

# Annotating Patterns of Reasoning about Medical Theories of Causation in Vaccine Cases: Toward a Type System for Arguments

**Vern R. Walker**  
Director,  
Research Laboratory for  
Law, Logic & Technology  
Maurice A. Deane School of  
Law at Hofstra University  
Hempstead, New York, USA  
Vern.R.Walker@  
Hofstra.edu

**Karina Vazirova**  
Protocols Coordinator,  
Research Laboratory for  
Law, Logic & Technology  
Maurice A. Deane School of  
Law at Hofstra University  
Hempstead, New York, USA

**Cass Sanford**  
Researcher,  
Research Laboratory for  
Law, Logic & Technology  
Maurice A. Deane School of  
Law at Hofstra University  
Hempstead, New York, USA

## Abstract

Automated argumentation mining requires an adequate type system or annotation scheme for classifying the patterns of argument that succeed or fail in a corpus of legal documents. Moreover, there must be a reliable and accurate method for classifying the arguments found in natural language legal documents. Without an adequate and operational type system, we are unlikely to reach consensus on argument corpora that can function as a gold standard. This paper reports the preliminary results of research to annotate a sample of representative judicial decisions for the reasoning of the factfinder. The decisions report whether the evidence adduced by the petitioner adequately supports the claim that a medical theory causally links some type of vaccine with various types of injuries or adverse medical conditions. This paper summarizes and discusses some patterns of reasoning that we are finding, using examples from the corpus. The pattern types and examples presented here demonstrate the difficulty of developing a type or annotation system for characterizing the logically important patterns of reasoning.

## 1 Introduction

This paper reports the preliminary results of research by the Research Laboratory for Law, Logic & Technology (LLT Lab) on a corpus of judicial decisions that we have annotated for patterns of argumentation. We first describe the sample of judicial decisions, and report the frequency of argument types using a coarse typology based on logical connectives. We then discuss three additional approaches to a finer-grained typology, based on types of inference, types of evidence, and types of evidentiary discrepancies. We conclude by discussing our working hypotheses for developing a type system for arguments, as well as discussing prior related work.

## 2 The Sample of Vaccine-Injury Compensation Decisions

The research in this paper is based on a sample of 10 judicial decisions in the United States, in which the petitioner was seeking compensation, under the National Vaccine Injury Compensation Program (NVICP), for injuries allegedly caused by a covered vaccine (Ashley and Walker 2013; Walker 2009; Walker et al. 2011, 2013). The sample is part of the Vaccine/Injury Project Corpus (V/IP Corpus), which comprises every decision filed during a 2-year period (a total of 35 decision texts, typically 15 to 40 pages each) that applied a 3-prong test of causation, enunciated by a federal court in *Althen* (2005). These decisions are authored by special masters attached to the Court of Federal Claims, who function as factfinders in contested cases. According to the

*Althen* test, in order to prevail the petitioner must establish three propositions, each by a preponderance of the evidence: (1) that a “medical theory causally connects” the type of vaccine with the type of injury; (2) that there was a “logical sequence of cause and effect” between the particular vaccination and the particular injury; and (3) that a “proximate temporal relationship” existed between the vaccination and the injury. Proving these causation conditions generally requires integrating expert, scientific evidence with non-expert evidence, and reconciling scientific standards of proof with non-scientific (legal or common-sense) standards of proof. In 5 of the decisions, the petitioner succeeded in proving all three *Althen* conditions and ultimately won the case (*Cusati, Roper, Casey, Werderitsh, and Stewart*), while in the remaining 5 cases the petitioner lost on the *Althen* first condition and the government won the case (*Meyers, Sawyer, Wolfe, Thomas, and Walton*).

This paper examines patterns of argument and reasoning found in the factfinding portions of these vaccine-compensation cases. And we have examined primarily the patterns of reasoning provided by the factfinder in support of the findings of fact. There are several reasons for taking this approach. First, focusing on the reported findings and reasoning of the factfinder also provides information about which arguments were successful (persuasive) and which were not. Second, in writing a decision, the factfinder is more likely to report her own reasoning with more care and detail than she might use in relating the argument of a witness or party. Third, in probably many situations, the reasoning reported by a factfinder was in fact also an argument made originally by a party, which was then adopted by the factfinder. Fourth, in most decisions it is easier to identify and count *all* of the reported findings and reasons of the factfinder, because they are often gathered together in a “Discussion” section of the decision; by contrast, counting the total number of arguments of the parties is more difficult.

