Hofstra Law Faculty Scholarship

2011

# A framework for the extraction and modeling of fact-finding reasoning from legal decisions: lessons from the Vaccine/Injury Project Corpus

Vern R. Walker
*Maurice A. Deane School of Law at Hofstra University*

Nathaniel Carie

Courtney C. Dewitt

Eric Lesh

Follow this and additional works at: https://scholarlycommons.law.hofstra.edu/faculty_scholarship

# A framework for the extraction and modeling of fact-finding reasoning from legal decisions: lessons from the Vaccine/Injury Project Corpus

**Vern R. Walker · Nathaniel Carie ·
Courtney C. DeWitt · Eric Lesh**

**Abstract**   This article describes the Vaccine/Injury Project Corpus, a collection of legal decisions awarding or denying compensation for health injuries allegedly due to vaccinations, together with models of the logical structure of the reasoning of the factfinders in those cases. This unique corpus provides useful data for formal and informal logic theory, for natural-language research in linguistics, and for artificial intelligence research. More importantly, the article discusses lessons learned from developing protocols for manually extracting the logical structure and generating the logic models. It identifies sub-tasks in the extraction process, discusses challenges to automation, and provides insights into possible solutions for automation. In particular, the framework and strategies developed here, together with the corpus data, should allow "top–down" and contextual approaches to automation, which can supplement "bottom-up" linguistic approaches. Illustrations throughout the article use examples drawn from the Corpus.

**Keywords**   Argumentation mining · Automation · Legal evidence · Legal rule · Logic schema · Vaccines

## 1 Introduction

Since Wigmore's early work on evidence charting, a core problem has been developing a methodology for modeling and measuring the "force of evidence" or

V. R. Walker (✉) · N. Carie · C. C. DeWitt · E. Lesh
Research Laboratory for Law, Logic and Technology, Hofstra University School of Law,
121 Hofstra University, Hempstead, NY 11549, USA
e-mail: Vern.R.Walker@Hofstra.edu
URL: http://www.LLTLab.org

"evidential support" for findings of fact in legal decisions (Anderson and Twining 1991; Schum 1994; Kadane and Schum 1996). One approach has been to analyze the logical structure of evidentiary reasoning—using either formal structures for default or defeasible inference (Toulmin et al. 1984; Kyburg and Teng 2001; Brachman and Levesque 2004), as a sub-area of nonmonotonic logic (Brewka et al. 1997), or informal logic theory in the form of argumentation schemes (Walton et al. 2008; Walton 1996). Unsurprisingly, artificial intelligence research has also been investigating formal structures useful in fact finding (Rissland et al. 2003; Ashley and Rissland 2003; Ashley 2009; Branting 2000), including default logic (Prakken 2004; Prakken and Sartor 2004; Verheij 2005) and argumentation schemes (Walton et al. 2008; Prakken et al. 2003; Gordon and Walton 2009). Such approaches encounter the problem of extracting the logical structure of reasoning from natural-language legal decisions.

The nature of law, moreover, suggests that theoretical work in this area must be grounded at some point in analyses of actual reasoning in documented legal decisions. Decision-making processes in law are distinctive because they involve a balancing of the epistemic objective against various non-epistemic objectives (Walker 2003, 2007a). The epistemic objective is to produce findings of fact that are as accurate as possible, and which are warranted by the evidence available to the factfinder. That is, factual conclusions must generally be based on evidence available in the "record" of the proceeding. Other objectives, on the other hand, are non-epistemic, and can be either generic to multiple legal areas (objectives such as procedural fairness to parties and administrative efficiency) or specific to a given area (objectives such as ensuring compensation for the side-effects of vaccines, and ensuring the integrity of the vaccine supply). Decision making in law, moreover, frequently involves fact finding under uncertainty, in the face of an acknowledged lack of complete information. In such situations, and given the need for fact finding in real time, a balance of competing objectives often fills the inferential gaps between incomplete evidence and findings of fact. Studying the balances that factfinders actually strike in such situations in different legal areas requires empirical research into human reasoning in real cases.

Studying actual legal decisions, however, faces the same challenges that natural language processing (NLP) encounters in other knowledge domains. Authors of natural-language documents can have highly variable styles of expression (allowed by the syntactic and semantic flexibility of natural languages), and such variability can make automated analysis extremely difficult (Ashley and Brüninghaus 2009, pp. 140, 154–155). Progress depends upon the use of simplified or annotated documents prepared by those with expertise in the law. For example, Ashley and Brüninghaus trained and evaluated their SMILE program for classifying case texts by using "squibs" developed by law students—that is, "narrative summaries of each case's important facts" (2009, pp. 141–142, 154–155; Brüninghaus and Ashley 2005, pp. 65–66). Saravanan and Ravindran trained and evaluated their program for assigning rhetorical roles to sentences by using a corpus of 200 decisions from Indian courts that was manually annotated by two lawyers (2010, pp. 51, 64–65). Other examples include Palau and Moens, using a set of legal texts of the European Court of Human Rights, with sentences manually classified with regard to playing a

role in argumentation, and utilizing a structured process to obtain agreement among lawyers as annotators (2009, pp. 102–103; Mochales and Moens 2008, pp. 12–14); Moens et al. using a set of sentences from the Araucaria corpus drawn from various sources, including newspapers and court reports, with the sentences manually classified as containing an argument or not (2007, p. 227); Biagioli et al. using a dataset of paragraphs selected and labeled by legal experts into eleven types of legislative "provision" (2005, p. 136); and Chorley and Bench-Capon, using a subset of cases and factors, with the researchers supplying the social values that are enhanced or promoted by deciding the case for the favored party (2005, pp. 333–336). In such work, the representativeness of the corpus and the validity of any annotations are critical for producing accurate automation that has general applicability. In addition, insights into the techniques used by human experts in producing accurate squibs, annotations, and classifications can provide insights into methods for automating the analysis of new legal decisions (see Wyner and Peters 2010).

This article has two major objectives. First, it describes a new corpus becoming available for developing and evaluating software for analyzing the fact-finding portions of legal decisions. The Vaccine/Injury Project Corpus ("V/IP Corpus") extracts the logical structure of the reasoning of the factfinders in complex decisions awarding or denying compensation for injuries allegedly caused by vaccines. The Corpus consists of the full decisions (which generally range from 20 to 40 double-spaced pages), as well as models of the reasoning in those decisions that connects the supporting evidence in the record to the findings of fact about causation. These case models contain annotations that allow mapping of the elements of the models into the sentences in the full decisions. The Corpus is being developed, maintained and made publicly available by the Research Laboratory for Law, Logic and Technology (LLT Lab) at Hofstra University School of Law in New York (URL: http://www.LLTLab.org).

Second, and more importantly, the article explains the process by which the V/IP Corpus is being generated, and analyzes the methodology developed at the LLT Lab for modeling the fact-finding reasoning found in those legal decisions. The Lab's primary goals in the modeling project (discussed in more detail in Sect. 3.7) are to produce models (1) that accurately represent the essential inference structure of the reported decision (descriptive accuracy); (2) that contain sufficient information such that a valid normative critique of the model is also a valid critique of the reported reasoning (normative completeness); (3) that capture important patterns in the reasoning (pattern discovery); and (4) that are useful as tools for understanding the reasoning (comprehension), and potentially for improving litigation support software. When these goals are in conflict, descriptive accuracy is always the most important goal. (For a discussion of the goals of argumentation mining, see Palau and Moens 2009, p. 102.) As the article discusses the protocols developed for manually extracting the logical structure from the natural-language documents, it identifies sub-tasks in the extraction process, discusses the challenges to automating those sub-tasks, and provides insights into the possibilities for successful automation. It illustrates those sub-tasks, challenges and insights by presenting a wide variety of examples drawn from the V/IP Corpus.

The ultimate goal of the research at the LLT Lab is to develop reliable, accurate and cost-effective methods of extracting logical structure from legal decisions. Such models of reasoning can be useful to judges and other participants in the legal process, so that legal decision making can become more transparent, more predictable, more efficient, and more accessible to the public. (For a discussion of reasons to focus on the legal domain when studying argumentation mining, see Palau and Moens 2009, p. 103.) If extraction and modeling methods are to be useful in conducting empirical research in a variety of different legal areas, they would have to be flexible enough to apply to a variety of decisions written in a variety of natural languages, should produce models and annotations that are standardized for further use in a variety of contexts, and should produce results that are susceptible to empirical testing for reliability and validity. Automated extraction of fact-finding reasoning from documented cases would be an important contribution to achieving this goal. This article advances research in this direction by describing the analytic framework of the V/IP Corpus, and by summarizing the numerous lessons gained to date while generating that Corpus.

## 2 The Vaccine/Injury Project Corpus

The Research Laboratory for Law, Logic and Technology is modeled on research laboratories in the sciences. Its mission is to employ teams of faculty and students to develop and disseminate practical tools for legal research, education and practice; to develop methodologies that enable the upward scaling of research projects, by increasing the number of researchers and collaborating research laboratories; and to advance the use of technology in law. These mission objectives are combined in the LLT Lab's Vaccine/Injury Project (V/IP). The project investigates the fact-finding reasoning in legal decisions made within the Vaccine Injury Compensation Program (VICP) in the United States—a claims process for compensation due to injuries caused by vaccinations. One part of that project is developing a corpus of modeled and annotated legal decisions (the V/IP Corpus) that has a variety of uses. This section of the article provides a brief overview of the Vaccine Injury Compensation Program and the Vaccine/Injury Project. It briefly describes the structure of a "logic model" created to represent the reasoning in a legal decision, as well as the Lab's research goals in generating such models. Section 3 of the article will discuss the project's protocols for extracting the logical structure of reasoning from the legal decisions of the VICP, as well as challenges and insights for automating sub-tasks within that process of extraction. Throughout the discussion, cross-references will be made to recent parallel work on argumentation mining by Ashley and Brüninghaus (2009), Saravanan and Ravindran (2010), and Palau and Moens (2009).

2.1 The Vaccine Injury Compensation Program (VICP) and V/IP Corpus

The VICP is a hybrid administrative-judicial system in the United States for providing compensation to persons who have sustained vaccine-related injuries

(Walker 2009b). Compensation awards are paid out of the Vaccine Injury Compensation Trust Fund, which is funded by an excise tax on each dose of covered vaccine. Petitioners claiming compensation file claims in the United States Court of Federal Claims. Claims contested by the Secretary of the Department of Health and Human Services (HHS) are decided by one of eight special masters in the Office of Special Masters, which Congress established for this purpose within the Court of Federal Claims. Contested VICP cases often involve complex issues of fact, which require taking into account medical, scientific, and other expert evidence along with non-expert evidence. Moreover, the special masters carefully document their evidence assessment and fact finding in their decisions, which must include "findings of fact and conclusions of law." Under certain conditions, the Court of Federal Claims has jurisdiction to review a special master's decision, and a judgment of that Court can be appealed to the United States Court of Appeals for the Federal Circuit. However, the findings of fact made by special masters may be set aside only if they are "arbitrary and capricious"—a very deferential standard of review (Walker 2009b). Such factfinders have wide discretion in assessing the probative value of the evidence and in drawing conclusions about the facts.

A major factual issue in contested petitions for compensation is often causation: whether the vaccine caused the injury alleged. The statute provides two ways for a petitioner to establish causation. The first involves a statutory presumption of causation, which is triggered when the person who was vaccinated with a covered vaccine sustained a type of injury that is listed for that vaccine in the Vaccine Injury Table, and did so within the time period after vaccination specified in that Table (see 42 U.S.C. § 300aa-11(c)(1)(C)(i) (2006)). The Vaccine Injury Table is an administratively maintained table of covered vaccines, injuries recognized as related to particular vaccines, and recognized time periods for the onset of those injuries. If the petitioner demonstrates that these conditions are satisfied, then the statute directs that the vaccine is rebuttably presumed to be a cause of the injury.

The second way of proving causation applies if the injury complained of is not listed in the Vaccine Injury Table, or the onset occurred outside the time period listed in that Table (see 42 U.S.C. § 300aa-11(c)(1)(C)(ii) (2006)). The rules for proving causation in this second way (for "Off-Table" injuries) were enunciated by the lead case of *Althen v. Secretary of HHS*, 418 F.3d 1274 (Fed. Cir. July 29, 2005), in which the Court stated that a petitioner must prove

> that the vaccination brought about her injury by providing: (1) a medical theory causally connecting the vaccination and the injury; (2) a logical sequence of cause and effect showing that the vaccination was the reason for the injury; and (3) a showing of a proximate temporal relationship between vaccination and injury.

(*Althen*, 418 F.3d at 1278.) The study sample for the Vaccine/Injury Project consists of decisions filed by special masters after the *Althen* decision, and in which the special master applies this three-pronged test of causation-in-fact. This study sample was selected for the societal importance of its decisions, the focus on the important concept of causation, and the complexity of evidence and reasoning involving both scientific and non-scientific evidence. The causation-in-fact rules established by

*Althen* have also remained stable over the years following that decision, with the three-pronged test of *Althen* remaining intact (see, for example, *Broekelschen v. Secretary of Health and Human Services*, 618 F.3d 1339, 1345 (Fed. Cir. 2010)). Currently, about 30 decisions are completed or nearing completion, and eventually the Corpus should contain at least double that number of decisions. The LLT Lab is modeling decisions chronologically beginning with the date of the *Althen* decision, thus providing a growing database of natural-language decisions and logic models that will be useful not only in research and in legal education, but also in actual legal practice.

