly humans reason and cognize in many additional ways, effectively. These additional ways range from the familiar and everyday, to the rarefied heights of cutting-edge formal science. In the former case, prominently, there is reasoning that makes crucial use of pictorial elements, and hence is reasoning that simply cannot be captured by the kind of symbolic structures we have hitherto brought to bear. As alert readers will have noticed, the universe $\mathscr{U}$ depicted in Figure 3 does include logics that offer machinery for representing and reasoning over diagrams and images. For a simple but relevant example, consider the question as to whether

❏

or

□

is more likely to have in front of it and shining upon it a light. Here, the two things centered just above aren't symbols; they are diagrams, and as such denote not as symbols do, but — to use the apt terminology of Sloman (1971) and Barwise (1995), resp. — in a manner that is *analogical* or *homomorphic*. Clearly, humans do routinely reason with diagrams — and yet the logics that have been employed above from $\mathscr{U}$ have no diagrams. Therefore further work in LCCM is clearly in order.[34] This work must bring to bear the spaces of pictorial logics indicated in the universe $\mathscr{U}$.

Now, finally, what about the latter challenge, that of applying LCCM to rarefied reasoning in the formal sciences? Here a key fact must be confronted: viz., that reasoning in logic and mathematics

[34] There are very few formal logics that allow, in addition to the standard symbolic/linguistic alphabets and grammars, diagrams/images. For such a logic, see (Arkoudas & Bringsjord 2009), which provides comprehensive references to the relevant literature.

often makes use of expressions and structures that are infinitary in nature. For example, there can be very good reason to make use of formulae that are infinitely long, such as a disjunction like

$$\delta := \exists^{=1} x Rx \vee \exists^{=2} x Rx \vee \ldots,$$

which — using a variation on the existential quantifier used repeatedly above — says that there is exactly one thing that is an R, or exactly two things each of which is an R, or exactly three things each of which is an R, and so on *ad infinitum*. It turns out that however exotic δ may seem, this is about the only way to express that there exist a finite number of Rs; and yet this way is utterly beyond the reach of first-order logic = $\mathscr{L}_1$. And yet there has been no discussion above of logics that allow for infinitely long disjunctions to be constructed; what are classified as "infinitary logics" in the universe $\mathscr{U}$, which are the logics needed, have been untouched in the foregoing discussion. Of course, as the reader will rationally suspect, the need for formulae of this nature, given the infinitary expressions presented even in textbooks devoted to bringing human students into serious cognizing about (say) analysis (e.g. see Heil 2019), is undeniable. So again, it would seem that if the general program of logic-based cognitive modeling is to succeed in capturing human reasoning and human-level reasoning across the board, additional effort of a different nature than has so far been carried out will be required of relevant researchers. This effort will need to tap other logics in $\mathscr{U}$ shown in Figure 3, which as the reader can now note by returning to that figure does indeed refer to the space of infinitary logics.[35]

[35]Readers wanting a short, cogent introduction to infinitary logic, should see presentation and explanation of the straightforward infinitary logic $\mathscr{L}_{\omega_1\omega}$ (which can express δ); and those with some logico-mathematical maturity can see (Dickmann 1975).

References

Adam, C., Herzig, A. & Longin, D. (2009), 'A Logical Formalization of the OCC Theory of Emotions', *Synthese* **168**(2), 201–248.

Andréka, H., Madarász, J. X., Németi, I. & Székely, G. (2011), 'A Logic Road from Special Relativity to General Relativity', *Synthese* pp. 1–17.
URL: *http://dx.doi.org/10.1007/s11229-011-9914-8*

Arkoudas, K. & Bringsjord, S. (2009), 'Vivid: An AI Framework for Heterogeneous Problem Solving', *Artificial Intelligence* **173**(15), 1367–1405.
URL: *http://kryten.mm.rpi.edu/KA_SB_Vivid_offprint_AIJ.pdf*

Arora, S. & Barak, B. (2009), *Computational Complexity: A Modern Approach*, Cambridge University Press, Cambridge, UK.