Moreover, we have limited ourselves in this paper primarily to patterns that address the first of these *Althen* conditions: that is, whether there was at the time of the litigation a medical theory that causally linked the type of vaccine involved with the type of injury alleged. We have selected this issue for this paper because it involves an issue and style of proof that is general in nature, and less dependent upon the plausibility of particular facts that are peculiar to the specific case.

Indeed, proving that “the vaccine can cause this type of injury, at least sometimes” is likely to exhibit patterns of reasoning common in many domains, both inside and outside of law.

In general, we expect both the arguments by the parties or witnesses and the reasoning given by the factfinder for a finding of fact to exhibit the same “argument patterns.” That is, we expect the same types of patterns to occur, whether a party puts forth an argument for the factfinder to adopt, or the factfinder reports certain reasoning as being persuasive. As a matter of terminology, the term “argument” is typically applied to the argumentative reasoning of a party, whether or not it proves to be persuasive to the factfinder, and the term “reasoning” is often reserved for the supporting reasoning provided by the legal decision maker. However, from the perspective of exhibiting reasoning patterns, we consider “arguments” and “reasoning” to be equivalent – the only difference being attribution (the agent using the pattern, or to whom the pattern is attributed).

### **3 The Frequency of Arguments in the Sample Cases**

In order to provide quality assurance in identifying the structure of the factfinder’s reasoning, our methodology integrated analyses by three annotators in three steps. First, a student researcher trained in the LLT Lab’s logic modeling protocols annotated a legal decision for elements of the factfinder’s reasoning. Second, another student (who was usually more experienced than the first student) then reviewed those annotations, and the two researchers reached a consensus on any discrepancies. Third, Lab Director Walker performed an independent analysis, and he and the two student researchers discussed and documented any annotation issues, and decided on the final annotations. The resulting “logic model” of the reasoning for a single case integrates numerous units of reasoning into a single logical structure, with each unit consisting of one conclusion and one or more immediately supporting reasons (premises).

Walker et al. (2011, pp. 296-300) provide details on the default-logic framework and on the logical connectives used in the LLT Lab’s logic models to connect the supporting reasons (premises) to the conclusion. Because evidentiary propositions (both conclusions and premises) have plausibility-values based on a seven-valued scale (from “highly plausible” through “undecided” to “highly implausible”), the logical connec-

tives must operate on a many-valued scale. The four logical connectives we use in our logic models are:

- “MIN” assigns to the conclusion the *lowest* plausibility-value possessed by any of its supporting premises (MIN functions like a conjunctive AND);
- “MAX” assigns to the conclusion the *highest* plausibility-value possessed by any of its supporting premises (MAX functions like a disjunctive OR);
- “EVIDENCE FACTORS” merely lists relevant reasons or premises, but does not provide a computable formula for producing a plausibility-value for the conclusion as a function of the values of the premises; and
- “REBUT” assigns to the conclusion a degree of implausibility inverse to the degree of plausibility of the rebutting (defeating) premise, when (but only when) the rebutting premise is plausible to some degree (for example, if the rebutting premise is “highly plausible,” then the conclusion is “highly implausible”; but if the rebutting premise is only

“slightly plausible,” then the conclusion is only “slightly implausible”).

Using the kind of logical connective employed as a coarse typology, we can classify the arguments found within the factfinding. Table 1 summarizes the results of the argument frequencies within the LLT Lab’s logic models for the 10 decisions in the sample, under *Althen* Prong 1 only, by type of connective. A single argument is defined as a single conclusion supported by an immediate level of reasoning – that is, a single conclusion supported by one or more premises or reasons. Where a single conclusion rests on both *prima facie* supporting premises and a defeater, we classified that argument by the connective occurring in the *prima facie* line of reasoning. In Table 1, for example, the reasoning of the factfinder under *Althen* Prong 1 in the *Roper* decision consisted of 7 arguments containing the EVIDENCE FACTORS connective, 2 arguments connected by MIN, and 1 argument connected by REBUT. The numeral “1” in square brackets in the REBUT column of Table 1 indicates that there was a second REBUT connective in the decision, but it occurred as a defeater attached to some other *prima facie* line of reasoning.