## 2.2 Rule trees: capturing core legal rules

The LLT Lab utilizes the Default-Logic Paradigm as the formal methodology for its modeling (Walker 2007a, b, 2009a). (For the formal needs of argumentation mining, and the major alternative theories valuable for argumentation mining, see Palau and Moens 2009, pp. 99–100.) The Default-Logic Paradigm is designed to provide a flexible and practical logical framework for analyzing legal reasoning in reported cases and for formulating strategies for proving issues of fact in future cases. For the descriptive task of modeling documented decisions, the Paradigm provides a minimum set of logical tools that is functionally sufficient to capture most reasoning. New logical elements can be developed and added as the need arises.

The first step in applying the Paradigm in the Vaccine/Injury Project is to model the logical structure of the legal rules on causation applicable during the study sample. The Default-Logic Paradigm models systems of rules as inverted "rule trees." Figure 1 shows part of the rule tree for the VICP, including the three sub-issues that the petitioner must prove under the *Althen* rule for causation: (1) that a "medical theory causally connects" the vaccination and the injury; (2) that a "logical sequence of cause and effect" shows that the vaccination "was the reason for" the injury; and (3) that a "proximate temporal relationship" exists between the vaccination and the injury. (*Althen*, 418 F.3d at 1278.) The top node of the tree is the ultimate issue to be proved by the petitioner. Each level of each branch extending downward from the top node states the logical conditions for proving the immediately higher proposition. Inferences proceed upward. The use of such a tree structure to represent a set of inferences is a very common form of logical representation (Ashley and Brüninghaus 2009, p. 135, Fig. 3), so the lessons learned from the process of extracting logical structure discussed in Sect. 3 are applicable to many analytic systems, not just the Default-Logic Paradigm. As Sect. 3 will demonstrate, the rule tree plays an important role in "top–down" analysis of evidence assessment. (The rule tree, to which evidentiary assessment attaches, helps to address the problem of determining argumentation structure, discussed by Palau and Moens 2009, pp. 101–102, 105.) Moreover, there is empirical evidence to suggest that the information in the rule tree can add to the predictive accuracy of the model (Ashley and Brüninghaus 2009, p. 151).

Figure 1 also illustrates three logical connectives: "AND" (the conclusion is true if but only if all supporting conditions are true); "OR" (the conclusion is true if but
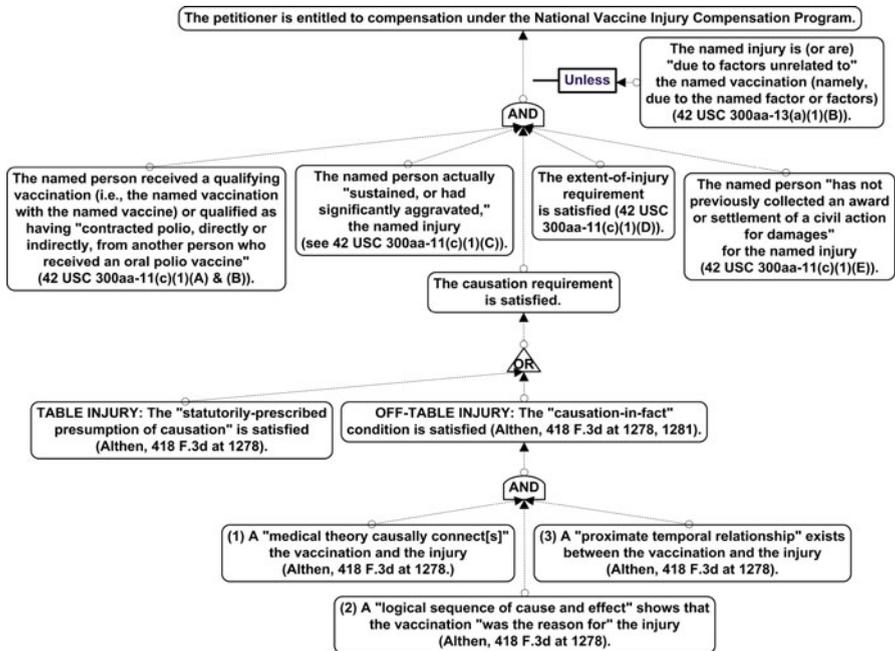
**Fig. 1** Part of the VICP rule tree, containing the three *Althen* sub-issues for causation-in-fact, and illustrating three logical connectives

only if at least one of the supporting conditions is true); and "UNLESS" (if the defeating condition is true, then the conclusion is false, even if the supporting *prima facie* conditions are true). A fourth logical connective used to model legal rules is "RULE FACTORS," which is used when the rule requires the factfinder to take the listed factors into account when making a finding, but does not specify an algorithm for assigning a truth-value to the conclusion as a function of the truth-values of those factors. The overall shape of a rule tree tends to be triangular, with the single ultimate issue as the apex at the top, which is dependent for its truth or falsehood upon combinations of the factual issues that terminate the branches of the tree, along the triangle's base at the bottom. (For an alternative approach to connecting premises, whether "support" or "against," to conclusions, see, e.g., Palau and Moens 2009, pp. 102–105.)

Every proposition in a rule tree has one of three values: "true"/"undecided"/ "false". When a legal proceeding begins, all propositions in the rule tree are "undecided." In any particular vaccine-compensation case, the petitioner and the Secretary of Health and Human Services produce evidence and try to prove or disprove the applicable issues of fact. The special master's role as factfinder is deciding which evidence is relevant to proving which issues of fact, evaluating the plausibility of the evidence, organizing the evidence and making reasonable inferences, and making findings of fact. Often this requires deciding between two

conflicting expert opinions or integrating expert and non-expert evidence into a single line of reasoning. When the factfinder makes findings of fact about the "terminal propositions" at the ends of the branches, the tree's logical connectives propagate those values upward, determining the truth-values of propositions throughout the tree, including the truth-value of the ultimate issue at the top (in vaccine cases, either awarding compensation or not).

## 2.3 Modeling the logic of evidence assessment

This sub-section provides a very brief overview of the modeling process for any particular decision. Lab researchers first select the appropriate rule tree containing the legal rules of causation governing the particular case to be modeled. Although the three-pronged test of causation enunciated by *Althen* has remained stable over the ensuing years (see Sect. 2.1), the Lab follows normal legal practice in always using a rule tree updated to incorporate reference to all relevant Federal Circuit appellate decisions that had been decided by the date of the vaccine decision being modeled. (By October 2010, for example, there were eight such appellate decisions.) Researchers then begin to model the evidence assessment, by extracting from the decision all assertions that the special master reported as relevant to the three conditions of causation. Such assertions may be statements made by testifying witnesses, statements contained in documents, or statements of the special master (e.g., describing an evidentiary exhibit or a witness's demeanor). The special master's findings of fact are also modeled as assertions. Assertions are assigned plausibility-values using an ordinal, seven-valued scale, with the values being: "highly plausible"/"very plausible"/"slightly plausible"/"undecided"/ "slightly implausible"/"very implausible"/"highly implausible". In vaccine-compensation cases, where findings are made by a preponderance of the evidence, an ordinal scale with seven values provides sufficient evaluative precision to achieve the goals of the Vaccine/Injury Project, discussed in Sect. 1. (Also see Walker 2007a.)

The next step is to attach the special master's findings of fact to the appropriate propositions in the rule tree. Researchers then attach to those findings the special master's supporting reasoning from the evidence, organizing the extracted evidentiary assertions into logical structures using "plausibility connectives." Plausibility connectives are similar to logical connectives, but operate on seven plausibility-values. The plausibility counterpart to AND is "MIN," which assigns to the conclusion the lowest plausibility-value possessed by any of its conditions (supporting premises). The "MAX" connective (the counterpart to OR) assigns to the conclusion the highest plausibility-value possessed by any of its conditions (supporting premises). The plausibility connective "REBUT" is the counterpart to UNLESS: if the rebutting assertion is plausible to any degree, then the REBUT connective assigns to the conclusion a degree of implausibility inverse to the degree of plausibility of the rebutting (defeating) assertion. For example, if the rebutting defeater is "highly plausible," then the conclusion is "highly implausible"; but if the defeater is only "slightly plausible," then the conclusion is only "slightly implausible." The project also uses the connective "EVIDENCE FACTORS" to

model reasoning where the special master indicates that she or he took certain assertions into account, but where the decision does not provide information about how the special master organized the assertions to produce a logical chain of reasoning.

Finally, researchers assign plausibility-values to the terminal assertions in the model, as well as to sets of evidence factors, so that the plausibility-values and truth-values throughout the model reflect the evaluation of the factfinder. (The approach for assuring reliability and validity of modeling is discussed in Sect. 3.) In deciding a case, the special master explicitly or implicitly evaluates the relevant evidence with respect to plausibility, and uses that evidence to reason to findings of fact. The objective of every model is to accurately represent the reported reasoning of the factfinder in the legal decision. Doing so involves extracting the relevant evidentiary assertions, assembling them into logical structures of proof using the plausibility connectives, and assigning plausibility-values that reflect the actual assessment of the factfinder.

In producing logic models of fact-finding reasoning, the LLT Lab uses the Legal Apprentice software from Apprentice Systems, Inc. The software creates XML-formatted files of the logic models, which is a standard format used in Internet-based programs. The project also makes such models available in HTML format. At the present time, the software provides an efficient environment in which researchers manually produce the logic models of decisions. The discussions of possible automation throughout Sect. 3 indicate some approaches to developing automation to assist and gradually replace manual modeling. Figure 2 is a screen shot of the Legal Apprentice model for the vaccine decision *Casey v. Secretary of Health and Human Services*, Case No. 97-612V (December 12, 2005). On a color computer display or a page printed in color, the round icon before each assertion has a color that indicates the plausibility-value of the assertion on that line (shades of green for plausibility, white for undecided, and shades of red for implausibility). Figure 2 illustrates a finding of the special master that "there is an adequate medical theory of causation," supported by two alternative lines of reasoning (connected by the plausibility connective MAX) based on two causal pathways (direct viral
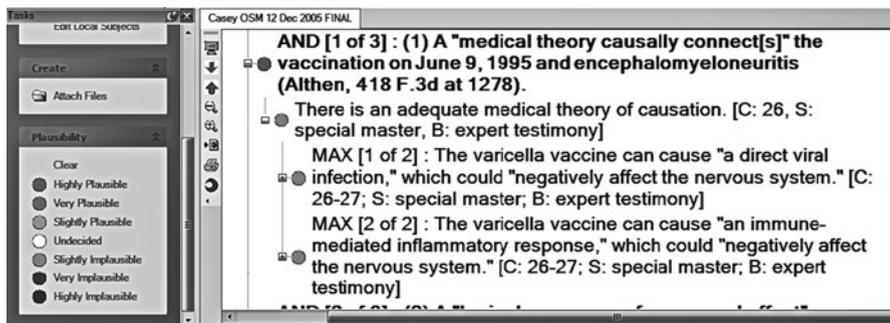


**Fig. 2** Illustration of a portion of the logic model for the *Casey* decision in the Legal Apprentice software

infection and immune-mediated inflammatory response). Section 3 will provide many examples drawn from the models in the V/IP Corpus.

## 2.4 Abstracting reasoning patterns from the V/IP Corpus and constructing plausibility schemas

Once descriptively accurate models have been constructed, a major objective is to abstract from the logic models those patterns of reasoning that do or might recur in multiple decisions. It is hypothesized that patterns and plausibility schemas will sometimes reflect a particular balancing of the epistemic objective (accurate and warranted findings) against non-epistemic objectives (e.g., encouraging vaccine production, administrative efficiency). For example, one hypothesis is that the special masters, as knowledgeable and repeat factfinders, would have a tendency to explain their fact finding under the first prong of *Althen* (namely, that a "medical theory causally connects" the vaccination and the injury) using the rationales generally accepted by scientists. In tension with this tendency, however, is a recognition that requiring scientific proof would result in very few compensation awards. In the vaccine decisions we find an interesting interplay between fact finding and policy implementation. As the *Casey* decision explained:

> In essence, the special master is looking for a reputable medical explanation of a logical sequence of cause and effect (Grant, 956 F.2d at 1148), and medical probability rather than certainty (Knudsen, 35 F.3d at 548–549). As the Federal Circuit explained in Knudsen, medical probability means biologic credibility or plausibility: "Causation in fact under the Vaccine Act is thus based on the circumstances of the particular case, having no hard and fast per se scientific or medical rules." 35 F.3d at 547.

*Casey*, at 9. Discovering how special masters translate such policies into fact finding requires empirical investigation.

When the data display fact-finding patterns, the Vaccine/Injury Project tries to develop plausibility schemas that can be useful in future cases. While patterns of reasoning might describe how factfinders have repeated the same method of assessing the evidence, "plausibility schemas" involve a refinement of those patterns, and may reflect a normative critique of the actual reasoning. In logic, a "schema" is a formal linguistic pattern containing variables, such that appropriate substitutions for the variables create instances of the pattern (Walker 2007a). "Plausibility" schemas are general patterns of evidentiary reasoning, consisting of evidentiary assertions and plausibility connectives. When a plausibility schema is instantiated with substitutions for the particular case, it warrants a default inference that the conclusion is plausible, provided the assertions that function as premises are plausible. Plausibility schemas can therefore state the conditions under which the residual uncertainty (potential for error) in making the inference is acceptable.

