

0

Introduction Sergei Artemov and Melvin Fitting

“Why is this thus? What is the reason of this thusness?”¹

Modal operators are commonly understood to qualify the truth status of a proposition: necessary truth, proved truth, known truth, believed truth, and so on. The ubiquitous possible world semantics for it characterizes things in universal terms: $\Box X$ is true in some state if X is true in *all* accessible states, where various conditions on accessibility are used to separate one modal logic from another. Then $\Box(X \rightarrow Y) \rightarrow (\Box X \rightarrow \Box Y)$ is valid, no matter what conditions are imposed, by a simple and direct argument using universal quantification. Suppose both $\Box(X \rightarrow Y)$ and $\Box X$ are true at an arbitrary state. Then both X and $X \rightarrow Y$ are true at all accessible states, whatever ‘accessible’ may mean. By the usual understanding of \rightarrow , Y is true at all accessible states too, and so $\Box Y$ is true at the arbitrary state we began with. Although arguments like these have a strictly formal nature, and are studied as modal model theory, they also give us some insights into our informal, everyday use of modalities. Still, something is lacking.

Suppose we think of \Box as epistemic, and to emphasize this we use \mathbf{K} instead of \Box for the time being. For some particular X , if you assert the colloquial counterpart of $\mathbf{K}X$, that is, if you say you know X , and I ask you why you know X , you would never tell me that it is because X is true in all states epistemically compatible with this one. You would, instead, give me some sort of explicit reason: “I have a mathematical proof of X ,” or “I read X in the encyclopedia,” or “I observed that X is the case.” If I asked you why $\mathbf{K}(X \rightarrow Y) \rightarrow (\mathbf{K}X \rightarrow \mathbf{K}Y)$ is valid you would probably say something like “I could use my reason for $X \rightarrow Y$ and combine it with my reason for X , and infer Y .” This, in effect, would be your reason for Y , given that you had reasons for X and for $X \rightarrow Y$. Notice

¹ Charles Farrar Browne (1834 – 1867) was an American humorist who wrote under the pen name Artemus Ward. He was a favorite writer of Abraham Lincoln, who would read his newspaper columns to his Cabinet. This quote is from a piece called *Moses the Sassy*.

that this neatly avoids the *logical omniscience* problem: that we know all the consequences of what we know. It replaces it with the more acceptable claim that there are reasons for the consequences of what we know, based on the reasons for what we know, but reasons for consequences are more complicated things. In our example, the reason for Y has some structure to it. It combines reasons for X , reasons for $X \rightarrow Y$, and inference, which is a kind of operation on reasons. We will see more examples of this sort; in fact we have just seen a fundamental paradigm.

In place of a modal operator, \Box , justification logics have a family of *justification terms*, informally intended to represent reasons, or justifications. Instead of $\Box X$ we will see $t.X$, where t is a justification term and the formula is read “ X is so for reason t ,” or more briefly, “ t justifies X .” At a minimum, justification terms are built up from justification variables, standing for arbitrary justifications. They are built up using a set of operations which, again at a minimum, contains a binary operation \cdot . For example, $x \cdot (y \cdot x)$ is a justification term, where x and y are justification variables. The informal understanding of \cdot is that $t \cdot u$ justifies Y provided t justifies an implication with Y as its consequent, and u justifies the antecedent. In justification logics the counterpart of

$$\Box(X \rightarrow Y) \rightarrow (\Box X \rightarrow \Box Y)$$

is

$$t:(X \rightarrow Y) \rightarrow (u:X \rightarrow [t \cdot u]:Y)$$

where, as we will often do, we have added square brackets to enhance readability. Note that this exactly embodies the informal explanation we gave in the previous paragraph for the validity of $\mathbf{K}(X \rightarrow Y) \rightarrow (\mathbf{K}X \rightarrow \mathbf{K}Y)$. That is, Y has a justification built from justifications for X and for $X \rightarrow Y$ using an inference that amounts to a *modus ponens* application—we can think of the \cdot operation as an abstract representation of this inference. Other behaviors of modal operators, $\Box X \rightarrow \Box \Box X$ for instance, will require operators in addition to \cdot , and appropriate postulated behavior, in order to produce justification logics that correspond to modal logics in which $\Box X \rightarrow \Box \Box X$ is valid. Examples, general methods for doing this, and what it means to ‘correspond’ all will be discussed during the course of this book.