Name of Case (Filing Date)	Prong-1 Finding	EVIDENCE FACTOR Args	MIN Args	MAX Args	REBUT Args
Cusati (9/22/05)	For petitioner	5			
Roper (12/9/05)	For petitioner	7	2		1 [1]
Casey (12/12/05)	For petitioner		4	1	
Werderitsh (5/26/06)	For petitioner	2			3 [1]
Stewart (3/19/07)	For petitioner	5			
Meyers (5/22/06)	For government	5			[1]
Sawyer (6/22/06)	For government	10	1		[1]
Wolfe (11/9/06)	For government	5			[1]
Thomas (1/23/07)	For government	3			1
Walton (4/30/07)	For government	14	1		1 [3]

**Table 1.** Frequency of Arguments in Ten-Case Sample under *Althen* Prong 1, by Type of Connective

An examination of the results in Table 1 shows one reason why we regard this classification of arguments by logical connective as providing only a high-level or coarse typology, but not an adequately informative or useful typology. By far the most common form of argument stated is simply a conclusion, supported by a list of relevant considerations (the arguments containing simply the EVIDENCE FACTORS connective). In these arguments, no other struc-

ture is expressly indicated, beyond a listing of supporting reasons. Nearly 79% of the arguments (56 out of 71) contained no internal truth-functional structure, but were merely lists of supporting information considered by the factfinder to be relevant to drawing that conclusion. This provides motivation for developing a more informative typology for arguments or reasoning patterns, beyond the connectives normally used in propositional logic.

However, we express a word of caution about maintaining descriptive accuracy in annotation. We have considered it critical to annotate patterns of reasoning in a way that accurately represents the reported reasoning of the factfinder. If there exists no semantic cue indicating a more structured form than merely a list of supporting reasons, then we believe that accurate annotation would represent the list form of the original document. It might be possible to *interpret* a list as a more structured line of reasoning, but the *data themselves* (as contrasted with the interpretation of those data) should not be contaminated with information not already expressed in the original source. Thus, we believe that it should always be possible to distinguish between annotations that are strictly faithful to the text in representing the author's stated meaning, and annotations that add the interpretations of commentators. The kind of type system we are discussing in this paper is the former kind, with which we can accurately capture the meaning of the author of the text.

#### 4 Patterns by Types of Inference

This section discusses the possible approach of identifying patterns of argumentation or reasoning that exhibit some type of inference from premises to conclusion (beyond the propositional connectives discussed in Section 3). This approach would draw upon inference methods studied in fields other than law, such as deductive logic, probability or statistics, science or medicine. We discuss some of these types of inference that we find in the vaccine cases.

##### 4.1 Deductive Reasoning

Occasionally reasoning is deductive in form – that is, if the premises are true then the conclusion must be true as well, and the sole avenue for undermining the argument is attacking the truth of the premises (Copi and Cohen 1998, p. 25). In such patterns, the supporting reasons are not merely a list, but rather a list of jointly sufficient reasons for drawing the inference as a necessary conclusion. For example, in *Casey* (p. 26), the conclusion that the varicella vaccine can negatively affect the nervous system was supported by a conjunction of two causal relations: that the vaccine can cause a direct viral infection, and that a direct viral infection can negatively affect the nervous system.

Deductive patterns of reasoning, however, would have premises connected to the conclusion by the propositional connective MIN, because

the conclusion would be true (or plausible) *whenever all* of the premises are true (plausible). At most, therefore, 8 of the 71 arguments found in the 10 sample cases would be deductive in form. We find that it is extremely rare for the factfinders in the vaccine cases to explicitly lay out reasoning in a deductively valid format.

##### 4.2 Probabilistic or Statistical Reasoning

Reasoning that is probabilistic or statistical in form could be sub-divided into many types – e.g., reasoning based on premises that are explicitly regarded as merely probable, or reasoning proceeding from a premise that most (or some percentage of) members of one class are members of another class. For example, in *Sawyer* (p. 10), the petitioner's expert relied on the generalization that "it would be reasonable for someone with [the petitioner's] condition to have some days that are less painful than others, but it should generally be a constant pain." And in *Walton* (p. 33), when the petitioner's expert argued that the MMR vaccine can cause myocarditis, the government's expert rebutted that if it were possible, then "we would have seen it by now because millions of doses of the vaccine have been given and this has not been reported."