For example, a preliminary plausibility schema for fact finding under *Althen* prong 1, based on reasoning in *Casey*, is shown in Example 2.4 (variables for substituting subjects are enclosed in {} brackets):

**Example 2.4:**

**Conclusion:** {The named vaccine} can cause {the named medical condition}.

    **MIN [1 of 2]** {The named vaccine} contains a live but attenuated virus.

    **MIN [2 of 2]** A wild virus of the same type as the virus contained in {the named vaccine} can cause {the named medical condition}.

    **REBUT** There is sufficiently probative defeating evidence.

        **MAX [1 of 2]** The virus in {the named vaccine} cannot reproduce once inside the human body.

        **MAX [2 of 2]** The theoretical possibility of causation is disproved.

When subjects are substituted for variables in a particular case and the resulting *prima facie* premises (the two assertions connected by MIN) are plausible, this schema produces a warranted inference and plausible conclusion. For example, in *Casey*, where the Special Master found for the petitioner, the vaccine involved was the varicella vaccine and the medical condition was a direct viral infection. The petitioner's expert established the two MIN assertions (the *prima facie* case) to the satisfaction of the Special Master. The government as respondent had the option of proving the rebutting assertion, by establishing either of the two assertions connected by MAX. In *Casey*, however, the government's expert did not dispute the theoretical possibility of causation, but merely contended that such a reaction would occur very rarely. Thus, the government failed to rebut the petitioner's *prima facie* case.

This particular plausibility schema reflects a pro-petitioner policy, and does not impose on the petitioner a scientifically rigorous standard of proof. An empirical research question is how such an argument *in fact fares* in *other* cases before *other* special masters, and a normative question is how it *should fare*, given the policies of the Vaccine Act. This article does not pursue either of those questions—this plausibility schema merely furnishes an example of how the V/IP Corpus provides an empirical basis for abstracting patterns of reasoning and generating plausibility schemas, particularly those that integrate fact finding with legal policy. It is hoped, however, that schemas discovered within the V/IP Corpus can furnish the basis for guiding and critiquing the reasoning of factfinders in future cases, by explaining how a reasonable factfinder could assess the available evidence and make findings of fact. They can also inform policy makers of how statutory wording is being interpreted—and is likely to be interpreted—when the statute is applied to particular cases. Finally, such schemas also increase the likelihood of automating various sub-tasks of fact finding. This effort to develop plausibility schemas will be assisted by, and will complement, the ongoing development of argumentation schemes in logic and artificial intelligence. For example, the development of argumentation schemes by Walton et al. on appeals to expert opinions (2008, pp. 13–34, 87–94) is highly relevant to the kinds of evidence found in the V/IP Corpus.

## 3 Lessons for automating the extraction of logical structure from fact finding

In order to abstract recurring patterns from actual decisions, we need a corpus of case models that accurately reflect the reasoning found in actual decisions. A major objective of the Vaccine/Injury Project is to develop and test protocols for generating, from the full text of legal decisions, logic models that are sufficiently reliable and valid. "Protocols" here are merely written directions that provide guidance for modelers in constructing logic models and checklists for quality assurance of the modeling. "Reliability" in this context refers to the degree of consistency between models when different researchers model the same decision, and "validity" refers to the degree to which a model accurately captures the reasoning reported by the factfinder (Walker 2004). Once we develop protocols by which human researchers can generate reliable and valid models for natural-language decisions, then we can hope to automate specific sub-tasks of that modeling process, or at least develop software that will assist human researchers and provide some methods of increasing quality assurance. This section of the article describes those modeling protocols and sub-tasks. Understanding the protocols is necessary in order to understand the nature of the logic models contained in the V/IP Corpus. All examples (corpora) are taken directly from the Corpus.

For developing and testing reliable protocols for logic models, the project assigns a pair of researchers to each decision in the study sample. LLT researchers are upper-level law students trained to apply the Lab's modeling protocols. Each researcher independently reads the decision; extracts the relevant evidentiary assertions; annotates those assertions as to citation, source, and other metadata fields; and assembles those assertions into a model of the factfinder's reasoning. The two researchers then confer on any differences of opinion, document the nature of those differences, and reach a consensus as to the appropriate model. Finally, the Lab Director and the two researchers discuss and document any modeling issues, and decide on the final model. One of the two researchers prepares the final model for publication to the Lab's website, while the other researcher (the more senior or experienced of the two) documents any agreed changes to the modeling protocols. Each time a team models the reasoning in a decision, lessons are learned for developing more reliable and valid modeling protocols—that is, for developing protocols that will produce more accurate models with higher inter-modeler consistency. For example, an iterative process resulted in the refinement of annotations or tags for evidentiary assertions, discussed in Sect. 3.3. (Compare the structured annotation process in Palau and Moens (2009).) In the future, after the protocols have been sufficiently refined, the Lab will design a reliability study from a wide variety of English passages drawn from actual decisions, test the consistency of modeling across multiple modelers, and employ standard statistics for measuring the reliability of the resulting modeling protocols (see Saravanan and Ravindran 2010, pp. 64–66).

For determining the validity of the logic modeling, the project currently relies primarily upon the deliberations among the two researchers and the Lab Director, and ultimately upon the latter's expertise and experience in logical theory, scientific evidence, and legal fact-finding processes. To the extent that the evidence and

reasoning in each decision is unique, it is difficult to employ other methods of establishing validity, such as criterion, content or construct validity (Carmines and Zeller 1979). When the corpus will be used to predict the outcomes of new decisions, success will be testable using experimental designs similar to those employed by Ashley and Brüninghaus (2009). In addition, independent investigators could confirm validation, and evaluation may include seeking confirmation from the special masters themselves.

The important point for this article is that the collaborative modeling process described above provides experience and documentation that is already being used to refine protocols for modeling. Such protocols are useful not only for training new researchers, and for providing quality assurance in modeling, but also for gaining insights into the automation of the modeling process. This collaborative and iterative process has already yielded important insights into the sub-tasks involved in modeling the logical structure of natural-language legal decisions. The following subsections discuss some of the most important insights from this work on protocol development. First, each section identifies a sub-task in the modeling process, and summarizes related challenges to automation. Taken as a set, these sections cover the major steps in the modeling process. Second, each section summarizes some of the project's insights into solving those problems. While the primary focus of any proposed solution is on the formulation of protocols that human researchers can utilize, a secondary objective is always to consider whether the protocols are susceptible to automation. Understanding in an explicit way the instructions for how humans can successfully complete the sub-task should provide important insights into possible approaches to automation.

## 3.1 Identifying and representing legal rules

Legal rules play a central role in legal decision making. The governing legal rules determine what conditions must be satisfied before governmental action is justified, what issues of fact are relevant to the decision, and what inferences from evidence are required or permitted as a matter of law. Those legal rules are generally adopted before the case begins, whether through statute, regulation, or case law, although in some cases new legal rules are adopted in the course of the legal case. It is important, therefore, that a human researcher or any modeling software be able to accurately identify which sentences in the decision state legal rules, as contrasted with those sentences that play a role in applying those rules to the specific case (through evidence assessment and fact finding).

Legal rules are relatively easy to define in terms of logical structure, and human modelers have a fairly high rate of success in identifying sentences that state legal rules, but it is far more difficult to state criteria for identifying rule-stating sentences and to automate their identification in legal decisions. Legal rules are conditional propositions that have been authoritatively adopted as universally applicable and are binding, or have the force of law (Branting 2000; Walker 2007a; Brewer 1996). A conditional proposition has the logical form "if p, then q," where p and q stand for two constituent propositions, and p can itself contain constituent propositions connected by a logical connective. A rule is a statement that finding proposition p

(the condition) to be true warrants finding proposition q (the conclusion) to be true also. As described in Sect. 2.2, the Default-Logic Paradigm models sets of legal rules as inverted rule trees consisting of three-valued propositions as nodes and three-valued logical connectives.

The definition of a legal rule provides several criteria for locating rules in a legal decision: (i) they are conditional propositions; (ii) they have been authoritatively adopted; (iii) they are intended to be universally applicable; and (iv) they have the force of law. In addition, the fact that rules connect together in sets is an important attribute. These five characteristics will be discussed in their order of usefulness in searching documents for legal rules.

First, there is often a convention to cite the relevant legal authority when stating a legal rule. In U.S. legal decisions, for example, the legal authority is often cited immediately after the rule, often using a standard format. For example:

**Example 3.1a (from *Casey v. Secretary of Health and Human Services*, Office of Special Masters, No. 97-612V, Dec. 12, 2005, pp. 8–9):**

To make the requisite showing, petitioner must offer "proof of a logical sequence of cause and effect showing that the vaccination was the reason for the injury." Shyface v. Sec'y of HHS, 165 F.3d 1344, 1353 (Fed. Cir. 1999) (quoting Grant v. Sec'y of HHS, 956 F.2d 1144, 1148 (Fed. Cir. 1992)).

Not all statements of legal rules, however, are necessarily accompanied by citations to authority. The extent to which decision authors follow this convention depends upon the particular author and the legal institution generating the decision. (See e.g., Saravanan and Ravindran 2010, p. 61.)

Second, because legal rules are intended to be universally applicable, they are often stated using only indefinite subjects, without any definite subjects. Indefinite subjects refer to objects or events only as a class or by their general characteristics, and are usually expressed in English by common nouns ("petitioners"), indefinite descriptions ("an injury"), and other grammatical phrases that name groups or classes of individuals. By contrast, definite subjects denote specific individuals, as by proper names ("Ms. Casey") or by definite descriptions ("the Special Master"). Pronouns can denote either indefinite or definite subjects, depending upon the context. There are difficulties, however, both in determining whether a word or phrase is definite or indefinite, and in determining whether the sentence expresses a legal rule. The word "petitioner" in Example 3.1a is accompanied by neither a definite article nor an indefinite article. Yet in context we understand that the sentence expresses a universally applicable legal rule. Moreover, the mere presence of definite subjects in the sentence does not guarantee that the sentence does not express a legal rule. On occasion, a writer of English will efficiently state a rule by substituting definite subjects for the normal indefinite subjects.

Third, while it must always be possible to reformulate an English statement of a legal rule into its logical conditional form, this sometimes requires a great deal of interpretation. Occasionally the decision's drafter will use the words "if" and "then" in an informative way to signal the condition(s) and conclusion of a rule, but this is far from always the case, as the following example illustrates:

**Example 3.1b (from *Casey v. Secretary of Health and Human Services*, Office of Special Masters, No. 97-612V, Dec. 12, 2005, p. 9):**

Without more, "evidence showing an absence of other causes does not meet petitioners' affirmative duty to show actual or legal causation." Grant, 956 F.2d at 1149.

In this example, the rule to be abstracted is that *if* a petitioner in a vaccine compensation case produces only "evidence showing an absence of other causes," *then* that petitioner has necessarily failed to prove causation, as a matter of law. This task of rule-content interpretation may remain one of the more difficult to automate.

Fourth, the fact that legal rules have the force of law means that sometimes they contain deontic (modal) words expressing obligation or permission. For example, expressions of obligation include "must," "shall" and "duty," while expressions of permission include "may" or "permitted." In addition, legal rules sometimes contain phrases denoting deontic roles that are fairly peculiar to law, such as "burden of proof" and "burden of producing evidence." Examples of rule-stating sentences that contain deontic words or phrases are Examples 3.1a and 3.1b. More difficult is:

**Example 3.1c (from *Werderitsch v. Secretary of the Department of Health and Human Services*, Office of Special Masters, No. 99-319V, May 26, 2006, p. 39):**

Close calls are to be resolved in favor of petitioners. Capizzano, supra, at 1327; Althen, supra, at 1280. See generally, Knudsen v. Secretary of HHS, 35 F.3d 543, 551 (Fed. Cir. 1994).

An example of a rule-stating sentence that contains a variation on standard deontic words is:

**Example 3.1d (from *Scott v. Secretary of the Department of Health and Human Services*, Office of Special Masters, No. 03-2211V, Aug. 21, 2006, p. 3):**

Because the injury alleged is not one listed on the Vaccine Injury Table (42 C.F.R. § 100.3), Mrs. Scott has the burden of demonstrating by preponderant evidence that her injury was caused by a vaccination.

This is also an example of a sentence that states a rather direct deduction from the structure of the rule tree: *if* a petitioner like Mrs. Scott alleges an injury not listed on the Vaccine Injury Table, *then* the petitioner has the burden of proving that the vaccine in fact caused the alleged injury. This proposition follows directly and necessarily from the statutory rules discussed in Sect. 2.1.

Finally, the fact that rules occur in integrated sets or systems is an important characteristic in identifying rules. For example, some rules provide definitions for legal terms that occur in other rules. In modeling such a rule in the Default-Logic Paradigm, the defining conditions (the definiens) occur in the immediately lower level of the rule tree, connected to the conclusion (the definiendum) by a logical connective (in the case of definitions, usually AND). As a search strategy for

identifying sentences that state rules, therefore, the words that occur in the terminal propositions of the rule tree provide important search words for identifying any new rules that might be stated or adopted in the decision.