One more important point. Suppose X and Y are equivalent formulas, that is, we have $X \equiv Y$. Then in any normal modal logic we will also have $\Box X \equiv \Box Y$. Let us interpret the modal operator epistemically again, and write $\mathbf{K}X \equiv \mathbf{K}Y$. In fact, $\mathbf{K}X \equiv \mathbf{K}Y$ can be quite an absurd assertion. Consider some astronomically complicated tautology X of classical propositional logic. Since it is a tautology, it is equivalent to $P \vee \neg P$, which we may take for Y . Y is hardly

astronomically complicated, but since $X \equiv Y$ we will have $\mathbf{K}X \equiv \mathbf{K}Y$. But in fact, we have $\mathbf{K}Y$ because we know Y essentially by inspection, while $\mathbf{K}X$ may involve an astronomical amount of work. Informally we see that, while both X and Y are tautologies, and so both are knowable in principle, any justification we might give for knowing one, *combined with quite a lot of formula manipulation*, can give us some justification for knowing the other. The two justifications may not, indeed will not, be the same. One is simple, the other very complex. Modal logic is about *propositions*. Propositions are, in a sense, the content of formulas. Propositions are not syntactical objects. “It’s good to be the king” and “Being the king is good” express the same proposition, but not in the same way. Justifications apply to *formulas*. Equivalent formulas determine the same proposition, but can be quite different as formulas. Syntax must play a fundamental role for us, and you will see that it does, even in our semantics. Consider one more very simple example. $A \rightarrow (A \wedge A)$ is an obvious tautology. We might expect $\mathbf{K}A \rightarrow \mathbf{K}(A \wedge A)$. But we should not expect $t:A \rightarrow t:(A \wedge A)$. If t does, in fact, justify A , a justification of $A \wedge A$ may involve t , but also should involve facts about the redundancy of repetition; t by itself can not be expected to suffice.

Modal logics can express, more or less accurately, how various modal operators behave. This behavior is captured axiomatically by proofs, or semantically using possible world reasoning. These sorts of justifications for modal operator behavior are not within a modal logic, but are outside constructs. Justification logics, on the other hand, can represent the whys and wherefores of modal behavior quite directly, and from within the formal language itself. We will see that most standard modal logics have justification counterparts that can be used to give a fine-grained, internal analysis of modal behavior. Perhaps this will help make clear why we used the quotation we did, at the beginning of this Introduction.

How did justification logics originate? It is an interesting story, with revealing changes of direction along the way. Going back to the days when Gödel was a young logician, there was a dream of finding a provability interpretation for intuitionistic logic. As part of his work on that project, in Gödel (1933), Gödel showed that one could abstract some of the key features of provability and make a propositional modal logic using them. Then, remarkably but naturally, one could embed propositional intuitionistic logic into the resulting system. C. I. Lewis had pioneered the modern formal study of modal logics, Lewis (1918); Lewis and Langford (1932), and Gödel observed that his system was equivalent to the Lewis system **S4**. All modern axiomatizations of modal logics follow the lines pioneered in Gödel’s note, while Lewis’s original for-

mulation is rarely seen today. Gödel showed that propositional intuitionistic logic embedded into **S4** using a mapping that inserted \Box in front of every subformula. In effect, intuitionistic logic could be understood using classical logic plus an abstract notion of provability: a propositional formula X is an intuitionistic theorem if and only if the result of applying Gödel's mapping is a theorem of **S4**. (This story is somewhat simplified. There are several versions of the Gödel translation—we have used the simplest one to describe. And Gödel did not use the symbol \Box but rather an operator **Bew**, short for *beweisbar*, or provability in the German language. None of this affects our main points.) Unfortunately, the story breaks off at this point because Gödel also noted that **S4** does not behave like formal provability (e.g. in arithmetic), by using the methods he had pioneered in his work on incompleteness. Specifically, **S4** validates $\Box X \rightarrow X$, so in particular we have $\Box \perp \rightarrow \perp$ (where \perp is falsehood). This is equivalent to $\neg \Box \perp$, which is thus provable in **S4**. If we had an embedding of **S4** into formal arithmetic under which \Box corresponded to Gödel's arithmetic formula representing provability, we would be able to prove in arithmetic that falsehood was not provable. That is, we would be able to show provability of consistency, violating Gödel's second incompleteness theorem. So, work on an arithmetic semantics for propositional intuitionistic logic paused for a while.

