Justification Logic

Classical logic is concerned, loosely, with the behavior of truths. Epistemic logic similarly is about the behavior of known or believed truths. Justification logic is a theory of reasoning that enables the tracking of evidence for statements and therefore provides a logical framework for the reliability of assertions. This book, the first in the area, is a systematic account of the subject, progressing from modal logic through to the establishment of an arithmetic interpretation of intuitionistic logic. The presentation is mathematically rigorous but in a style that will appeal to readers from a wide variety of areas to which the theory applies. These include mathematical logic, artificial intelligence, computer science, philosophical logic and epistemology, linguistics, and game theory.

SERGEI ARTEMOV is Distinguished Professor at the City University of New York. He is a specialist in mathematical logic, logic in computer science, control theory, epistemology, and game theory. He is credited with solving long-standing problems in constructive logic that had been left open by Gödel and Kolmogorov since the 1930s. He has pioneered studies in the logic of proofs and justifications that render a new, evidence-based theory of knowledge and belief. The most recent focus of his interests is epistemic foundations of game theory.

MELVIN FITTING is Professor Emeritus at the City University of New York. He has written or edited a dozen books and has worked in intensional logic, semantics for logic programming, theory of truth, and tableau systems for nonclassical logics. In 2012 he received the Herbrand Award from the Conference on Automated Deduction. He was on the faculty of the City University of New York from 1969 to his retirement in 2013, at Lehman College, and at the Graduate Center, where he was in the Departments of Mathematics, Computer Science, and Philosophy.

CAMBRIDGE TRACTS IN MATHEMATICS

Recent titles include the following:

181. *Totally Positive Matrices*. By A. Pinkus
182. *Nonlinear Markov Processes and Kinetic Equations*. By V. N. Kolokoltsov
183. Period Domains over Finite and p-adic Fields. By J.-F. Dat, S. Orlik, and M. Rapoport
184. *Algebraic Theories*. By J. Adámek, J. Rosický, and E. M. Vitale
185. *Rigidity in Higher Rank Abelian Group Actions I: Introduction and Cocycle Problem*. By A. Katok and V. Nițică
186. *Dimensions, Embeddings, and Attractors*. By J. C. Robinson
187. *Convexity: An Analytic Viewpoint.* By B. Simon
188. *Modern Approaches to the Invariant Subspace Problem*. By I. Chalendar and J. R. Partington
189. *Nonlinear Perron–Frobenius Theory*. By B. Lemmens and R. Nussbaum
190. *Jordan Structures in Geometry and Analysis*. By C.-H. Chu
191. *Malliavin Calculus for Lévy Processes and Infinite-Dimensional Brownian Motion*. By H. Osswald
192. *Normal Approximations with Malliavin Calculus*. By I. Nourdin and G. Peccati
193. *Distribution Modulo One and Diophantine Approximation*. By Y. Bugeaud
194. *Mathematics of Two-Dimensional Turbulence*. By S. Kuksin and A. Shirikyan
195. *A Universal Construction for Groups Acting Freely on Real Trees*. By I. Chiswell and T. Müller
196. *The Theory of Hardy's Z-Function*. By A. Ivić
197. *Induced Representations of Locally Compact Groups*. By E. Kaniuth and K. F. Taylor
198. *Topics in Critical Point Theory*. By K. Perera and M. Schechter
199. *Combinatorics of Minuscule Representations*. By R. M. Green
200. *Singularities of the Minimal Model Program*. By J. Kollár
201. *Coherence in Three-Dimensional Category Theory*. By N. Gurski
202. *Canonical Ramsey Theory on Polish Spaces*. By V. Kanovei, M. Sabok, and J. Zapletal
203. *A Primer on the Dirichlet Space*. By O. El-Fallah, K. Kellay, J. Mashreghi, and T. Ransford
204. *Group Cohomology and Algebraic Cycles*. By B. Totaro
205. *Ridge Functions*. By A. Pinkus
206. *Probability on Real Lie Algebras*. By U. Franz and N. Privault
207. *Auxiliary Polynomials in Number Theory*. By D. Masser
208. *Representations of Elementary Abelian p-Groups and Vector Bundles*. By D. J. Benson
209. *Non-homogeneous Random Walks*. By M. Menshikov, S. Popov, and A. Wade
210. *Fourier Integrals in Classical Analysis (Second Edition)*. By C. D. Sogge
211. *Eigenvalues, Multiplicities and Graphs*. By C. R. Johnson and C. M. Saiago
212. *Applications of Diophantine Approximation to Integral Points and Transcendence*. By P. Corvaja and U. Zannier
213. *Variations on a Theme of Borel*. By S. Weinberger
214. *The Mathieu Groups*. By A. A. Ivanov
215. *Slender Modules and Rings I: Foundations*. By R. Dimitric