##### 4.3 Scientific or Medical Reasoning

While deductive and probabilistic inferences do not rely upon methods developed within any particular discipline, legal factfinders are often persuaded by inference methods familiar from science or medicine. For example, in *Walton* (p. 35), an expert for the petitioner and an expert for the government agreed that "an acute reaction from a vaccine-caused myocarditis would be expected to manifest within days to two weeks of infection." Yet the petitioner's "symptoms occurred well over three weeks after her vaccination." The special master found that "[p]erhaps the most significant problem" with the petitioner's theory of causation was "the lack of temporal connection between the MMR vaccination and evidence of a cardiac illness."

Scientific, medical and other expert witnesses sometimes reach a conclusion by balancing various factors and arriving at a considered professional judgment, and this process itself might be persuasive to the legal factfinder. For example, the opinions of two medical experts that the special master found credible in *Stewart* (pp. 36, 38) were supported by lists of reasons. An expert's scientific argument may be presented in a way that lends itself to a legal and logical structure,

providing an expert judgment in weighing the same evidence that the Special Master can rely on in reaching a conclusion.

Sometimes the appeal to scientific reasoning is mediated by evidence that scientists themselves have already reached a conclusion on the issue. In *Meyers* (p. 10), the special master noted that the “scientific community has rejected [the expert’s] theories as detailed in his articles because the human studies that have been conducted do not support his conclusions and his analytical methods do not comport with the *Daubert* requirement of reliability.” *Daubert* was a decision by the U.S. Supreme Court discussing factors relevant to assessing the evidentiary reliability of a scientific expert opinion in court.

## 5 Patterns by Types of Evidence

This section discusses another approach to classifying patterns of argument and reasoning, one that is based upon the type of evidence in one or more of the premises. The form of reasoning or inference that relies on that evidence to arrive at a conclusion can be varied (as discussed in Section 4).

This section illustrates some of the types of evidence we find recurring in vaccine cases.

### 5.1 Legal Precedent as Basis

One of the most common patterns of legal reasoning involves the citation of prior legal decisions as precedents. Precedent-based reasoning occurs when judges or factfinders utilize prior cases as providing a binding rule or applicable principle, or as providing guidance by analogy to explain or justify an outcome in the undecided matter before them (Cross et al., 2010, pp. 490-512; Levi 1949, pp. 8-27).

In vaccine decisions, the special masters have utilized precedent-based reasoning in various ways. In *Wolfe* (pp. 9-11), for example, the petitioner’s expert based his theory of causation solely on the temporal relationship between the vaccination and the onset of the injury, together with a lack of alternative theories of causation. The special master found that argument to fall short of the established burden, citing the Federal Circuit’s decision in *Grant* (p. 1148) for the proposition that mere temporal relationship and lack of alternative causes is not enough to create a *prima facie* case. In *Werderitsh* (p. 44), the government challenged the petitioner’s *prima facie* case by pointing to “the failure of valid epidemiologic studies to show a relationship” and

“the absence of the knowledge of the appropriate biologic mechanisms responsible.” The special master countered that “[l]egally, the absence of epidemiologic support for linking hepatitis B vaccine and MS, and the lack of identification of the specific biologic mechanism at work if hepatitis B vaccine causes MS do not prevent petitioner from satisfying her burden of proof,” citing the Federal Circuit’s decision in *Knudsen*.

### 5.2 Legal Policy as Basis

Sometimes an important basis for the reasoning is not merely a precedent, but an authoritatively established legal policy that is considered operative. For example, in *Casey* (p. 26), the special master decided that the petitioner “provided sufficient proof of a medical theory of causation,” and explained in part that “[i]t is precisely because individuals experience adverse reactions to safe vaccines on rare occasions that Congress created the Vaccine Program.” It is possible, for example, that a weaker statistical inference (Section 4.2) might combine with a policy objective to produce a persuasive argument.