Although it did not solve the problem of a provability semantics for intuitionistic logic, an important modal/arithmetic connection was eventually worked out. One can define a modal logic by requiring that its validities are those that correspond to arithmetic validities when reading \Box as Gödel's provability formula. It was shown in Solovay (1976) that this was a modal logic already known in the literature, though as noted above, it is not **S4**. Today the logic is called **GL**, standing for Gödel-Löb logic. **GL** is like **S4** except that the **S4** axiom $\Box X \rightarrow X$ is replaced by a modal formula abstractly representing Löb's theorem: $\Box(\Box X \rightarrow X) \rightarrow \Box X$. **S4** and **GL** are quite different logics.

By now the project for finding an arithmetic interpretation of intuitionistic logic had reached an impasse. Intuitionistic logic embedded into **S4**, but **S4** did not embed into formal arithmetic. **GL** embedded into formal arithmetic, but the Gödel translation does not embed intuitionistic logic into **GL**.

In his work on incompleteness for Peano arithmetic, Gödel gave a formula

$$\text{Bew}(x, y)$$

that represents the relation: x is the Gödel number of a proof of a formula with Gödel number y . Then a formal version of provability is

$$(\exists x)\text{Bew}(x, y)$$

which expresses that there is a proof of (the formula whose Gödel number is)

y . If this formula is what corresponds to \Box in an embedding from a modal language to Peano arithmetic, we get the logic GL. But in a lecture in 1938 Gödel pointed out that we might work with *explicit proof representatives* instead of with provability, Gödel (1938). That is, instead of using an embedding translating every occurrence of \Box by $(\exists x)\text{Bew}(x, y)$, we might associate with each occurrence of \Box some formal term t that somehow represents a particular proof, allowing different occurrences of \Box to be associated with different terms t . Then in the modal embedding, we could make the occurrence of \Box associated with t correspond to $\text{Bew}(\hat{t}, y)$, where \hat{t} is a Gödel number for t . For each occurrence of \Box we would need to find some appropriate term t , and then each occurrence of \Box would be translated into arithmetic differently. The existential quantifier in $(\exists x)\text{Bew}(x, y)$ has been replaced with a meta-existential quantifier, outside the formal language. We provide an explicit proof term, rather than just asserting that one exists. Gödel believed that this approach should lead to a provability embedding of S4 into Peano arithmetic.

Gödel's proposal was not published until 1995 when Volume 3 of his collected works appeared. By this time the idea of using a modal-like language with explicit representatives for proofs had been rediscovered independently by Sergei Artemov, see Artemov (1995) and Artemov (2001). The logic that Artemov created was called LP, which stood for *logic of proofs*. It was the first example of a justification logic. What are now called justification terms were called proof terms in LP. Crucially, Artemov showed LP filled the gap between modal S4 and Peano arithmetic. The connection with S4 is primarily embodied in a *Realization Theorem*, which has since been shown to hold for a wide range of justification logic, modal logic pairs. It will be extensively examined in this book. The connection between LP and formal arithmetic is Artemov's *Arithmetic Completeness Theorem*, which also will be examined in this book. Its range is primarily limited to the original justification logic, LP, and a few close relatives. This should not be surprising, though. Gödel's motivation for his formulation of S4 was that \Box should embody properties of a formal arithmetic proof predicate. This connection with arithmetic provability is not present for almost all modal logics, and is consequently also missing for corresponding justification logics, when they exist. Nonetheless, the venerable goal of finding a provability interpretation for propositional intuitionistic logic had been attained. The Gödel translation embeds propositional intuitionistic logic into the modal logic S4. The Realization Theorem establishes an embedding of S4 into the justification logic LP. And the Arithmetic Completeness Theorem shows that LP embeds into formal arithmetic.

It was recognized from the very beginning that the connection between S4 and LP could be weakened to sublogics of S4 and LP. Thus there were jus-

tification logic counterparts for the standard modal logics, K, K4, T, and a few others. These justification logics had arithmetic connections, because they were sublogics of LP. The use of *proof term* was replaced with *justification term*, and the general name *justification logic* was introduced. Although the connection with arithmetic was weaker than it had been with LP, justification terms still had the role of supplying explicit justifications for epistemically necessary statements. One can consult Artemov (2008) and Artemov and Fitting (2012) for survey treatments, though the present book includes the material found there.

Almost all of the early work on justification logics was proof-theoretically based. Realization theorems were shown constructively, making use of a sequent calculus. The existence of an algorithm to compute what are called *realizers* is important, but this proof-theoretic approach limits the field to those logics known to have sequent calculus proof systems. A semantics for LP and its near relatives was introduced in Mkrtychev (1997). It did not use possible worlds, and had a strong syntactic flavor. Possible worlds were added to the mix in Fitting (2005), producing something that potentially applied much more broadly than the earlier semantics. Using this, a non-constructive, semantic based, proof of realization was given. It was now possible to avoid the use of a sequent calculus, though the algorithmic nature of realization is lost.