Justification Logic

Reasoning with Reasons

SERGEI ARTEMOV
Graduate Center, City University of New York

MELVIN FITTING
Graduate Center, City University of New York

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE
UNIVERSITY PRESS

University Printing House, Cambridge CB2 8BS, United Kingdom

One Liberty Plaza, 20th Floor, New York, NY 10006, USA

477 Williamstown Road, Port Melbourne, VIC 3207, Australia

314–321, 3rd Floor, Plot 3, Splendor Forum, Jasola District Centre,
New Delhi – 110025, India

79 Anson Road, #06–04/06, Singapore 079906

Cambridge University Press is part of the University of Cambridge.

It furthers the University's mission by disseminating knowledge in the pursuit of education, learning, and research at the highest international levels of excellence.

www.cambridge.org
Information on this title: www.cambridge.org/9781108424912
DOI: 10.1017/9781108348034

First published 2019

Printed and bound in Great Britain by Clays Ltd, Elcograf S.p.A.

A catalogue record for this publication is available from the British Library.

Library of Congress Cataloging-in-Publication Data
Names: Artemov, S. N., author. | Fitting, Melvin, 1942- author.
Title: Justification logic : reasoning with reasons / Sergei Artemov (Graduate Center, City University of New York), Melvin Fitting (Graduate Center, City University of New York).
Description: Cambridge ; New York, NY : Cambridge University Press, 2019. | Series: Cambridge tracts in mathematics ; 216 | Includes bibliographical references and index.
Identifiers: LCCN 2018058431 | ISBN 9781108424912 (hardback : alk. paper)
Subjects: LCSH: Logic, Symbolic and mathematical. | Inquiry (Theory of knowledge) | Science–Theory reduction. | Reasoning.
Classification: LCC QA9 .A78 2019 | DDC 511.3–dc23
LC record available at https://lccn.loc.gov/2018058431

ISBN 978-1-108-42491-2 Hardback

To our wives, Lena and Roma.

Contents

Introduction

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 distinguish one modal logic from another. Then $\Box(X \to Y) \to (\Box X \to \Box Y)$ is valid, no matter what conditions are imposed, by a simple and direct argument using universal quantification. Suppose both $\Box(X \to Y)$ and $\Box X$ are true at an arbitrary state. Then both X and $X \to Y$ are true at all accessible states, whatever "accessible" may mean. By the usual understanding of $\to$, 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 $\mathbf{K}(X \to Y) \to (\mathbf{K}X \to \mathbf{K}Y)$ is valid you would probably say something like "I could use my reason for X and combine it with my reason for $X \to Y$, and infer Y." This, in effect, would be your reason for Y, given that you had reasons for X and for $X \to Y$.

[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 articles to his Cabinet. This quote is from a piece called *Moses the Sassy*, Ward (1861).