### 5.3 Medical or Scientific Studies as Basis

In vaccine decisions, it is often the case that arguments are based upon medical or scientific studies, either published in medical or scientific journals or reported in medical treatises. For example, in *Stewart* (pp. 38-39), the special master relied in part on medical literature reporting a connection between the hepatitis A virus and cerebellar ataxia, in finding for the petitioner in a case involving hepatitis A vaccine and the same adverse medical condition. On the other hand, in *Meyers* (pp. 12-14), the special master refused to credit an expert opinion that was based on articles and reports that failed to address the relevant vaccine or injuries in the case. And in *Werderitsh* (p. 43), the special master was not persuaded by an article whose authors admitted that their study was small and its statistical power was reduced.

### 5.4 Case Reports as Basis

A case report is a descriptive study of a single patient’s experience (Cetrulo 2013). Such anecdotal evidence is extremely weak evidence of causation, due to the lack of a control group for testing comparisons (Kaye et al., 2014). However, despite their obvious statistical shortcomings, case reports have been utilized by special masters in their reasoning. In *Roper* (pp. 5-9), the special master explicitly addressed the inability of case reports to provide “scientific certainty” by noting

that the petitioner’s burden is subject to a much lower “more probable than not” standard, and that accumulated circumstantial evidence of probability can become sufficient to prove causation. In *Stewart* (p. 36), the special master found the case report filed by the government to be relevant for showing a plausible medical theory. And in *Werderitsh* (p. 46), the special master was convinced of causation in large part by the expert’s analogy with another vaccine case, and by the similarity of the timing pattern of the relevant symptoms in the two cases.

### 5.5 Fact Testimony as Basis

We mention here reliance upon the fact testimony of a lay witness as an evidentiary basis, although this is less common when the issue is the existence of a medical theory.

## 6 Patterns Based on Evidentiary Discrepancies

This section discusses a third approach to classifying patterns of argumentation and reasoning, one that is based on the insight that in law, perhaps more than in other domains involving fact-finding, cases are often decided primarily by resolving inconsistencies or discrepancies in the evidence. When the parties or their expert witnesses agree on a proposition, then the legal factfinder generally accepts that proposition as uncontested fact, for purposes of the litigation. For example, in *Cusati* (pp. 11-14), the opposing parties’ experts agreed on the fact that the MMR vaccine causes fever, and, in turn, that fever causes seizures. The special master utilized that consensus as part of the basis for concluding that the petitioner had met the *prima facie* standard with respect to causation. Similarly, in *Casey* (p. 26), all experts agreed that the temporal sequence of petitioner’s symptoms was appropriate. The special master deemed that consensus entirely determinative as to the third condition of *Althen* and did not even engage in further temporal analysis.

### 6.1 Credibility of Source: Expert vs. Expert

When opposing expert witnesses disagree on some proposition or issue within their expertise, then one approach of the decision maker is to weigh the credibility of the experts, either individually or in comparison. For example, in *Sawyer* (pp. 16-20), the special master found the expert to be so unreliable that this ultimately became the main basis for the decision: “This is a

case of unreliable expert testimony.” The special master devoted 4½ pages of his single-spaced decision to detailing his supporting reasoning. In *Stewart* (pp. 41-42), the special master looked to the lack of credibility of the government’s expert in reaching a finding for the petitioner. The expert was found to be “less than candid or credible” due to an insistence on an assumption that was directly contradicted by medical records, and his failure to take videotape evidence sufficiently into account. And in *Walton* (p. 35), the special master discussed why she “found Dr. Charash [the petitioner’s expert] to be far less persuasive than Dr. Glezen or Dr. Brinker [experts for the government].”

### 6.2 Credibility of Source: Inadequate Explanation

Sometimes the credibility of a source is predicated upon that source’s taking irrelevant factors into account, or failing to take relevant factors into account. In *Walton* (p. 35), for example, the petitioner’s expert “relied upon assertions of fact not supported by contemporaneous medical records, failed to address the significance of negative cardiac testing, relied upon a temporal relationship between vaccination and onset of symptoms not established by the evidence, failed to demonstrate any support for his theories in research, and failed to address the contrary research evidence submitted by respondent’s experts.”