Ashcraft, M. & Radvansky, G. (2013), *Cognition*, Pearson, London, UK. This is the 6th edition.

Barwise, J. & Etchemendy, J. (1994), *Hyperproof*, CSLI, Stanford, CA.

Barwise, J. & Etchemendy, J. (1995), Heterogeneous Logic, *in* J. Glasgow, N. Narayanan & B. Chandrasekaran, eds, 'Diagrammatic Reasoning: Cognitive and Computational Perspectives', MIT Press, Cambridge, MA, pp. 211–234.

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

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

Bringsjord, S. (2014), 'Review of P. Thagard's *The Brain and the Meaning of Life*', *Religion & Theology* **21**, 421–425.
URL: *http://kryten.mm.rpi.edu/SBringsjord_review_PThagard_TBTMOL.pdf*

Bringsjord, S., Govindarajulu, N. & Giancola, M. (2021), 'Automated Argument Adjudication to Solve Ethical Problems in Multi-Agent Environments', *Paladyn, Journal of Behavioral Robotics* **12**, 310–335. The URL here goes to a *rough*, *uncorrected*, *truncated* preprint as of 071421.
URL: *http://kryten.mm.rpi.edu/AutomatedArgumentAdjudicationPaladyn071421.pdf*

Bringsjord, S. & Govindarajulu, N. S. (2018), Artificial Intelligence, *in* E. Zalta, ed., 'The Stanford Encyclopedia of Philosophy'.
URL: *https://plato.stanford.edu/entries/artificial-intelligence*

Bringsjord, S., Licato, J. & Bringsjord, A. (2016), The Contemporary Craft of Creating Characters Meets Today's Cognitive Architectures: A Case Study in Expressivity, *in* J. Turner, M. Nixon, U. Bernardet & S. DiPaola, eds, 'Integrating Cognitive Architectures into Virtual Character Design', IGI Global, Hershey, PA, pp. 151–180.

Byrne, R. (1989), 'Suppressing Valid Inferences with Conditionals', *Journal of Memory and Language* **31**, 61–83.

Charniak, E. & McDermott, D. (1985), *Introduction to Artificial Intelligence*, Addison-Wesley, Reading, MA.

Chisholm, R. (1966), *Theory of Knowledge*, Prentice-Hall, Englewood Cliffs, NJ.

Davis, M., Sigal, R. & Weyuker, E. (1994), *Computability, Complexity, and Languages: Fundamentals of Theoretical Computer Science*, Academic Press, New York, NY. This is the second edition, which added Sigal as a co-author.

Dickmann, M. A. (1975), *Large Infinitary Languages*, North-Holland, Amsterdam, The Netherlands.

Dietz, E.-A., Hölldobler, S. & Ragni, M. (2012), A Computational Logic Approach to the Suppression Task, *in* 'Proceedings of the Annual Meeting of the Cognitive Science Society', Vol. 34.

Dietz, E.-A., Hölldobler, S. & Wernhard, C. (2014), 'Modeling the Suppression Task Under Weak Completion and Well-Founded Semantics', *Journal of Applied Non-Classical Logics* **24**(1-2), 61–85.

Ebbinghaus, H. D., Flum, J. & Thomas, W. (1994), *Mathematical Logic (second edition)*, Springer-Verlag, New York, NY.

Feferman, S. (1995), Turing in the Land of O(Z), *in* R. Herken, ed., 'The Universal Turing Machine (2nd Ed)', Springer-Verlag, Secaucus, NJ, pp. 103–134.

Francez, N. (2015), *Proof-theoretic Semantics*, College Publications, London, UK.

Genesereth, M. & Nilsson, N. (1987), *Logical Foundations of Artificial Intelligence*, Morgan Kaufmann, Los Altos, CA.