Notice that this neatly avoids the *logical omniscience* problem: that we know all the consequences of what we know. It replaces logical omniscience 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 $X \to Y$, and inference as 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 that, 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

$$\Box(X \to Y) \to (\Box X \to \Box Y)$$

is

$$t{:}(X \to Y) \to (u{:}X \to [t \cdot u]{:}Y)$$

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 $\mathbf{K}(X \to Y) \to (\mathbf{K}X \to \mathbf{K}Y)$. That is, Y has a justification built from justifications for X and for $X \to Y$ 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, $\Box X \to \Box\Box X$ 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 $\Box X \to \Box\Box X$ 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 $X \leftrightarrow Y$. Then in any normal modal logic we will also have $\Box X \leftrightarrow \Box Y$. Let us interpret the modal operator epistemically again, and write $\mathbf{K}X \leftrightarrow \mathbf{K}Y$. In fact, $\mathbf{K}X \leftrightarrow \mathbf{K}Y$, when read in the usual epistemic way, can sometimes be quite an absurd assertion. Consider some astronomically complicated tautology X of classical propositional logic. Because it is a tautology, it is equivalent

to $P \vee \neg P$, which we may take for Y. Y is hardly astronomically complicated. However, because $X \leftrightarrow Y$, we will have $\mathbf{K}X \leftrightarrow \mathbf{K}Y$. Clearly, we know Y essentially by inspection and hence $\mathbf{K}Y$ holds, while $\mathbf{K}X$ on the other hand will involve an astronomical amount of work just to read it, let alone to verify it. 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 cannot 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.

1 What Is This Book About?

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 formulation 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 $\Box X \to X$, so in particular we have $\Box\bot \to \bot$ (where $\bot$ is falsehood). This is equivalent to $\neg\Box\bot$, 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 earlier, it is not S4. Today, the logic is called GL, standing for Gödel–Löb logic. GL is like S4 except that the T axiom $\Box X \to X$, an essential part of S4, is replaced by a modal formula abstractly representing Löb's theorem: $\Box(\Box X \to X) \to \Box X$. 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

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

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

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

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 $\exists x \mathsf{Bew}(x, y)$, 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 $\mathsf{Bew}(\ulcorner t \urcorner, y)$, where $\ulcorner t \urcorner$ 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 $\exists x \mathsf{Bew}(x, y)$ 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, 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 justification 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*. 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. For a time it was hoped that various extensions of sequent and tableau calculi would be useful and, to some extent, this has been the case. The most optimistic version of this hope was expressed in Artemov (2001) quite directly, "Gabbay's Labelled Deductive Systems, Gabbay (1994), may serve as a natural framework for LP." Unfortunately this seems to have been too optimistic. While the formats had similarities, the goals were different, and the machinery did not interact well.

A semantics for LP and its near relatives, not based on arithmetic provability, was introduced in Mkrtychev (1997) and is discussed in Chapter 3. (A constructive version of the canonical model for LP with a completeness theorem can be found already in Artemov (1995).) Mkrtychev's semantics 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. This is the subject of Chapter 4. Using this possible world semantics, a nonconstructive, 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 was lost. More recently, a semantics with a very simple structure was created, Artemov's *basic* semantics (Artemov, 2012). It is presented in Chapter 3. Its machinery is almost minimal

for the purpose. In this book, we will use possible world semantics to establish very general realization results, but basic models will often be used when we simply want to show some formula fails to be a theorem.

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) and 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 nonconstructive, 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 are read t is a *justification* for F, with a broader understanding of the word justification going 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 (2016a) the subject expanded abruptly. Using nonconstructive 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 nonconstructive, 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) and Negri and von Plato (2001) will extend the constructive range still further, perhaps even to the infinite family that nonconstructive 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. Earlier we discussed a justification formula as a simple, representative exam-

ple: $t{:}(X \to Y) \to (u{:}X \to [t \cdot u]{:}Y)$. 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 that 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 a proof of $\varphi(x)$. 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 (2011a, 2014b). A connection with formal arithmetic was established. There is a constructive proof of a Realization Theorem, connecting first-order LP with first-order S4. Unlike propositionally, no nonconstructive proof is currently known The possible world semantics includes the familiar monotonicity condition on world domains. It is likely 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 made, and a semantics, but not yet a Realization Theorem, can be found in Fitting and Salvatore (2018). Much involving quantification is still uncharted territory.

This book will cover the whole range of topics just described. 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.