## 7 Discussion: Developing a Type System for Arguments

Our objective is to develop a type system for argument patterns using high-level categories such as those illustrated in Sections 4-6, and developing sub-types based on lower-level categories. The features of sub-categories that are important are those that help to identify arguments that were successful or unsuccessful in the vaccine cases. For example, under the category of medical or scientific studies as basis (see Section 5.3 above), a sub-category might be studies that report negative results (no evidence of a statistically significant causal relationship), and important features might include sample size and statistical power (see *Werderitsh*, p. 43). In developing a type system, we have formulated several working hypotheses from the approaches and examples discussed above.

First, it might be difficult to construct a well-defined taxonomy for types of argument patterns

in a specific legal domain. That is, it might not be feasible to devise a classification system under which every argument instance would fit in one and only one category, even in principle. An argument instance might well fit under multiple categories. Moreover, this theoretical point is quite separate from the difficulty of devising a methodology for reliably and accurately placing each argument instance into the correct category.

As a result, it might be more realistic to develop a list of significant categories and features of arguments, and to score a profile for any particular argument instance on those categories and features. For example, an argument instance might involve a combination of attacking the credibility of a source through inadequate explanation (Section 6.2), by relying on a scientific study (Section 5.3) while employing probabilistic reasoning (Section 4.2). Then, instead of an identity relation (two arguments being “the same type”), a fuzzier relation of similarity might be useful, computed as a function of the profile scores of two argument instances.

Second, we cannot exclude the possibility that the optimal list of argument categories and features might be in part a function of the task to be performed, and some a priori list might not be optimal for all uses. This is an empirical question that remains to be answered.

Third, if it proves to be the case that the optimal list of argument categories and features is both a function of the task to be performed and widely variant from domain area to domain area, then the most promising approach might be machine learning on training texts that have been only lightly parsed in standard ways. That is, a coarser-grained semantic markup (e.g., sentences or clauses tagged merely as “evidence” and “finding,” or “premise” and “conclusion”) might be less costly to achieve, might have higher inter-annotator reliability, and might be entirely adequate for machine learning. In order to test this hypothesis, however, we still need to develop an adequate annotation scheme and a gold standard corpus of argument patterns.

Finally, once an adequate annotation scheme is developed, we will face the challenge of outcome evaluation. Argument patterns may differ in weight with respect to the overall conclusion and ultimate outcome of the case. For example, in the context of vaccine decisions, a finding that petitioner’s expert was entirely unreliable may singlehandedly support dismissal of the entire claim (see *Sawyer*, pp. 16-20). In addition, it is far from clear whether case reports alone can

support a medical theory connecting the vaccine to the injury (see *Roper*, pp. 5-9; *Stewart*, p. 36). Thus, comparative research and assignment of weights to argument patterns may be useful.

In addition to keeping these working hypotheses in mind, when deciding upon the tentative adoption of any annotation scheme or type system, we give consideration to: (1) the cost (in terms of both resources and risk of error) of manually applying the scheme to the number of legal documents needed to allow machine training and testing; (2) the feasibility of successfully automating the detection and annotation of new texts using the scheme; (3) the adequacy of the scheme as a means of performing various tasks, such as generating new arguments in new cases, or predicting ultimate case outcomes; and (4) the interoperability of the scheme with existing ontologies and datasets. Although many researchers have tested systems for automatically annotating legal texts (see references discussed in Section 8), we believe that critical work remains to be done on empirically developing an annotation scheme that is adequate for representing natural language arguments in legal texts, for the purpose of assisting in the generation of new arguments in new cases.

## 8 Prior Related Work

Our strategic approach is both empirical and logical in nature. Our approach is empirical because we consider it crucial to use a corpus of diverse and linguistically rich legal decisions to gain insights into actual argument patterns. We also take the typical approach of logic in looking for patterns of successful and unsuccessful argument at a “local” level, in the relationships among premises and conclusions (the distinction between “local” and “global” is due to Mochales and Moens, 2011, pp. 3-8). Local argument patterns (which frequently occur within a single paragraph) are distinguished from the global argument pattern that supports the ultimate decision on a claim (often the conclusion of an entire decision) (*id.*). We briefly mention here recent research bearing on our work.