### 3.2 Dynamic subject variables and roles

In the Default-Logic Paradigm, when a rule tree is constructed to represent the legal rules of a statute or case law, certain logical subjects are incorporated as variables, using the subject-predicate distinction of predicate logic (Walker 2007a, pp. 204, 217, 241). For example, in the V/IP rule tree, the petitioner, the vaccination, the vaccine, and the alleged injury vary from case to case, and are represented as variables ("dynamic subjects") within the conditions of the rule tree. Assigning values (specific individuals or instances) to variable subjects in a particular case gives definite meaning to the variables by providing denotations—for example, by assigning "Shannon E. Casey" as the petitioner. In contrast to subject variables, the logical predicates contain "constant" terms, such as "causation" and "the court." Such words are given meaning by other rules, or by their use within the rule tree itself—for example, by further rules that identify a particular court, or relevant factors for proving causation, or the conditions for finding causation. This integration of legal constants and subject variables is consistent with our common understanding of how legal rules apply to particular cases: the legal concepts and their relationships within the rule tree remain invariant for all cases, but the variables acquire values specific to the particular case.

This function of assigning definite subjects as values for subject variables (which, in the English text, appears as a substitution) is one function of quantification in predicate logic. When the same phrase receives the same value throughout a decision or model, this accomplishes a cross-referencing of logical subjects. Throughout the logic model for a particular decision, "the petitioner," "Shannon E. Casey," "Shannon," and many personal pronouns all refer to the same specific individual (Shannon Casey). Similarly, "her encephalomyeloradiculoneuropathy" is "the alleged injury" of Shannon Casey. Automation could definitely assist in developing case-specific lexicons for the various natural-language terms that name the logically important subjects.

When variables occur within rule trees and logic models, it is important to determine their range or scope as being either "global" or "local." In the Vaccine/Injury Project Corpus, for example, there are five "global dynamic subjects," with the assigned specific value substituted for the variable wherever the variable occurs within the model: "the named person," "the named vaccination," "the named vaccine," "the named injury," and "the named factor or factors." (The adjective "named" is inserted in the variable so that inadvertent substitutions do not occur in the model in inappropriate places, such as within quotations.) "Local" subject variables, by contrast, range over only a portion of the model, generally a portion of a particular branch. There are two occasions where this function is particularly useful. First, there may be legal rules that are adopted generally to govern all cases, but which rarely occur in the Corpus. These can be modeled as "rule sub-trees"— pre-formed portions of rule trees that can be inserted into the case model only as and

where they are needed. For example, in the Vaccine Injury Compensation Program, the administratively maintained Vaccine Injury Table defines certain "Table injuries," which grant the petitioner a statutory presumption of causation if certain conditions are alleged and met (see Sect. 2.1). A rule sub-tree that captures additional legal rules can be inserted into the core rule tree as needed in a particular case. Such rule sub-trees may incorporate additional subject variables whose range of substitution is only the sub-tree itself, not the entire model. A second application for local subject variables is in the development of plausibility schemas for evidence assessment. For example, the LLT Lab has developed deductive schemas (such as modus ponens and modus tollens) for insertion as needed into the evidence assessment portion of a case model, and it is in the process of empirically abstracting and formulating policy-based schemas as well (as discussed in Sect. 2.4). Just as rule sub-trees can incorporate local subject variables, plausibility schemas can also incorporate local subject variables (such as "assertion p" or "individual x"), which take values only for the scope of the schema.

The use of subject variables, both global and local, presents both challenges and opportunities for automation. Ashley and Brüninghaus have tested SMILE (a program that classifies sentences of legal text, drawn from trade-secret misappropriation cases, into a set of 26 factors) using "roles-replacement representations," in which case-specific names of parties and products in sentences were replaced with their roles in the case (e.g., "defendant" replacing a specific defendant's name) (2009, pp. 143–144). Their expectation was "that replacing names by roles would help generalize from examples, detect patterns that would otherwise remain hidden, and prevent spurious generalizations based on coincidental occurrences of the same name in different cases." With respect to accurately classifying sentences under factors, their evaluation provided support for the hypothesis that "abstracting from names and individual entities in a case text to their roles in the case allows a learning algorithm to better generalize from training examples" (2009, pp. 154, 158). Thus, there is reason to believe that, for decisions in the V/IP Corpus, utilizing information about dynamic-subject variables (in contrast to case-specific names) could assist the search for recurring patterns within evidence assessment.

## 3.3 Identifying and representing evidentiary assertions

As with legal rules, the basic building blocks of evidence assessment are propositions (statements capable of being either true or false), but the Default-Logic Paradigm models these propositions using seven, ordered plausibility-values instead of three truth-values, and refers to them as "evidentiary assertions," or simply "assertions" (see Sect. 2.3). Examples of evidentiary assertions are statements made by testifying witnesses, statements contained in documents that are admitted into evidence, statements of the factfinder summarizing or characterizing the evidence, and the findings of the factfinder. Unlike legal rules, evidentiary assertions do not have a single logical form that can be useful in identifying them, and they can contain either definite or indefinite logical subjects. Evidentiary assertions can also contain legal terms (particularly within findings of fact), resulting in what are called "mixed issues of law and fact" (see Aman and Mayton 2001, p. 451).

A working guideline at the LLT Lab is that any sentence or clause in a decision that does not state a legal rule, a legal policy, or a legal principle should be modeled as an evidentiary assertion. Even in a corpus where the legal rules are stable and seldom change, therefore, such as in the Vaccine/Injury Project Corpus, it is necessary to identify sentences that state legal rules (see Sect. 3.1), so as to exclude them from consideration as evidentiary assertions. Similarly, one objective of the project is to identify sentences stating legal policies or principles, so that they can be excluded from consideration in the modeling. Examples of statements involving policy from the vaccine cases are:

**Example 3.3a (from *Casey v. Secretary of Health and Human Services*, Office of Special Masters, No. 97-612V, Dec. 12, 2005, p. 26):**

It is precisely because individuals experience adverse reactions to safe vaccines on rare occasions that Congress created the Vaccine Program.

**Example 3.3b (from *English v. Secretary of Health and Human Services*, Office of Special Masters, No. 01-61V, Sept. 28, 2005, p. 6):**

The Federal Circuit explained that requiring medical literature "prevents the use of circumstantial evidence envisioned by the preponderance standard and negates the system created by Congress, in which close calls regarding causation are resolved in favor of the injured claimants." Id. at 1280 (citing Knudsen, 35 F.3d 543, 549 (Fed. Cir. 1994)).

Such sentences may occur with or without accompanying citations to legal authority. Fortunately, in a particular legal area, such as the vaccine compensation cases, the list of relevant policies and principles tends to be fairly short and it can be compiled by examining decisions in a reasonably small sample. Eventually, of course, the goal should be to develop protocols for identifying statements of policy and principle, and eventually automating their detection, both in general and within specific databases. Such protocols will become even more important when empirical research begins to extend beyond the logic of rules and evidence assessment, and into the logic of rule justification on the basis of legal policies or legal principles.

Once sentences stating evidentiary assertions have been identified in a particular legal decision, they must be represented accurately within the logic model. This section discusses the LLT Lab's general approach to representing evidentiary assertions, and then discusses several complications encountered in the V/IP decisions. First, to the maximum extent possible, the content of an assertion is represented in the model by quotations directly from the decision. There are several reasons for adopting this protocol. One is to avoid the obvious danger of paraphrasing the wording of the decision, and failing to include some nuance of meaning that may be important. Another reason is future efficiency in developing natural-language algorithms—it may be more efficient to calibrate and test algorithms directly against case models, before testing the algorithms on the full decision. To the extent that the model contains the natural-language patterns of the original document, it is easier to test algorithms working with the essential language

contained in the models than to work with the complete decision, with all of its linguistic noise. Whether this turns out to be true is an empirical question, but representing assertions by quotation to the extent possible loses little information. Presumably for similar reasons, Ashley and Brüninghaus also encouraged a "cut-and-paste" approach to generating "squibs," which researchers created to summarize the important facts in a case opinion (2009, p. 142). Whereas the models in the V/IP Corpus incorporate assertions into a logical structure of inference, the squibs incorporated the short passages into a "readable narrative."

Second, once the content of the assertion has been identified, the LLT Lab protocols require that certain annotations or tags be added in brackets, which contain logically important metadata about the particular assertion. The principle annotations considered important are: citation, source, modality, and basis. Two examples of annotated assertions represented in the V/IP Corpus are:

**Example 3.3c (from the V/IP case model for *Casey v. Secretary of Health and Human Services*, Office of Special Masters, No. 97-612V, Dec. 12, 2005, p. 13):**

"[T]he virus in the vaccine still can multiply." [Citation: p. 13; Source: finding of Special Master, see p. 26; Modality: "it is reasonable to assume"; Basis: testimony of Dr. Tornatore].

**Example 3.3d (from the V/IP case model for *Howard v. Secretary of Health and Human Services*, Office of Special Masters, No. 03-550V, March 22, 2006, p. 21):**

"Sierra's HIS was caused by a viral infection, unrelated to the vaccination, the first symptoms of which appeared about May 24." [Citation: p. 21; Source: finding of Special Master; Modality: "it is likely that"; Basis: opinion of Dr. Berger]

In these examples, the quotation marks indicate the extent to which the words for the content of the assertion have been extracted directly from the natural-language decision. "Citation" annotates the page number of the decision where the text for the assertion can be found. The "source" tag references the person making the assertion, and often the role of the assertion—for example, a source might be a special master making a conclusion or finding, a lay witness or medical expert testifying for the record, or a notation found in a medical record or published medical article. "Modality" captures the textual justification for assigning a particular degree of plausibility to the assertion. Models of reasoning in the V/IP Corpus are designed to reflect the degree of plausibility assigned by the special master as factfinder. The LLT Lab's protocol is to assign plausibility-values from the seven-value scale of the Default-Logic Paradigm (see Sect. 2.3), using the middle values for plausibility ("very plausible") or implausibility ("very implausible") as the default values (the task of accurately assigning values of plausibility as opposed to implausibility is discussed in Sect. 3.8). Assigning a higher or lower degree of plausibility ("highly plausible" or "slightly plausible," respectively) or implausibility requires an explicit or implicit textual indication. For example,

sentences with phrasing such as "it is uncontested that" or "one must conclude with an overwhelming degree of medical certainty" warrant the highest plausibility-value ("highly plausible"), while sentences with modal hedges (such as the phrase "it is reasonable to assume that") justify assigning a lower degree of plausibility ("slightly plausible"). Finally, the "basis" for an assertion annotates the origin of the assertion into the record, or to the grounds given for the assertion's plausibility or implausibility. For example, an assertion by a medical expert regarding causation might be based on medical records, medical literature or previous professional experience.

This two-step approach to representing English sentences as evidentiary assertions (namely, identifying the content and then annotating the logical metadata) runs into difficulty when representing actual English sentences in complete decisions. Three recurring challenges will be mentioned here. First, an important grammatical aspect of English and other natural languages is the embedding of propositions within propositions. Embedded propositions are constituents of other propositions, for example:

**Example 3.3e (from *Casey v. Secretary of Health and Human Services*, Office of Special Masters, No. 97-612V, Dec. 12, 2005, p. 13):**

However, Dr. Tornatore stated that it is reasonable to assume that the virus in the vaccine still can multiply.

In this example, the clause "the virus in the vaccine still can multiply" is embedded in the clause "it is reasonable to assume that …" (stating a modal hedge), which is embedded in turn in the main sentence "However, Dr. Tornatore stated that …" (stating the original source of the assertion). In legal decisions, the author is often merely recounting the testimony of a witness, and many sentences contain embedded propositions about what the witness said. The modeling problem with embedded propositions is deciding which of the propositions to represent as the content of the assertion in the logic model. Example 3.3c contains the V/IP Corpus model for the sentence in Example 3.3e, and includes the embedded substantive statement as the assertion, annotated or tagged with the remaining information. In that model, the content selected for the modeled reasoning is the most-embedded statement ("the virus in the vaccine still can multiply"), with the statement immediately containing it ("it is reasonable to assume that …") being used to annotate the modality of the assertion, and the main statement ("Dr. Tornatore stated that …") being used to annotate the basis for the assertion. In the logic model, the assertion is annotated with the Special Master as its source, because the Special Master adopted the assertion as a conclusion.

Which clause in a sentence should provide the content of the assertion in the logic model, however, depends upon the inferential context. (For a similar problem when dealing with quotations within texts, see Saravanan and Ravindran 2010, p. 59.) In some contexts, the important evidentiary information might be the fact that a particular witness made the statement, rather than the statement itself. For example, the factfinder in a particular case might have decided that Dr. Tornatore was more credible in his assertions than the opposing expert witness. In such a case,

the factfinder's reasoning might rest on the assertion that "Dr. Tornatore stated that it is reasonable to assume that the virus in the vaccine still can multiply," which should then become the content of the assertion in the logic model. Thus, a challenge for automation is that the reasoning context often provides important information about the content of the assertions that are constituents of the reasoning, and about whether an embedding proposition or an embedded proposition should provide the content of the assertion in the model.