2 What Is Not in This Book?

There are several historical works and pivotal developments in justification logic that will not be covered in the book due to natural limitations, and in this section we will mention them briefly. We are confident that other books and surveys will do justice to these works in more detail.

Apart from Gödel's lecture, Gödel (1938), which remained unpublished

until 1995 and thus could not influence development in this area, the first results and publications on the logic of proofs are dated 1992: a technical report, Artemov and Straßen (1992), based on work done in January of 1992 in Bern, and a conference presentation of this work at CSL'92 published in Springer Lecture Notes in Computer Science as Artemov and Straßen (1993a). In this work, the *basic logic of proofs* was presented: it had proof variables, and the format *t is a proof of F*, but without operations on proofs. However, it already had the first installment of the fixed-point arithmetical completeness construction together with an observation that, unlike provability logic, the logic of proofs cannot be limited to one standard proof predicate "from the textbook" or to any single-conclusion proof predicate.

This line was further developed in Artemov and Straßen (1993b), where the logic of single-conclusion proof predicates (without operations on proofs) was studied. This work introduced the unification axiom, which captures single-conclusioness by propositional tools. After the full-scale logic of proofs with operations had been discovered (Artemov, 1995), the logic of single-conclusion proofs with operations was axiomatized in V. Krupski (1997, 2001). A similar technique was used recently to characterize so-called sharp single-conclusion justification models in Krupski (2018).

Another research direction pursued after the papers on the basic logic of proofs was to combine provability and explicit proofs. Such a combination, with new provability principles, was given in Artemov (1994). Despite its title, this paper did not introduce what is known now as The Logic of Proofs, but rather a fusion of the provability logic GL and the basic logic of proofs without operations, but with new arithmetical principles combining proofs and provability and an arithmetical completeness theorem. After the logic of proofs paper (Artemov, 1995), the full-scale logic of provability and proofs (with operations), LPP, was axiomatized and proved arithmetically complete in Sidon (1997) and Yavorskaya (Sidon) (2001). A leaner logic combining provability and explicit proofs, GLA, was introduced and proved arithmetically complete in Nogina (2006, 2014b). Unlike LPP, the logic GLA did not use additional operations on proofs other than those inherited from LP. Later, GLA was used to find a complete classification of reflection principles in arithmetic that involve provability and explicit proofs (Nogina, 2014a).

The first publication of the full-scale logic of proofs with operations, LP, which became the first justification logic in the modern sense, was Artemov (1995). This paper contains all the results needed to complete Gödel's program of characterizing intuitionistic propositional logic IPC and its BHK semantics via proofs in classical arithmetic: internalization, the realization theorem for S4 in LP, arithmetical semantics for LP, and the arithmetical completeness the-

orem. It took six years for the corresponding journal paper to appear: Artemov (2001). In Goris (2008), the completeness of LP for the semantics of proofs in Peano arithmetic was extended to the semantics of proofs in Buss's bounded arithmetic S_2^1. In view of applications in epistemology, this result shows that explicit knowledge in the propositional framework can be made computationally feasible. Kuznets and Studer (2016) extend the arithmetical interpretation of LP from the original finite constant specifications to a wide class of constant specifications, including some infinite ones. In particular, this "weak" arithmetical interpretation captures the full logic of proofs LP with the total constant specification.

Decidability of LP (with the total constant specification) was also established in Mkrtychev (1997), and this opened the door to decidability and complexity studies in justification logics using model-theoretic and other means. Among the milestones are complexity estimates in Kuznets (2000), Brezhnev and Kuznets (2006), Krupski (2006a), Milnikel (2007), Buss and Kuznets (2012), and Achilleos (2014a).

The arithmetical provability semantics for the Logic of Proofs, LP, naturally generalizes to a first-order version with conventional quantifiers and to a version with quantifiers over proofs. In both cases, axiomatizability questions were answered negatively in Artemov and Yavorskaya (2001) and Yavorsky (2001). A natural and manageable first-order version of the logic of proofs, FOLP, has been studied in Artemov and Yavorskaya (Sidon) (2011), Fitting (2014a), and Fitting and Salvatore (2018) and will be covered in Chapter 10.