Though its significance was not properly realized at the time, in 2005 the subject broadened when a justification logic counterpart of S5 was introduced in Pacuit (2005); Rubtsova (2006a,b), with a connecting realization theorem. There was no arithmetical interpretation for this justification logic. Also there is no sequent calculus for S5 of the standard kind, so the proof given for realization was non-constructive, using a version of the semantics from Fitting (2005). The semantics needed some modification to what is called its *evidence function*, and this turned out to have a greater impact than was first realized. Eventually constructive proofs connecting S5 and its justification counterpart were found. These made use of cut-free proof systems that were not exactly standard sequent calculi. Still, the door to a larger room was beginning to open.

Out of the early studies of the logics of proofs and its variants a general logical framework for reasoning about epistemic justification at large naturally emerged and the name, *Justification Logic*, was introduced (cf. Artemov (2008)). Justification Logic is based on justification assertions: $t:F$ that read t is a *justification* for F , with a broader understanding of the word justification that goes beyond just mathematical proofs. The notion of justification, which has been an essential component of epistemic studies since Plato, had been conspicuously absent in the mathematical models of knowledge within

the epistemic logic framework. The Justification Logic framework fills in this void.

In Fitting (2016) the subject expanded abruptly. Using non-constructive semantic methods it was shown that the family of modal logics having justification counterparts is infinite. The justification phenomenon is not the relatively narrow one it first seemed to be. While that work was non-constructive, there are now cut-free proof systems of various kinds for a broader range of modal logics than was once the case, and these have been used successfully to create realization algorithms, in Kuznets and Goetschi (2012) for instance. It may be that the very general proof methodologies of Fitting (2015) and especially Negri (2005); Negri and von Plato (2001) will extend the constructive range still further, perhaps even to the infinite family that non-constructive methods are known to work for. This is active current work.

Work on quantified justification logics exists, but the subject is considerably behind its propositional counterpart. An important feature of justification logics is that they can, in a very precise sense, internalize their own proofs. Doing this for axioms is generally simple. Rules of inference are more of a problem. Above we discussed a justification formula as a simple, representative example: $t:(X \rightarrow Y) \rightarrow (u:X \rightarrow [t \cdot u]:Y)$. This, in effect, internalizes the axiomatic *modus ponens* rule. The central problem in developing quantified justification logics was how to internalize the rule of universal generalization. It turned out the key was the clear separation between two roles played by individual variables. On the one hand they are formal symbols, and one can simply infer $(\forall x)\varphi(x)$ from $\varphi(x)$. On the other hand they can be thought of as open for substitution, that is, throughout a proof one can replace free occurrences of x with a term t to produce a new proof (subject to appropriate freeness of substitution conditions, of course). These two roles for variables are actually incompatible. It was the introduction of specific machinery to keep track of which role a variable occurrence had that made possible the internalization of proofs, and thus a quantified justification logic.

An axiomatic version of first-order LP was introduced in Artemov and Yavorskaya (Sidon) (2011), and a possible world semantics for it in Fitting (2003, 2014). A connection with formal arithmetic was established. As was the case with propositional LP, there are both constructive and non-constructive proofs of a Realization Theorem. The possible world semantics includes the familiar monotonicity condition on world domains. It is clear that all this can be extended to a much broader range of quantified modal logics than just first-order S4, provided monotonicity is appropriate. A move to constant domain models, to quantified S5 in particular, has been much harder and a semantics, but not

yet a Realization Theorem, can be found in Fitting and Salvatore (2017). Much of this is still uncharted territory.

This book will cover the whole range of topics described above. It will not do so in the historical order that was followed in this Introduction, but will make use of the clearer understanding that has emerged from study of the subject thus far. We will finish with the current state of affairs, standing on the edge of unknown lands. We hope to prepare some of you for the journey, should you choose to explore further on your own.