A second challenge with representing English sentences as evidentiary assertions is identifying the "source" of the assertion in the assertion's metadata. As will be discussed in more detail in Sect. 3.4, dealing with findings of fact, the primary objective of modeling in the Vaccine/Injury Project is to capture the findings and reasoning of the special master as factfinder. It is therefore critically important to identify the special master as the "source" of all assertions in the model that represent the special master's own reasoning. The plausibility-values assigned to assertions in the model (discussed in Sect. 3.8) represent the plausibility-values actually assigned by the factfinder—not the plausibility determinations of a witness or of the modeler. Unfortunately, decision authors often state the testimony of a witness while implicitly adopting the witness's reasoning as the author's own reasoning. In such situations, the correct model not only contains the embedded proposition that the witness asserted as the content of the model assertion, but also contains as metadata the factfinder as the source, rather than the witness. For example, if in the context the researcher has good reason to believe that the special master agrees with and adopts the assertion of Dr. Tornatore in Example 3.3e, then the appropriate representation is that found in Example 3.3c. The source of the assertion in the model is the special master ("finding of Special Master"), and the "testimony of Dr. Tornatore" is the basis of the assertion, not its source. Automating this modeling sub-task is rendered more difficult because the reasoning context must supply important information, not merely the single English sentence being represented. On the other hand, automation is assisted when the modeling context provides information that is useful in representing assertions—such as contextual or "top–down" information about the rule tree (Sect. 3.1), the findings of fact of the special master (Sect. 3.4), the structure of the special master's reasoning in particular branches of the logic model (Sect. 3.6), and possible plausibility schemas in vaccine decisions (Sect. 2.4).

A third challenge for modeling assertions is identifying "generalizations," which are a particularly important type of evidentiary assertion in inference (Schum 1994; Toulmin et al. 1984; Walton 1996; Walker 2007a). Examples of generalizations are: "most witnesses testifying under oath tell the truth," "one-third of Americans are overweight," and "20% of the at-risk population develop the disease"—which have the following logical forms (respectively): "most As are Bs," "X/Y of As are Bs," and "X% of the members of group A are (also) members of group B." A generalization typically includes an asserted degree of logical quantification over the reference class (such as "most," "X/Y," or "X%"), and sometimes contains an explicit modal "hedge" limiting the generalizability of the entire assertion—such as expressions of frequency ("often"), typicality ("normally"), temporal limitation ("in the past"), or degree of confidence

("perhaps") (Walker 2007a). Identifying English sentences that state generalizations, however, as well as adequately representing in a model all of the important information contained in generalizations and all relevant metadata, presents important challenges for developing protocols and eventual automation. A lexicon of quantification and modal terms, together with syntactic parsing for terms referring to indefinite subjects (common nouns, adjectives, indefinite descriptions), might provide useful search tools for generalizations, and help to distinguish generalizations from (for example) legal rules.

## 3.4 Modeling findings of fact and the core structure of counterarguments

A critical subset of the assertions in a decision consists of the factfinder's findings of fact. The terminal propositions at the ends of all the branches of the rule tree identify the factual issues to be decided. The next step in modeling a particular decision is to identify and model all of the evidentiary assertions that are findings of fact on those issues. This is followed by selecting the logical structure of the reasoning for and against those findings of fact. These two steps are related, and this section discusses the challenges for automating this sub-task in the modeling process.

The most important findings in a decision in the Vaccine/Injury Project Corpus are the ultimate finding of fact at the top of the rule tree (whether or not "the petitioner is entitled to compensation under the National Vaccine Injury Compensation Program"), and the findings on the three branches of the causation test under the rules of *Althen*: (1) that a "medical theory causally connects" the vaccination and the injury; (2) that a "logical sequence of cause and effect" shows that the vaccination "was the reason for" the injury; and (3) that a "proximate temporal relationship" exists between the vaccination and the injury (see Sect. 2.2). Any accurate model of the reasoning of the factfinder in a V/IP case must, at a minimum, capture the findings of fact on these major issues. The sub-task is identifying sentences that state the findings of fact of the special master on these issues, modeling those sentences as evidentiary assertions, and attaching those assertions to the appropriate branches using the appropriate logical structures.

In identifying a finding of fact for a particular issue, an important consideration is consistency with higher findings, with other findings in the decision, and particularly with the ultimate outcome of the case. For example, if the ultimate outcome or finding is that the petitioner is entitled to compensation, and the case is an "Off-Table" case (see Sect. 2.1), then there must be findings in favor of the petitioner on all three conditions of the *Althen* test of causation. We know that, absent a glaring logical error on the part of the factfinder, there must be a finding for the petitioner under each of these three factual issues. If, by contrast, the ultimate finding is that the petitioner is not entitled to compensation, then the petitioner has failed to persuade the factfinder on at least one necessary element of the *prima facie* case. It would be perfectly reasonable (and efficient) for the special master to make a single subsidiary finding against the petitioner, on only one of the necessary conditions for causation, and to leave all other issues undecided. So determining the

ultimate outcome in the case informs the question of how many subsidiary findings to search for. Given the selection criteria for cases to include in the V/IP Corpus (discussed in Sect. 2.1), an ultimate decision *for* the petitioner ordinarily indicates that findings in favor of the petitioner exist under each of the three *Althen* prongs, and an ultimate decision *against* the petitioner indicates that one or more of the three *Althen* prongs contains a finding against the petitioner. The first priority, therefore, is to identify the finding in the decision on the ultimate issue of compensation.

There are generally many reliable cues for locating and modeling the ultimate finding in a V/IP decision. (Compare Saravanan and Ravindran 2010, p. 59; Mochales and Moens 2008, p. 15.) Experience shows that a special master generally recites the ultimate finding in the last paragraph or two of the decision, and often within a section entitled "Conclusion," although some (also) recite it near the beginning of the decision. Moreover, identification is assisted because the special masters generally use ritualistic words such as "finds" or "finding" (or related words with similar meanings, such as "determines," "concludes," or "proved"), in addition to formulaic words and phrases that track the language of the statute captured in the rule tree (such as "entitled to compensation"). (On the use of the adjective "formulaic" in a similar context, see Ashley and Brüninghaus 2009, p. 145.) Despite the variability of styles of individual special masters in stating subsidiary findings, there is a lower degree of variability in stating the ultimate findings.

There is considerably more variability of author style, however, in stating findings of fact for the three conditions of causation. The following examples illustrate the great variability found in sentences relevant to the second *Althen* condition—proving that a "logical sequence of cause and effect" shows that the vaccination "was the reason for" the injury:

**Example 3.4a (from *Stewart v. Secretary of the Department of Health and Human Services*, Office of Special Masters, No. 06-287V, March 19, 2007, pp. 38–39):**

The logical sequence of cause and effect is that hepatitis A virus has been causally linked to cerebellar ataxia and the vaccine contains an inactivated form of the hepatitis A virus to which, in rare cases, individuals may respond with cerebellar ataxia.

**Example 3.4b (from *Bowes v. Secretary of Health and Human Services*, Office of Special Masters, No. 01-481V, Sept. 8, 2006, p. 9):**

The evidence is sufficient, however, in my view, to show that it is probable that these specific types of vaccinations can cause ATM in extremely rare instances, and that Michael's own case of ATM was vaccine-caused.

**Example 3.4c (from *Casey v. Secretary of Health and Human Services*, Office of Special Masters, No. 97-612V, Dec. 12, 2005, p. 27):**

Thus, there is a logical sequence of cause and effect connecting petitioner's varicella vaccination to her subsequent injuries.

Example 3.4a contains a declarative statement of the "logical sequence of cause and effect," but does not contain definite subjects referring to the specifics of the case; 3.4b does not use these formulaic words but in fact ascribes such a sequence to the particular case; and 3.4c contains both the formulaic words and the definite references. To overcome variability, accurate automation will probably search for ritualistic words indicating that a finding is being made, as well as search for the formulaic legal terminology in the rule condition to which the finding relates. In addition, contextual information about the evidence assessment in other parts of the rule tree is important. Determining the ultimate outcome of the case is critical. Moreover, in the above examples involving the second *Althen* condition, the phrasing from the finding and reasoning under the first *Althen* condition specifying a general medical theory (e.g., "immune-mediated inflammatory response") can assist the search for findings under the second condition. Another useful indication of a finding is if it contains definite subjects referring to the specific persons or events in the particular case (which a rule statement itself would seldom contain). Taken together, such ritualistic and formulaic words, contextual information, and linguistic cues often provide sufficient guidance, but difficult ambiguous situations still arise. (For a discussion of a similar coordinated approach to argumentation mining, see Palau and Moens 2009, pp. 103–106.) As with other modeling issues, the variety of author styles and the variety of English grammatical forms provide a challenge to automation, but knowledge of the logical context can provide valuable assistance. This is also a reason why empirical research on an actual corpus of decisions is necessary, and a suggestion for how software can improve by learning the writing styles of individual authors.

After identifying a sentence in a decision that states the finding of fact that corresponds to a particular rule condition, the next sub-task is to represent the finding by an appropriate assertion in the model, along with the appropriate metadata. Section 3.3 contains a discussion on how to model a finding of fact, and Example 3.3c provides an illustration. Clearly, it is essential to determine whether a finding is in favor of the petitioner or against the petitioner—e.g., to determine accurately whether the petitioner is entitled to compensation or is not entitled to compensation. A critical function of any software to automate modeling must be to accurately identify negations within findings of fact (see Ashley and Brüninghaus 2009, p. 145.)

After identifying a finding of fact for a particular rule condition and modeling it as an assertion of the factfinder, the next sub-task is to select a logic structure for attaching that assertion (finding) to the condition of the rule tree. The paramount goal is achieving descriptive accuracy with respect to the reasoning expressed in the decision. Many findings on contested issues of fact follow a standard logic. Typically, the party that has the burden of proving the issue of fact (or a witness on behalf of that party) asserts that the rule condition is true (satisfied), while the opposing party (or a witness on behalf of that party) asserts that it is not true, and the factfinder finds in favor of one of these two sides—that is, adopts a finding in favor of one party or the other. Thus, for many contested issues of fact, the standard model has the core structure shown in Example 3.4d, and the finding of fact is either the assertion of the proponent or of the opponent.

**Example 3.4d:** {Contested rule condition}

{Assertion of the proponent}

**REBUT** {Assertion of the opponent}

As an illustration, consider the following excerpt from a case model for the second condition of the *Althen* test of causation:

> **Example 3.4e (from the V/IP case model for *Birdsell v. Secretary of the Department of Health and Human Services*, Office of Special Masters, No. 04-1755V, May 30, 2006):**
>
> **Rule Condition:** (2) A "logical sequence of cause and effect" shows that the vaccination on December 20, 2001 "was the reason for" idiopathic thrombocytopenic purpura or ITP (*Althen*, 418 F.3d at 1278).
>
>> The "hepatitis B vaccine is the most plausible cause of Heather's ITP." [Citation: p. 5; Source: Dr. Richard I. Schiff, a clinical immunologist who testified for petitioner; Basis: Heather's history and medical literature]
>>
>>> **REBUT:** "Heather's preceding cold, rather than the vaccination in December 2001, was the cause of her ITP." [Citation: p. 12; Source: finding of Special Master; Basis: Drs. Schiff, Nachman and Ward]

In this example, the Special Master adopted the reasoning of the government, offered in rebuttal to the petitioner's reasoning. This is indicated by the annotation of the rebuttal assertion showing the "finding of Special Master" as the source. In addition, the model would assign a positive plausibility-value to the rebuttal, which in turn would assign a value of "false" to the rule condition.

There are variations on this standard logical structure, however, particularly when the factfinder decides against the proponent of the issue. For example, in some cases it is more faithful to the reasoning of the special master to attach the proponent's allegation to the rule condition, then model the proponent's supporting reasoning, and then attach the opponent's and factfinder's REBUT defeaters to one or more of the proponent's supporting reasons. This is illustrated in the following format:

**Example 3.4f:** {Contested Rule Condition}

{Assertion of the Proponent}

**MIN [1 of 2]**{Premise 1 of the Proponent}

**REBUT** {Assertion of the Opponent}

**MIN [2 of 2]**{Premise 2 of the Proponent}

With this core logical structure, assigning a positive plausibility-value to the rebuttal assertion will still propagate a value of implausibility to the proponent's assertion, and therefore a truth-value of "false" to the rule condition, but as the result of a targeted rebuttal to one of the necessary premises, rather than as a result of an affirmative rebuttal to the main assertion. How to decide which pattern more

accurately represents the reasoning of the special master in the particular case is a challenge for designing protocols for human researchers, let alone for writing algorithms for automating software. This challenge of selecting the appropriate core logical structure is informed by the general theory of modes of attacking assertions (see e.g., Pollock 1990; Prakken 2008; Walton et al. 2008, pp. 30–34).

3.5  Sorting assertions into branches: "horizontal relevance"

Once we have identified findings of fact for the major issues in the case, and have formulated the core structure of the reasoning under those issues (Sect. 3.4), we are in a better position to sort the remaining evidentiary assertions with respect to "horizontal relevance." "Relevant evidence" is defined in United States law as "evidence having any tendency to make the existence of any fact … more probable or less probable that it would be without the evidence" (Federal Rule of Evidence 401). What we call "horizontal" relevance within the Default-Logic Paradigm is the relevance of any particular evidentiary assertion to one or more of the branches of the rule tree—thought of as "horizontal" when the tree expands downward and horizontally, as in Fig. 1. The sub-task is identifying the branch of the tree to which any particular assertion is relevant. This sub-task should be coordinated with the problem of deciding the appropriate depth of modeling, or the number of assertions from the decision to include in the final model (a topic discussed in Sect. 3.7). (For a related discussion of detecting units of argumentation, see Mochales and Moens 2008, pp. 15–16.) Only assertions that are relevant to a rule-tree branch of interest to the modeler need to be identified and represented in the tree. (Compare the task discussed by Palau and Moens of classifying all sentences in the decision as being "argumentative" or "non-argumentative" (2009, p. 103).)