Originally, the logic of proofs was formulated as a Hilbert-style axiomatic system, but this has gradually broadened. Early attempts were tableau based (which could equivalently be presented using sequent calculus machinery). Tableaus generally are *analytic*, meaning that everything entering into a proof is a subformula of what is being proved. This was problematic for attempts at LP tableaus because of the presence of the $\cdot$ operation, which represented an application of modus ponens, a rule that is decidedly not analytic. Successful tableau systems, though not analytic, for LP and closely related logics can be found in Fitting (2003, 2005) and Renne (2004, 2006). The analyticity problem was overcome in Ghari (2014, 2016a). Broader proof systems have been investigated: hypersequents in Kurokawa (2009, 2012), prefixed tableaus in Kurokawa (2013), and labeled deductive systems in Ghari (2017). Indeed some of this has led to new realization results (Artemov, 1995, 2001, 2002, 2006; Artemov and Bonelli, 2007; Ghari, 2012; Kurokawa, 2012).

Finding a computational reading of justification logics has been a natural research goal. There were several attempts to use the ideas of LP for building a lambda-calculus with internalization, cf. Alt and Artemov (2001), Artemov

(2002), Artemov and Bonelli (2007), Pouliasis and Primiero (2014), and others. Corresponding combinatory logic systems with internalization were studied in Artemov (2004), Krupski (2006b), and Shamkanov (2011). These and other studies can serve as a ground for further applications in typed programming languages. A version of the logic of proofs with a built-in verification predicate was considered in Protopopescu (2016a, b).

The aforementioned intuition that justification logic naturally avoids the logical omniscience problem has been formalized and studied in Artemov and Kuznets (2006, 2009, 2014). The key idea there was to view logical omniscience as a proof complexity problem: The logical omniscience defect occurs if an epistemic system assumes knowledge of propositions, which have no feasible proofs. Through this prism, standard modal logics are logically omniscient (modulo some common complexity assumptions), and justification logics are not logically omniscient. The ability of justification logic to track proof complexity via time bounds led to another formal definition of logical omniscience in Wang (2011a) with the same conclusion: Justification logic keeps logical omniscience under control.

Shortly after the first paper on the logic of proofs, it became clear that the logical tools developed are capable of evidence tracking in a general setting and as such can be useful in epistemic logic. Perhaps, the first formal work in this direction was Artemov et al. (1999), in which modal logic S5 was equivalently modified and supplied with an LP-style explicit counterpart. Applications to epistemology have benefited greatly from Fitting semantics, which connected justification logics to mainstream epistemology via possible worlds models. In addition to applications discussed in this book, we would like to mention some other influential work. Game semantics of justification logic was studied in Renne (2008) and dynamic epistemic logic with justifications in Renne (2008) and Baltag et al. (2014). In Sedlár (2013), Fitting semantics for justification models was elaborated to a special case of the models of general awareness. Multiagent justification logic and common knowledge has been studied in Artemov (2006), Antonakos (2007), Yavorskaya (Sidon) (2008), Bucheli et al. (2010, 2011), Bucheli (2012), Antonakos (2013), and Achilleos (2014b, 2015a, b). In Dean and Kurokawa (2010), justification logic was used for the analysis of Knower and Knowability paradoxes. A fast-growing and promising area is probabilistic justification logic, cf. Milnikel (2014), Artemov (2016b), Kokkinis et al. (2016), Ghari (2016b), and Lurie (2018).

We are deeply indebted to all contributors to the exciting justification logic project, without whom there would not be this book.

Very special thanks to our devoted readers for their sharp eyes and their useful comments: Vladimir Krupski, Vincent Alexis Peluce, and Tatiana Yavorskaya (Sidon).

> I think there is no sense in forming an opinion when there is no evidence to form it on. If you build a person without any bones in him he may look fair enough to the eye, but he will be limber and cannot stand up; and I consider that *evidence* is the bones of an opinion.[2]

[2] Mark Twain (1835–1910). The quote is from his last novel, *Personal Recollections of Joan of Arc*, Twain (1896).