References

- Artemov, Sergei N. 1995 (Dec.). *Operational modal logic*. Tech. rept. MSI 95–29. Cornell University.
- Artemov, Sergei N. 2001. Explicit Provability and Constructive Semantics. *Bulletin of Symbolic Logic*, **7**(1), 1–36.
- Artemov, Sergei N. 2008. The Logic of Justification. *The Review of Symbolic Logic*, **1**(4), 477–513.
- Artemov, Sergei N., and Fitting, Melvin. 2012. Justification Logic. In: Zalta, Edward N. (ed), *The Stanford Encyclopedia of Philosophy*. The Metaphysics Research Lab, Center for the Study of Language and Information, Stanford University.
- Artemov, Sergei N., and Yavorskaya (Sidon), Tatiana. 2011 (May). *First-Order Logic of Proofs*. Tech. rept. TR–2011005. City University of New York.
- Feferman, Solomon, Dawson Jr., John W., Kleene, Stephen C., Moore, Gregory H., Solovay, Robert M., van Heijenoort, Jean, Goldfarb, Warren D., Parsons, Charles, and Sieg, Wilfried (eds). 1986-2003. *Kurt Gödel Collected Works*. Oxford. Five volumes.
- Fitting, Melvin C. 2003. *A Semantics for the Logic of Proofs*. Tech. rept. TR-2003012. CUNY Ph.D. Program in Computer Science.
- Fitting, Melvin C. 2005. The Logic of Proofs, Semantically. *Annals of Pure and Applied Logic*, **132**, 1–25.
- Fitting, Melvin C. 2014. Possible World Semantics for First Order Logic of Proofs. *Annals of Pure and Applied Logic*, **165**, 225–240.
- Fitting, Melvin C. 2015. Cut-Free Proof Systems for Geach Logics. *IfCoLog Journal of Logics and their Applications*, **2**(2), 17–64.
- Fitting, Melvin C. 2016. Modal Logics, Justification Logics, and Realization. *Annals of Pure and Applied Logic*, **167**, 615–648.
- Fitting, Melvin C., and Salvatore, Felipe. 2017. *First-order justification logic with constant domain semantics*. to appear.
- Gödel, Kurt. 1933. Eine Interpretation des intuitionistischen Aussagenkalküls. *Ergebnisse eines mathematischen Kolloquiums*, **4**, 39–40. Translated as *An interpretation of the intuitionistic propositional calculus* in Feferman et al. (1986-2003) I, 296-301.
- Gödel, Kurt. 1938. *Vortrag bei Zilsel*. Translated as *Lecture at Zilsel's* in Feferman et al. (1986-2003) III, 62-113.

- Kuznets, Roman. 2009. Complexity through tableaux in justification logic. Page 121 of: *2008 European Summer Meeting of the Association for Symbolic Logic, Logic Colloquium '08, Bern, Switzerland, July 3–July 8, 2008*. Bulletin of Symbolic Logic, vol. 15(1). Association for Symbolic Logic. Abstract.
- Kuznets, Roman, and Goetschi, Remo. 2012. Realization for justification logics via nested sequents: Modularity through embedding. *Annals of Pure and Applied Logic*, **163**(9), 1271–1298.
- Lewis, C. I. 1918. *A Survey of Symbolic Logic*. Berkeley: University of California Press.
- Lewis, C. I., and Langford, C. H. 1932. *Symbolic Logic*. Second edition 1959, new york: Dover edn. London: Century.
- Mkrtychev, Alexey. 1997. Models for the Logic of Proofs. Pages 266–275 of: Adian, Sergei, and Nerode, Anil (eds), *Logical Foundations of Computer Science, 4th International Symposium, LFCS'97, Yaroslavl, Russia, July 6-12, 1997, Proceedings*. Lecture Notes in Computer Science, vol. 1234. Springer.
- Negri, Sara. 2005. Proof analysis in modal logic. *Journal of Philosophical Logic*, **34**, 507–544.
- Negri, Sara, and von Plato, Jan. 2001. *Structural Proof Theory*. Paperback edition 2008 edn. Cambridge University Press.
- Pacuit, Eric. 2005. A Note on some explicit modal logics. Pages 117–125 of: Dimitracopoulos, Costas (ed), *Proceedings of the Fifth Panhellenic Logic Symposium*.
- Rubtsova, Natalia M. 2006a. Evidence Reconstruction of Epistemic Modal Logic S5. Pages 313–321 of: Grigoriev, Dima, Harrison, John, and Hirsch, Edward A. (eds), *Computer Science – Theory and Application, First International Computer Science Symposium in Russia, CSR 2006, St. Petersburg, Russia, June 8-12, 2006, Proceedings*. Lecture Notes in Computer Science, vol. 3967. Springer.
- Rubtsova, Natalia M. 2006b. On Realization of S5-modality by Evidence Terms. *Journal of Logic and Computation*, **16**(5), 671–684.
- Solovay, Robert M. 1976. Provability interpretation of modal logic. *Israel Journal of Mathematics*, **25**, 287–304.