The three conditions for proving causation under *Althen* provide an example of how the content of the rule condition, as well as the content of the assertions attached to that condition (including the finding), can inform the search for relevant evidentiary assertions. The following discussion will address each of these three conditions in turn, and explore possible approaches to automating the search for decision sentences that might be relevant to each.

The first condition for causation that a petitioner must prove is that there is a "medical theory causally connecting" the vaccination and the injury (*Althen*, 418 F.3d at 1278)—that is, that the vaccine "can … cause the type of injury alleged" (*Pafford*, 451 F.3d at 1355–1356). There is normally some general causal theory explaining how a vaccine like the one administered in the particular case could cause this type of injury in the general population. An example with decision sentences relevant to proving such "general causation" is:

**Example 3.5a (from *Stewart v. Secretary of the Department of Health and Human Services*, Office of Special Masters, No. 06-287V, Mar. 19, 2007, pp. 25, 38–39):**

Thirdly, hepatitis A vaccine is designed to induce an immune response that mimics hepatitis A infection, an infection known to cause cerebellar ataxia, which is consistent with causation.

Fifthly, the mechanism is an immune-based central nervous system reaction.

The logical sequence of cause and effect is that hepatitis A virus has been causally linked to cerebellar ataxia and the vaccine contains an inactivated form of the hepatitis A virus to which, in rare cases, individuals may respond with cerebellar ataxia.

Generally, sentences making assertions relevant to the first *Althen* condition contain only indefinite subjects (see Sect. 3.1) and express generalizations, either explicitly or more often implicitly (see Sect. 3.3).

Moreover, in any particular decision, there is normally expert support for the asserted causal theory—generally an expert opinion, along with citations to the scientific literature. For example, the following reasoning supported a conclusion that the hepatitis A virus is a possible cause of cerebellar ataxia, which is in turn relevant to general causation between the hepatitis A vaccine and cerebellar ataxia:

**Example 3.5b (from the V/IP case model for *Stewart v. Secretary of the Department of Health and Human Services*, Office of Special Masters, No. 06-287V, Mar. 19, 2007, pp. 25, 36, 39):**

**Conclusion:** "[H]epatitis A virus has been causally linked to cerebellar ataxia." [Citation: p. 39; Source: conclusion of Special Master]

> **MIN [1 of 3]:** "[M]edical literature shows that hepatitis A virus can cause acute cerebellar ataxia." [Citation: p. 25; Source: conclusion of Special Master; Basis: Dr. Mark's testimony]
>
> **MIN [2 of 3]:** "The Nussinovitch article (R. Ex. G) lists hepatitis A virus as a cause of cerebellar ataxia." [Citation: p. 39; Source: conclusion of Special Master; Basis: Respondent's Exhibit G, M. Nussinovitch et al., "Post-Infectious Cerebellar Ataxia in Children," 42 Clin Pediatrics 7:581–584 (2003)]
>
> **MIN [3 of 3]:** "[T]he Tuthill case report … concludes that subclinical hepatitis A infection should be added to the list of documented causes of cerebellar ataxia." [Citation: p. 36; Source: conclusion of Special Master; Basis: Respondent Exhibit C, D. Tuthill, et al., "Acute Cerebellar Ataxia After Subclinical Hepatitis A Infection," 15 Ped Infectious Dis J 6: 546–557 (1996)]

These three conditions work together to establish the conclusion, and it is reasonable to think that each played a sufficiently necessary role in the reasoning, so they are therefore combined with a MIN connective. The evidentiary assertions that are relevant are those tending to prove or disprove that there is a medical theory that can connect the vaccination to the injury, but not assertions that are specific to the petitioner's injury. The key linguistic indicators are sentences with indefinite subjects, and information connecting the type of vaccine with the type of injury.

The second condition for causation under *Althen* that a petitioner must prove is a "logical sequence of cause and effect" that shows that the particular vaccination "was the reason for" the particular injury in the specific case (see *Althen*, 418 F.3d at 1278)—that is, that the particular vaccination involved in the case was a

"but-for" cause of the particular injury (see *Pafford*, 451 F.3d at 1355–1356). This condition requires evidence that links the specific vaccination to the specific injury. Often, in the decisions three kinds of evidence are produced to satisfy the requirements of this condition—fact testimony by lay witnesses, medical records and expert testimony. First, information about the particular sequence of events in the case is often presented, based upon medical records created by treating physicians. The following example illustrates a typical series of assertions found under the second condition of *Althen*:

**Example 3.5c (from the V/IP case model for *Casey v. Secretary of Health and Human Services*, Office of Special Masters, No. 97-612V, Dec. 12, 2005, pp. 2, 26–27):**

**Conclusion:** "[T]he varicella vaccine caused both a direct infection of her nervous system as well as an immune-mediated inflammatory response in her nervous system," which together "resulted in damage to her brain, spinal cord, and peripheral nervous system (i.e., her encephalomyeloneuritis)." [Citation: p. 26; Source: finding of Special Master]

> **MIN [1 of 4]:** "[P]etitioner was vaccinated against varicella on June 9, 1995." [Citation: p. 27; Source: finding of Special Master; Basis: petitioner's exhibit]
> **MIN [2 of 4]:** "The attenuated virus in the varicella vaccine … directly attacked petitioner's nervous system." [Citation: p. 27; Source: finding of Special Master; Basis: opinion of petitioner's expert]
> **MIN [3 of 4]:** "The attenuated virus in the varicella vaccine … caused an immune-mediated inflammatory response in her nervous system." [Citation: p. 27; Source: finding of Special Master; Basis: opinion of petitioner's expert]
> **MIN [4 of 4]:** "As a result, … petitioner began to experience the onset of symptoms of her encephalomyeloneuritis." [Citation: p. 27; Source: finding of Special Master; Basis: opinion of petitioner's expert]

Second, the petitioner almost always produces an expert's opinion about the particular sequence of events in the case, concluding that it is an instance of the general causal model established in the first condition. The following is an example of an expert opinion that helps support a finding that a petitioner had cerebellitis, a symptom of encephalomyeloneuritis (see last assertion in Example 3.5c):

**Example 3.5d (from the V/IP case model for *Casey v. Secretary of Health and Human Services*, Office of Special Masters, No. 97-612V, Dec. 12, 2005, pp. 27):**

"[C]erebellitis [is] a symptom particularly associated with a natural varicella infection." [Citation: p. 27; Source: finding of Special Master; Basis: Dr. Tornatore's testimony that "cerebellitis is a peculiar and specific characteristic of a natural varicella infection, and is not typically seen with other viruses," p. 14]

Ideally, we expect the sequence of events in the particular case (under Condition 2 of *Althen*) to sufficiently track the sequence expected under the general causal

model (under Condition 1 of *Althen*), and we expect the petitioner's expert to testify that the particular case is in fact an instance of that general model. (This *Althen* condition also lends itself to a story-based approach to proof (Bex et al. 2010, pp. 130–132).)

The third condition for causation under *Althen* that a petitioner must prove is that a "proximate temporal relationship" exists between the particular vaccination involved in the case and the particular injury (see *Althen*, 418 F.3d at 1278). Generally, we expect the petitioner to produce information about the timing between the vaccination and the onset of symptoms, as well as expert testimony about the appropriateness of that timing. First, information about the timing in the case often rests on medical records generated by treating physicians. For example:

**Example 3.5e (from the V/IP case model for *Stewart v. Secretary of the Department of Health and Human Services*, Office of Special Masters, No. 06-287V, Mar. 19, 2007, pp. 34, 37):**

**Conclusion:** "The fourth day after hepatitis A vaccination (counting the day of vaccination as the first day) was Friday, March 29, 2002 when Will's condition became noticeable." [Citation: p. 34; Source: finding of Special Master; Basis: medical records]

> **MIN [1 of 2]:** The "onset of Will's cerebellar ataxia did not occur on March 26, 2002, but some time thereafter." [Citation: p. 37; Source: Dr. Kane's testimony; Basis: watching a videotape made during Will's birthday party on the evening of vaccination]
>
> **MIN [2 of 2]:** "Will's behavior was unremarkable on Wednesday, March 27, 2002, and on Thursday, March 28, 2002." [Citation: p. 34; Source: finding of Special Master; Basis: medical records]

This example illustrates a finding, based on supporting evidence, concerning the actual timing of onset in the case.

Second, we expect there to be an expert's opinion that this timing is consistent with what would be expected if this is an instance of the general causal theory established in the first *Althen* condition. The following example displays part of the reasoning that illustrates this point:

**Example 3.5f (from the V/IP case model for *Stewart v. Secretary of the Department of Health and Human Services*, Office of Special Masters, No. 06-287V, Mar. 19, 2007, p. 39):**

**Conclusion:** There is a proximate temporal relationship. [Source: implicit finding of the Special Master]

…

> **MIN [2 of 3]:** "The appropriate temporal framework for causation between viruses and cerebellar ataxia runs from 1 to 21 days." [Citation: p. 39; Source: finding of Special Master; Basis: A.M. Connolly, et al., 35 Ann Neurol 673–679 (1994)]

**MIN [3 of 3]:** "If Will were exposed to a protein in the vaccine to which he was previously exposed, … four days would be an appropriate interval for causation." [Citation: p. 39; Source: finding of Special Master; Basis: opinion of Dr. Marks]

Examples 3.5e and 3.5f, taken together, illustrate the reasoning on Condition 3 of *Althen* in the *Stewart* case. The key linguistic indicators for Condition 3 are sentences about the timeframe of the injured person's injury and evidence showing that the timeframe does or does not fit within the general causal theory established in Condition 1.

This section has discussed and illustrated the concept of "horizontal relevance." The sub-task is to determine, for each evidentiary assertion to be represented within the case model (see Sect. 3.3), the branch or branches of the rule tree to which the assertion is relevant. Under a contested issue of fact (see Sect. 3.4), the sub-task is also to determine whether the assertion favors the proponent or the opponent of the issue. Automation will likely utilize information about the content of the different branches, together with linguistic cues, as discussed above. Once the researcher or future software compiles a list of relevant assertions for a particular tree branch, the next step is to organize the levels of reasoning under that branch—the topic of Sect. 3.6.

### 3.6 Constructing levels of reasoning ("vertical relevance") and selecting plausibility connectives for levels

Once we have identified the finding for a particular condition in the rule tree and have attached the finding to the rule tree using a core reasoning structure (Sect. 3.4), and we have identified the assertions relevant to that particular condition (including all subsidiary conclusions or findings of the factfinder) (Sect. 3.5), then the next sub-task is to organize the supporting reasoning for the finding, including the major arguments urged by the parties and rejected. This sub-task has two important and related aspects: determining which relevant assertions support other relevant assertions (the issue of "vertical relevance"), and selecting which plausibility connective to use to connect multiple assertions that occur on the same level. This section discusses insights and challenges for designing protocols and automating software for both of these aspects.

In general, one option is to simply attach a list of relevant evidentiary assertions to each side of a controverted issue, and ignore the rest of the logical structure reported in the decision. For example, if the core structure of the finding is the standard one illustrated in Example 3.4d, then the modeler might attach one EVIDENCE FACTORS connective to the proponent's assertion, and another EVIDENCE FACTORS connective to the opponent's assertion, and simply list all relevant evidentiary assertions under one connective or the other. In a sense, this appears to be similar to the approach used by Ashley and Brüninghaus, in which "factors" either favor the plaintiff or favor the defendant (2009, pp. 133–139; Chorley and Bench-Capon 2005, pp. 333–336), or the classification of premises by Palau and Moens as "support" or "against" (2009, pp. 102–105). As discussed in Sect. 2.3, the connective EVIDENCE FACTORS allows the modeler to attach the

stated reasons and evaluate them with respect to plausibility, but contains no algorithm for assigning a plausibility-value to the conclusion as a function of the plausibility of those factors. Such an assignment of plausibility must be made manually, after weighing the entirety of the competing lists. (For various algorithms for predicting outcomes based on relevant factors, see Ashley and Brüninghaus 2009, pp. 133–139, 148–149.)

Indeed, in many actual decisions, the decision's author merely recites as supporting reasons the factors that she considered relevant and took into account in reaching the finding, but the author does not report how she combined those factors to reach the conclusion. In such situations, descriptively accurate modeling requires use of the connective EVIDENCE FACTORS, which allows the modeler simply to attach the stated reasons and then assign them the plausibility-value actually assigned by the factfinder. To use a stronger connective in the model would be unwarranted given the reported reasoning. Moreover, listing the stated relevant factors in the model allows more reasoning to be added in support of individual factors, if such reasoning is present in the decision, and saves this important information for future normative critique. This is one illustration of how descriptive accuracy in modeling is a paramount objective over normative analysis—if the logic model accurately represents the actual decision, then analysts can separately make a normative critique of the decision's logic; but if the modeler simply adds logical structure to the model that was not present in the actual decision, then analysts could not reconstruct the original stated reasoning merely by inspecting the logic model.