Giancola, M., Bringsjord, S., Govindarajulu, N. S. & Varela, C. (2020), Ethical Reasoning for Autonomous Agents Under Uncertainty, *in* M. Tokhi, M. Ferreira, N. Govindarajulu, M. Silva, E. Kadar, J. Wang & A. Kaur, eds, 'Smart Living and Quality Health with Robots • Proceedings of ICRES 2020', CLAWAR, London, UK, pp. 26–41. Paper vailable at the URL given above. The ShadowAdjudicator system can be obtained here: `https://github.com/RAIRLab/ShadowAdjudicator`.
URL: *http://kryten.mm.rpi.edu/MG_SB_NSG_CV_LogicizationMiracleOnHudson.pdf*

Glymour, C. (1992), *Thinking Things Through*, MIT Press, Cambridge, MA.

Govindarajulu, N., Bringsjord, S. & Peveler, M. (2019), On Quantified Modal Theorem Proving for Modeling Ethics, *in* M. Suda & S. Winkler, eds, 'Proceedings of the Second International Workshop on Automated Reasoning: Challenges, Applications, Directions, Exemplary Achievements (ARCADE 2019)', Vol. 311 of *Electronic Proceedings in Theoretical Computer Science*, Open Publishing Association, Waterloo, Australia, pp. 43–49. The ShadowProver system can be obtained here: https://naveensundarg.github.io/prover/.
URL: *http://eptcs.web.cse.unsw.edu.au/paper.cgi?ARCADE2019.7.pdf*

Govindarajulu, N. S. & Bringsjord, S. (2017), Strength Factors: An Uncertainty System for Quantified Modal Logic, *in* V. Belle, J. Cussens, M. Finger, L. Godo, H. Prade & G. Qi, eds, 'Proceedings of the IJCAI Workshop on "Logical Foundations for Uncertainty and Machine Learning" (LFU-2017)', Melbourne, Australia, pp. 34–40.
URL: *http://homepages.inf.ed.ac.uk/vbelle/workshops/lfu17/proc.pdf*

Groarke, L. (1996/2017), Informal Logic, *in* E. Zalta, ed., 'The Stanford Encyclopedia of Philosophy'.
URL: *https://plato.stanford.edu/entries/logic-informal*

Hájek, P. (1998), *Metamathematics of Fuzzy Logic*, Trends in Logic, Volume 4, Kluwer, Dordrecht, The Netherlands.

Hayes, P. (1978), The Naïve Physics Manifesto, *in* D. Mitchie, ed., 'Expert Systems in the Microeletronics Age', Edinburgh University Press, Edinburgh, Scotland, pp. 242–270.

Hayes, P. J. (1985), The Second Naïve Physics Manifesto, *in* J. R. Hobbs & B. Moore, eds, 'Formal Theories of the Commonsense World', Ablex, Norwood, NJ, pp. 1–36.

Heil, C. (2019), *Introduction to Real Analysis*, Springer, Cham, Switzerland.

Hendricks, V. & Symons, J. (2006), Epistemic Logic, *in* E. Zalta, ed., 'The Stanford Encyclopedia of Philosophy'.
URL: *http://plato.stanford.edu/entries/logic-epistemic*

Hummel, J. (2010), 'Symbolic versus Associative Learning', *Cognitive Science* **34**(6), 958–965.

Hummel, J. E. & Holyoak, K. J. (2003), 'A symbolic-connectionist theory of relational inference and generalization', *Psychological Review* **110**, 220–264.

Johnson, G. (2016), *Argument & Inference: An Introduction to Inductive Logic*, MIT Press, Cambridge, MA.

Kemp, C. (2009), Quantification and The Language of Thought, *in* Y. Bengio, D. Schuurmans, J. Lafferty, C. Williams & A. Culotta, eds, 'Advances in Neural Information Processing

Systems', Vol. 22, Curran Associates.

URL: *https://proceedings.neurips.cc/paper/2009/file/82161242827b703e6acf9c726942a1e4-Paper.pdf*

Kleene, S. (1967), *Mathematical Logic*, Wiley & Sons, New York, NY. We recommend a 2002 Dover unabridged republication of the original 1967 book from Wiley, if you cannot obtain the original book.