By contrast, whenever the decision provides more information about the structure of the supporting reasoning, then it is the task of the modeler to capture that structure in the model. The first step is to organize relevant assertions into levels, asking whether some assertions tend to prove or disprove other assertions. In contrast to the concept of "horizontal relevance," which is central to the task of assigning an evidentiary assertion to a particular branch of a rule tree, the concept of "vertical relevance" is central to organizing evidentiary assertions into supporting levels within the same branch of a model. The problem can be posed in this fashion: presented with a (long) list of relevant assertions under a given conclusion, which assertions on the list support other assertions on the list, and therefore belong on a supporting level? And which assertions possess probative value because they are combined with other assertions, and therefore belong on the same level with those other assertions? If Assertion A is relevant to proving the conclusion only because it tends to prove Assertion B, then the model should have Assertion A on some level below Assertion B. If Assertions A and C are relevant to proving the conclusion because they in combination tend to prove the conclusion, then Assertions A and C should be modeled on the same level below the conclusion. (Compare the task addressed by Palau and Moens of classifying argumentative propositions as premises and conclusions (2009, p. 104).) What plausibility connective to select to connect multiple assertions that occur on the same level is the second part of this sub-task.

Automating the construction of levels of support within a set of relevant assertions will be a significant challenge. (See Mochales and Moens 2008, pp. 15–16; Palau and Moens 2009, p. 104; Prakken 2008, pp. 344–349.) The

reasoning patterns being documented for the V/IP Corpus may play an important role, however, in improving automation accuracy. As discussed and illustrated in Sect. 2.4, reasoning patterns can be abstracted from decisions in the Corpus—some being generic logical structures (such as modus tollens), while others are specific to particular branches found only in vaccine decisions (such as the pattern discussed in Sect. 2.4). Such patterns can be used to construct plausibility schemas, which could then be used to guide the structuring of assertions within the model. To use the example discussed in Sect. 2.4, if the software were searching for logical structure within the branch for Condition 1 of the *Althen* test of causation, it could check for assertions about whether the vaccine contains a "live" but "attenuated" virus, and whether the "wild virus" has been causally connected to the type of injury alleged in the case. If such assertions are present, then a working hypothesis would be that such assertions combine together on the same level and tend to establish a finding for the petitioner on that branch of the rule tree. Additional information that may prove useful is whether the ultimate findings on compensation and causation were made for or against the petitioner, as discussed in Sect. 3.8. In general, our current thinking is that the search for the logical structure of reported reasoning should proceed not merely from the "bottom up" (using linguistic information contained within discrete sentences), but also from the "top down" (using a library of documented patterns or schemas containing possible logical structures).

The second step, after sorting relevant assertions into levels by vertical relevance, is assigning a plausibility connective for combining assertions that occur on the same level. A plausibility connective is an operator that ideally assigns a plausibility-value to the conclusion as a function of the plausibility-values of its supporting assertions. Such operators play an important role in normative logical analysis and in automation. Although the connective EVIDENCE FACTORS accurately represents the fact that particular assertions support a stated conclusion, it is the weakest connective in terms of logical power, and provides no algorithm for combining plausibility-values of the connected assertions. Whenever warranted by the decision being modeled, the plausibility connectives MIN and MAX should be used in lieu of EVIDENCE FACTORS because they contain inferential information that is otherwise lost. In general, the MAX connective occurs less frequently than MIN in the V/IP Corpus, because MAX is reserved for connecting two or more alternative and independently sufficient lines of reasoning leading to the same conclusion. Syllogistic and other patterns of logical reasoning support their conclusions by sets of necessary premises (assertions connected together by MIN). The connective REBUT is used only when the assertion defeats the conclusion. The next paragraphs provide examples of each connective.

Whenever the assertions combined on the same level are all necessary for reaching the conclusion, then MIN should be used to connect them to the conclusion. In the final model, an evidentiary assertion has a plausibility-value drawn from the ordinal, seven-valued scale discussed in Sect. 2.3 ("highly plausible"/"very plausible"/"slightly plausible"/"undecided"/"slightly implausible"/"very implausible"/"highly implausible," with "highly plausible" being the highest value). The MIN connective uses the weakest assertion in the set to determine the plausibility of the conclusion, and assigns to the conclusion the lowest plausibility-value found in

the set. Therefore, the MIN connective should be used only when all of the conditions combine into a single argument of support for the conclusion, and each condition is reasonably necessary for reaching the conclusion. Examples 3.5b and 3.5c illustrate use of this connective. In these examples, the least plausible assertion determines the plausibility-value of the conclusion. The MIN connective therefore generalizes the AND connective (conjunction) for use in many-valued logics.

The connective MAX should be used only when the supporting assertions on the same level represent alternate and independently sufficient lines of reasoning. The MAX connective assigns to the conclusion the highest plausibility-value possessed by any of the supporting assertions. The assertions are equally acceptable alternatives, any one of which provides adequate support for the conclusion. Figure 2 in Sect. 2.3 provides an example of a MAX connective from the V/IP Corpus. The MAX connective therefore generalizes the "OR" connective (disjunction) for use in many-valued logics.

Finally, the plausibility connective REBUT, like MIN and MAX, can be thought of as combining two assertions, but the rebutting assertion states a condition for negating the conclusion. A REBUT connective is used whenever one party is trying to prove that an assertion of the other party is incorrect. If the rebutting assertion possesses any degree of plausibility, then the REBUT connective assigns to the conclusion the inverse of the plausibility-value possessed by the rebutting assertion. For example, if the rebutting assertion is "very plausible," then the conclusion would be "very implausible"; if the rebutting assertion is only "slightly plausible," then the conclusion would be "slightly implausible." If the rebutting assertion is implausible or undecided, then it has no effect on the plausibility-value of the conclusion. The REBUT connective therefore generalizes the "UNLESS" connective for use in many-valued logics. For example:

**Example 3.6 (from the V/IP case model for *Scott v. Secretary of Health and Human Services*, Office of Special Masters, No. 03-2211V, Aug. 21, 2006, pp. 25, 29, 31]:**

**Conclusion:** Mrs. Scott "developed neurological symptoms and blood tests that were consistent with [the type of] APS" associated with some viruses and vaccines. [Citation: p. 25; Source: reasoning of Dr. Anderson]

> **REBUT:** "[T]he nature of the antibodies produced and their persistency demonstrated the unlikelihood that the MMR vaccination triggered the development of APS in Mrs. Scott"—that is, "viral- or vaccine-associated APS most frequently presents with significant differences from the type of antibodies present in Mrs. Scott." [Citation: pp. 31, 29; Source: finding of Special Master; Basis: opinion of Dr. Brenner]

In this example, the rebutting assertion defeats the conclusion only if it is plausible; and if it has any degree of plausibility, then the degree of defeat is the inverse of that plausibility.

The sub-task of assigning the appropriate plausibility connective to combine assertions that are on the same level poses significant challenges for automation. It

is one thing to develop a protocol to be used by human researchers—such as stating that every assertion connected to a conclusion by MIN must be reasonably "necessary" to the reasoning—but it is quite another thing to operationalize that protocol for software development. It may be helpful to use information drawn from the context and schemas developed from patterns found in other decisions. Information drawn from the context includes whether the finding on the issue is for the respondent, not the petitioner—which would indicate the presence of at least one REBUT connective within the reasoning. Possible schemas for the branch of reasoning may suggest combinations of connectives to consider. A major goal of the V/IP Corpus is to identify such contextual information, and to abstract such patterns and develop such schemas.

Finally, it is an open hypothesis whether additional plausibility connectives should be defined and utilized. The pragmatic approach to modeling is not to design a new element, even one well-founded in the theoretical literature, until there is a need for the element in modeling actual decisions. For example, while the connective UNDERCUT is well defined (Pollock 1990, p. 79; Prakken and Sartor 2004; Prakken et al. 2003; Walker 2007a), so far it has not been needed in modeling the vaccine compensation decisions. Other examples are modal or temporal logics. While such theoretical elements might well have important uses in accurately representing or usefully improving legal reasoning, the pragmatic strategy is to incorporate more complexity in the modeling paradigm only when, and to the extent that, it is needed. Additional modeling elements add complexity, generally require additional information in application, and often increase the challenges to automation. The pragmatic approach is to incur such added costs only when there are proportional benefits to be gained.

### 3.7 Selecting the depth of modeling

While special masters strive to produce findings of fact and conclusions of law that are accurate and supported by the evidentiary record, researchers modeling the logic of those decisions strive to produce models that accurately represent the essential inference structure of the special master's reasoning, based on information in the reported decision. Once the special master has assessed the evidence and resolved contested issues of fact, then the objective of the written decision is to accurately report sufficient information about that evidence assessment. Once the decision does so, then it is complete. Similarly, a question of completeness arises for the logic model of the reported decision. In principle, a model could include assertions for every relevant sentence in the decision. In practice, however, the vast majority of the text is never represented in the model. (Moreover, if all relevant sentences were indeed represented in the model, then the logical structure of the reasoning would be obscured, thereby defeating one of the principle advantages of modeling.) So the question arises for both human protocols and automation: At what point in the modeling process does the model sufficiently represent the underlying logic of the written decision? At what point should the human modeler or modeling software stop extending branches by adding new levels to the inference tree?

Selecting the appropriate stopping point for a case model is a function of the goals of the modeling exercise—the resulting model should be adequate for the uses to which the model will be put. For the models in the V/IP Corpus, those goals include descriptive accuracy, normative completeness, pattern discovery, and usefulness as an analytic and management tool for human users.

First, descriptive accuracy means that a model for a decision must faithfully capture the logical structure of the reasoning as reported by the decision maker. To understand the logic model is to understand the actual reasoning as reported in the documented decision. The model must not read into the decision reasoning that is not there, yet it should include any assertion that explicitly or implicitly plays a role in the factfinder's reasoning. (On the difficulties of detecting implicit reasoning, see Prakken 2008, pp. 342, 358.) At a minimum, the model should include the necessary information that is reasonably implied by the reasoning that the factfinder has chosen to report. For example, if a petitioner was awarded compensation, then we can infer that all procedural conditions of the case must have been fulfilled (e.g., that the court had jurisdiction to decide the case), even if the special master did not record a finding on each essential procedural element.

Second, normative completeness means that the model contains sufficient information about the reported reasoning, so that a normative critique of the model is also a valid critique of the reported reasoning. If an inspection of the model uncovers a defect in the reasoning, then we should find that same defect in the original decision. So the model should contain not only the threshold level of minimal information, but also sufficient levels of structure to accurately represent the underlying logic of the decision.

Third, pattern discovery (and the resulting production of plausibility schemas) is a critical goal of the research (see Sect. 2.4). Recognizing that patterns of proof recur in different cases deepens our understanding of the issues litigated, and enables more effective and efficient argument in future cases. For example, a model that only includes the basic findings of the special master on each *Althen* condition would accurately display the outcome of the case, but it would not display the reasoning that supported the conclusion. The reader would see the bare justification for the ruling, but would not know which evidentiary assertions were most persuasive to the special master, and why they were persuasive. Patterns and resulting schemas are also expected to play a helpful role in automation, because they can help guide the construction of logical structure (see Sect. 3.6). If a model fails to capture important steps or levels of reasoning in the decision, it will not be easy to recognize patterns and it will be more difficult to develop plausibility schemas.

Finally, the usefulness of a model as a tool of human understanding, and as a management tool in litigation, derives in part from the abstraction of the essence of the reasoning from the noise of the natural-language decision. A model filled with too much information from the written decision would have little value as a practical tool.

Eventual automation of the modeling process, or even providing automated assistance to human researchers, requires a method for deciding when to terminate the search for additional assertions to put into the case model. Experience with the

V/IP Corpus suggests some approaches to solving this problem. First, it appears that formulaic legal terminology from a rule condition occurring within an assertion indicates that the assertion may be playing an important inferential role in the decision. The factfinder's task is to make findings on the issues of fact that are contested by the parties, and those findings are likely to include at least some legal terminology from the rule condition (see Sect. 3.4). It also appears likely, based on experience with the V/IP Corpus, that sentences directly supporting those findings will share some of the same legal terminology. At the other extreme, the numerous sentences in a decision that merely recount evidence from the record (e.g., a lay witness's testimony), without any stated connection to the factfinder's reasoning on any particular legal issue, are likely to contain no legal terminology.

Second, at least in the vaccine decisions of the V/IP Corpus, the location of the sentence within the decision appears to be highly associated with being part of the factfinder's supporting reasoning. (For a contrary experience with Indian court decisions, see Saravanan and Ravindran 2010, p. 50.) For example, being located in the same paragraph as the finding (or in immediately accompanying paragraphs) makes it far more likely that the sentence expresses information related directly to the finding. Moreover, special masters often provide the bulk of their supporting reasoning in a separate section of the decision labeled "Discussion" (or some similar heading), usually located at the end of the decision, just before the "Conclusion" section. If a sentence occurs in the "Discussion" section it is far more likely to play a role in the supporting reasoning, when compared to a sentence occurring earlier in the decision. Individual decision authors have their own styles for writing decisions, and a major advantage of modeling decisions in a decision-making process like vaccine compensation is that there are only a few repeat factfinders, whose styles the human researchers and eventual software can learn. Of course, it is an empirical question whether the vaccine-decision authors are sufficiently representative so that results can be generalized.

At times, however, it is also important for model completeness to incorporate assertions that occur early in the decision. A special master will often simply accept the reasoning of a medical expert, or adopt a statement from a medical record created by a treating physician. In cases such as these, it is appropriate to tag the assertion with the stated basis for it. As discussed in Sect. 3.3, representations of assertions sometimes include a "basis" annotation that points to the origin of the assertion in the record, or to the grounds given for the assertion's plausibility or implausibility. In such situations, the "basis" annotation to the assertion can provide important assistance in the search, and help provide completeness when a model stops. The wording of a basis annotation is not part of the content of the assertion, but it is useful metadata because it points in the direction of the next level of supporting reasoning. This is illustrated in Examples 3.5b ("Dr. Mark's testimony," "respondent's Exhibit G"); 3.5d ("review of the medical records," "examination of petitioner by Dr. Tornatore," "evidence of injury to the brain, spinal cord, and proximal nerve roots"); 3.5e ("watching a videotape made during Will's birthday party on the evening of vaccination"); and 3.5f ("A.M. Connolly, et al., 35 Ann Neurol 673–679 (1994)"). When generating a sufficiently complete model requires retrieving and modeling sentences that occur earlier in the decision, basis

annotations can be useful in conducting searches for those sentences. Another advantage of the basis annotation is that, when the model does stop at some particular level of expansion, the metadata for the terminal assertions suggest the kind of evidence that the next level of support would have involved.

Finally, the V/IP Corpus can provide other guidance for automating an optimal degree of modeling. Models in the database rarely extend beyond four or five levels of evidence assessment, but there also appear to be reasons why certain branches might extend beyond this norm in certain situations. For example, when the opponent contests sub-issues asserted within the argument of a party, with REBUT connectives occurring at deeper levels within the model, completeness may require modeling the reasoning that immediately supports both the challenged assertion and the rebutting defeater. It may be premature to stop modeling supporting reasoning until we have an adequate understanding of the arguments for and against contested sub-issues—at least, those sub-issues and controversies sufficiently critical to the case that the factfinder explicitly discusses them in the decision. Thus, the statistical aspects of the models in the Corpus supply some guidance on the optimal depth of modeling.

## 3.8 Assigning plausibility-values to terminal assertions in the logic model

The final step in the modeling process is to assign plausibility-values to the terminal assertions in the model. Those plausibility-values must accurately represent the evidence evaluation of the factfinder as expressed in the written decision. Moreover, as the plausibility and logical connectives propagate those values up the inference tree toward the ultimate issue at the top, the entire profile of plausibility-values and truth-values must represent the factfinder's assessment of all of the structured issues in the case. Assigning plausibility-values merely to the terminal propositions (and sets of evidence factors) and allowing those values to propagate up the model is also a major test of the logical coherence of the model's structure.

Once the human modeler (and eventually the modeling software) has identified the major findings in the decision and determined their values (for or against the petitioner) (see Sect. 3.4), then this information provides a "top–down" strategy for determining which branches must evaluate as "true" in the final model, which branches must evaluate as "false," and which can remain "undecided." If the ultimate finding in a V/IP decision is that the petitioner is "entitled to compensation," then every essential branch of the rule tree must evaluate as "true." In particular, each of the three *Althen* conditions of causation must evaluate as "true." Which means, in turn, that the special master must have found the petitioner's arguments on those three issues more plausible than the respondent's. If, on the other hand, the ultimate finding is that the petitioner is "not entitled to compensation," then one or more of the essential branches must evaluate as "false" (the remaining branches can remain "undecided"). Thus, a "top–down" analysis of the decision provides information critical to assigning plausibility-values to supporting assertions, as well as to testing the accuracy of the final model.

It would be a mistake, however, to simply "stipulate" appropriate values of "true" or "false" to a branch based upon this information. The goal is to properly model the reasoning that supports and warrants the finding on the issue, and the

finding's plausibility-value should be determined by the reasoning that supports it, not merely by a stipulated value. The Legal Apprentice software used to generate the logic models in the V/IP Corpus allows for stipulation of any proposition or assertion in the model—just as parties in real cases can stipulate the facts—but the stipulation then supplies the value to the proposition or assertion, and the software disregards any values that would otherwise be propagated upward from the supporting reasoning. Thus, a stipulation cuts off any influence from the supporting reasoning. The only time that a model should stipulate a truth-value to a proposition, or a plausibility-value to an assertion, is when there is no supporting reasoning (i.e., it is a "terminal" proposition or assertion).

If a proposition is not stipulated, but its truth-value is determined by the plausibility-value of a finding of fact, which is determined in turn by its supporting reasoning, then at bottom that reasoning rests on the "terminal assertions" at the ends of the supporting chain of reasoning. The next task is to assign plausibility-values to those terminal assertions, and let the plausibility connectives of the reasoning structure propagate the appropriate plausibility-value to the finding, which will propagate the appropriate truth-value to the proposition. This will occur correctly if the logic is properly modeled and the plausibility assignments are properly made. This strategy for evaluating terminal assertions and testing logical structure to see if it propagates values as expected also provides insights into possible automation of this sub-task. The strategy is another example of taking both "top–down" and "bottom-up" approaches. Sentences expressing terminal assertions sometimes contain linguistic cues that indicate whether the factfinder found those assertions to be persuasive and plausible. Far more often, however, it is the context (and only the context) that supplies this information. It is often the case that the question of which witness's testimony the special master finds more credible, and adopts as part of her own reasoning, can only be inferred from the context of the sentence—both the findings ultimately reached and the logical role of the witness's testimony in reaching those findings. Whether the ultimate finding in a branch is for or against the petitioner is information useful in assigning plausibility-values to the terminal assertions in the branch. Moreover, sentences elsewhere in the decision sometimes indicate that the special master found one witness more credible than another one. In such situations, the "basis" annotation recording the origin of the assertion in the record (e.g., which expert originally asserted it) is useful in deciding whether the special master considered the assertion plausible. The V/IP Corpus provides a wealth of materials from which software developers and the software itself can take a "top–down" approach and learn to evaluate assertions and propositions more accurately.

## 4 Conclusion

As this article demonstrates, the V/IP Corpus is a valuable database for developing and evaluating software that can analyze (or at least help humans to analyze) the logical structure of the evidentiary reasoning in legal decisions. The Corpus allows a pairing of natural-language sentences (in their normal context) with assertions representing the logical content of those sentences, within an inferential context.

Thus, the Corpus supplies data for developing and evaluating software for extracting logical structure from natural-language documents, as well as software that would use such logical structure to predict outcomes in similar cases or to formulate arguments for use in similar cases. The Corpus of modeled legal decisions, and the eventual libraries of reasoning patterns and schemas abstracted from those decisions, should provide useful resources for formal and informal logic theory, for natural-language research in linguistics, and for artificial-intelligence research.

But more importantly, the lessons learned from the process of manually extracting the logical structure and generating the logic models contained in the V/IP Corpus are valuable for designing software that can automate at least some sub-tasks in the extraction process. The Default-Logic Paradigm employed to generate the Corpus models provides a framework for defining what those sub-tasks are. Moreover, that framework, together with the data developed in the Corpus, allows "top–down" solutions to many automation problems, which can supplement a "bottom-up" approach based on linguistic cues. Information about inferential context and possible logic structures is often as important as linguistic information about the immediate text being represented. The effort to develop protocols for human modelers, based on experience with a wide variety of modelers and decisions, provides valuable insights into the challenges to, and possible solutions for, such automation. Until we study carefully how human experts perform the task of extraction, we are unlikely to gain sufficient insight into methods of accurate automation.

Finally, the V/IP Corpus, as well as the process being used to generate it, represents an important fusion of legal research, legal education, and legal practice. This article has surveyed many of the research insights already gained, and has pointed toward many gains anticipated for the future. But such research must be supported by society, and that support depends upon the value that the research can provide. The Vaccine/Injury Project and its V/IP Corpus demonstrate how that value need not be merely long-term. The process of generating the Corpus is itself an important educational component for law students and legal practitioners, because it systematically trains them in important logic skills, applied in the context of evidence assessment and proof. Moreover, every case model added to the Corpus, as well as any patterns and schemas abstracted from the models, helps make legal practice in the socially important area of vaccine-injury compensation more effective and more efficient. The Corpus, with its associated research and education, helps extend the rule of law by making legal decision making more transparent, more predictable, and more accessible to the affected population.

# References

Aman AC Jr, Mayton WT (2001) Administrative law, 2nd edn. West Group, St. Paul
Anderson T, Twining W (1991) Analysis of evidence: how to do things with facts based on Wigmore's Science of judicial proof. Northwestern University Press, Evanston

Ashley KD (2009). Ontological requirements for analogical, teleological, and hypothetical legal reasoning. In: Proceedings of 12th international conference on artificial intelligence and law (ICAIL-09). ACM, New York, pp 1–10

Ashley KD, Brüninghaus S (2009) Automatically classifying texts and predicting outcomes. Artif Intell Law 17:125–165

Ashley KD, Rissland EL (2003) Law, learning and representation. Artif Intell 150:17–58

Bex FJ, van Koppen PJ, Prakken H (2010) A hybrid theory of arguments, stories and criminal evidence. Artif Intell Law 18:123–152

Biagioli C, Francesconi E, Passerini A, Montemagni S, Soria C (2005). Automatic semantics extraction in law documents. In: Proceedings of tenth international conference on artificial intelligence and law (ICAIL-05). ACM, New York, pp 133–140

Brachman RJ, Levesque HJ (2004) Knowledge representation and reasoning. Elsevier, Amsterdam

Branting LK (2000) Reasoning with rules and precedents: a computational model of legal analysis. Kluwer, Dordrecht

Brewer S (1996) Exemplary reasoning: semantics, pragmatics, and the rational force of legal argument by analogy. Harv Law Rev 109:923

Brewka G, Dix J, Konolige K (1997) Nonmonotonic reasoning: an overview. CSLI Publications, Stanford

Brüninghaus S, Ashley KD (2005). Generating legal arguments and predictions from case texts. In: Proceedings of 10th international conference on artificial intelligence and law (ICAIL-05). ACM, New York, pp 65–74

Carmines EG, Zeller RA (1979) Reliability and validity assessment. Sage, Newbury Park

Chorley A, Bench-Capon T (2005) An empirical investigation of reasoning with legal cases through theory construction and application. Artif Intell Law 13:323–371

Gordon TF, Walton D (2009). Legal reasoning with argumentation schemes. In: Proceedings of 12th international conference on artificial intelligence and law (ICAIL-09). ACM, New York, pp 137–146

Kadane JB, Schum DA (1996) A probabilistic analysis of the Sacco and Vanzetti evidence. Wiley, New York

Kyburg HE Jr, Teng CM (2001) Uncertain inference. Cambridge University Press, Cambridge

Mochales R, Moens M-F (2008). Study on the structure of argumentation in case law. In: Legal knowledge and information systems (Jurix 2008). IOS Press, pp 11–20

Moens M-F, Boiy E, Palau R, Reed C (2007). Automatic detection of arguments in legal texts. In: Proceedings of 11th international conference on artificial intelligence and law (ICAIL-07). ACM, New York, pp 225–230

Palau RM, Moens M-F (2009) Argumentation mining: the detection, classification and structure of arguments in text. In: Proceedings of 12th international conference on artificial intelligence and law (ICAIL-09). ACM, New York, pp 98–107

Pollock JL (1990) Nomic probability and the foundations of induction. Oxford University Press, New York

Prakken H (2004) Analysing reasoning about evidence with formal models of argumentation. Law Probab Risk 3(1):33–50

Prakken H (2008) Formalising ordinary legal disputes: a case study. Artif Intell Law 16:333–359

Prakken H, Sartor G (2004) The three faces of defeasibility in the law. Ratio Juris 17(1):118–139

Prakken H, Reed C, Walton D (2003) Argumentation schemes and generalisations in reasoning about evidence. In: Proceedings of 9th international conference of artificial intelligence and law (ICAIL-03). ACM, New York, pp 32–41

Rissland EL, Ashley KD, Loui RP (2003) AI and law: a fruitful synergy. Artif Intell 150:1–15

Saravanan M, Ravindran B (2010) Identification of rhetorical roles for segmentation and summarization of a legal judgment. Artif Intell Law 18:45–76

Schum DA (1994) Evidential foundations of probabilistic reasoning. Wiley, New York

Toulmin S, Rieke R, Janik A (1984) An introduction to reasoning. Macmillan, New York

Verheij B (2005) Virtual arguments: on the design of argument assistants for lawyers and other arguers. TMC Asser Press, The Hague

Walker VR (2003) Epistemic and non-epistemic aspects of the factfinding process in law. Am Phil Assoc Newsl Law Phil 3(1):132–136

Walker VR (2004) Restoring the individual plaintiff to tort law by rejecting "Junk Logic" about specific causation. Ala Law Rev 56(2):381–481

Walker VR (2007a) A default-logic paradigm for legal fact-finding. Jurimetrics 47:193–243

Walker VR (2007b) Visualizing the dynamics around the rule-evidence interface in legal reasoning. Law Probab Risk 6(1–4):5–22

Walker VR (2009a) Emergent reasoning structures in law. In: Trajkovski G, Collins SG (eds) Handbook of research on agent-based societies: social and cultural interactions. Information Science Reference, Hershey, pp 305–324

Walker VR (2009b) Designing factfinding for cross-border healthcare. Opinio Juris in Comparatione 3(1):1–40

Walton DN (1996) Argument schemes for presumptive reasoning. Lawrence Erlbaum, Mahwah

Walton D, Reed C, Macagno F (2008) Argumentation schemes. Cambridge University Press, Cambridge

Wyner A, Peters W (2010) Lexical semantics and expert legal knowledge towards the identification of legal case factors. In: Legal knowledge and information systems (Jurix 2010). IOS Press