Konyndyk, K. (1986), *Introductory Modal Logic*, University of Notre Dame Press, Notre Dame, IN.

Markman, A. & Gentner, D. (2001), 'Thinking', *Annual Review of Psychology* **52**, 223–247.

McCarthy, J. (1980), 'Circumscription—A Form of Non-Monotonic Reasoning', *Artificial Intelligence* **13**, 27–39.

McKeon, R., ed. (1941), *The Basic Works of Aristotle*, Random House, New York, NY.

McKinsey, J., Sugar, A. & Suppes, P. (1953), 'Axiomatic Foundations of Classical Particle Mechanics', *Journal of Rational Mechanics and Analysis* **2**, 253–272.

Nelson, M. (2015), Propositional Attitude Reports, *in* E. Zalta, ed., 'The Stanford Encyclopedia of Philosophy'.

URL: *https://plato.stanford.edu/entries/prop-attitude-reports*

Newell, A. & Simon, H. (1956), 'The Logic Theory Machine: A Complex Information Processing System', *P-868 The RAND Corporation* pp. 25–63. An almost exactly similar version of this paper can be found in *IRE Transactions on Information Theory*, vol **2**, pages 61–79.

URL: *http://shelf1.library.cmu.edu/IMLS/BACKUP/MindModels.pre_Oct1/logictheorymachine.pdf*

Núñez, R., Murthi, M., Premaratne, K., Bueno, O. & Scheutz, M. (forthcoming), 'Uncertain First-Order Logic: A Classically-Consistent Logic Framework for Modeling, Quantifying, and Tracking Uncertainty Using Dempster-Shafer Theory'.

Paris, J. & Vencovská, A. (2015), *Pure Inductive Logic*, Cambridge University Press, Cambridge, UK.

Partee, B. (2013), The Starring Role of Quantifiers in the History of Formal Semantics, *in* Vít Punčochář and Petr Švarný, ed., 'The Logica Yearbook 2012', College Publications, London, UK.

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

Pollock, J. L. (1992), 'How to Reason Defeasibly', *Artificial Intelligence* **57**(1), 1–42.
URL: *citeseer.ist.psu.edu/pollock92how.html*

Prakken, H. & Vreeswijk, G. (2001), Logics for Defeasible Argumentation, *in* D. Gabbay & F. Guenthner, eds, 'Handbook of Philosophical Logic', Springer, Dordrecht, The Netherlands, pp. 219–318.

Reiter, R. (1980), 'A Logic for Default Reasoning', *Artificial Intelligence* **13**, 81–132.

Russell, S. & Norvig, P. (2020), *Artificial Intelligence: A Modern Approach*, Pearson, New York, NY. Fourth edition.

Saldanha, E.-A. D. & Kakas, A. (2020), 'Cognitive Argumentation and the Suppression Task', *arXiv:2002.10149* .

Simpson, S. (2010), *Subsystems of Second Order Arithmetic*, Cambridge University Press, Cambridge, UK. This is the 2nd edition.

Sloman, A. (1971), 'Interactions Between Philosophy and AI: The role of Intuition and Non-logical Reasoning in Intelligence', *Artificial Intelligence* **2**, 209–225.

Smith, P. (2013), *An Introduction to Gödel's Theorems*, Cambridge University Press, Cambridge, UK. This is the second edition of the book.

Stenning, K. & van Lambalgen, M. (2008), *Human Reasoning and Cognitive Science*, MIT Press, Cambridge, MA.

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

Sun, R. & Bringsjord, S. (2009), Cognitive Systems and Cognitive Architectures, *in* B. W. Wah, ed., 'The Wiley Encyclopedia of Computer Science and Engineering Vol 1', Wiley, New York, NY, pp. 420–428.

URL: *http://kryten.mm.rpi.edu/rs_sb_wileyency_pp.pdf*

Understanding Quantifiers in Language (2009).

URL: *http://www.jakubszymanik.com/papers/understanding_qua.pdf*

Zadeh, L. (1965), 'Fuzzy Sets', *Information and Control* **8**(3), 338–353.