Mochales and Moens (2011, pp. 5-6; 2008, pp. 12-14) annotated a corpus of 47 judicial decisions from the European Court of Human Rights using a system of argumentation schemes developed by Walton (1996; Walton et al., 2008). They focused on “The Law” sections of the judicial opinions, which discuss the arguments of the parties and the court’s reasons supporting its de-

cision. Thus, their focal point within the judicial decisions is similar to that in our study of the vaccine cases (factfinder reasoning) and their attention to the local argument patterns is similar to ours. One major difference might be that the reasoning in our judicial decisions usually rests heavily upon scientific and medical evidence, including expert opinions.

Saravanan and Ravindran (2010, pp. 47-53, 65-66) manually annotated sentences in a corpus of 200 decisions (approximately 16,000 sentences) from Indian courts using a “rhetorical” annotation scheme containing 7 categories: “identifying the case,” “establishing facts of the case,” “arguing the case,” “history of the case,” “arguments (analysis),” “ratio decidendi (ratio of the decision),” and “final decision (disposal).” It is unclear how many of these rhetorical categories will play a role in defining useful local argument patterns in U.S. cases.

Ashley and Brüninghaus (2009, pp. 132-43; Brüninghaus and Ashley 2005, pp. 65-67) investigated 146 cases involving trade secret misappropriation. “Squibs” of these cases (manually prepared textual descriptions of the case facts) contained sentences that were manually annotated with respect to being positive instances of 26 factors (a positive instance was a sentence “from which it could be reasonably inferred” that the factor “applied in the case”). A “factor” is a category of facts that helps to predict the case outcome for either the plaintiff or the defendant – for example, Factor F4 represents the fact pattern in which the defendant entered into a non-disclosure agreement with the plaintiff, and F4 favors an ultimate decision for the plaintiff. Chorley and Bench-Capon (2005) supplemented these factors with “values,” in the context of building a “theory.” Wyner and Peters (2010) selected 39 of these cases and a limited number of base factors, and developed semantically salient terms and synonyms for these factors. Even assuming that case squibs retain the linguistic richness of the original documents and contain a sufficient number of negative instances (sentences that are irrelevant to argumentation), the goal of annotating sentences for factors is to predict the ultimate outcome in domain cases, and such factors may or may not be relevant to local argument patterns within the case.

Biagioli et al. (2005), using a dataset of paragraphs selected from Italian legislative texts, classified the paragraphs into eleven types of legislative “provision” (e.g., definition, obligation, prohibition, permission). Each type of pro-

vision takes various arguments – for example, the provision type “obligation” takes as arguments the “addressee,” the “action,” and a “third-party.” Such types of provisions might well appear in local argument patterns, which apply legal rules to evidence in a particular case.

Wyner et al. (2013, p. 167) annotated intellectual property appellate cases using 32 annotations, which were selected as being “those used in practice in the analysis of cases in law schools.” Annotation types ranged from “Judge Name” to “Legal Facts” (“the legally relevant facts of the case that are used in arguing the issues”). While some of these annotation types might be relevant to local argument patterns, others probably are not.

## 9 Conclusion

We have reported the preliminary results of our efforts to develop an adequate type system or annotation scheme for marking up successful and unsuccessful patterns of argument in U.S. judicial decisions. We are working from a corpus of vaccine-injury compensation cases that report factfinding about causation, based on both scientific and non-scientific evidence and reasoning. We have summarized and illustrated the patterns of reasoning we are finding, and have discussed our strategy for future research. What seems clear is that the task of developing an adequate type or annotation system is both difficult and important. Without an adequate and operational type system, we are unlikely to reach consensus on argument corpora that can function as a gold standard, or to make robust and useful progress on automating the annotation of judicial decisions for argumentation.

## References

- Althen v. Secretary of Health and Human Services, 418 F.3d 1274 (Fed.Cir. 2005).
- Ashley Kevin D., and Stefanie Brüninghaus. 2009. Automatically classifying texts and predicting outcomes. *Artificial Intelligence and Law* 17:125-165.
- Ashley, Kevin D., and Vern R. Walker. 2013. From Information Retrieval (IR) to Argument Retrieval (AR) for Legal Cases: Report on a Baseline Study. *Proceedings of the 26<sup>th</sup> International Conference on Legal Knowledge and Information Systems (JURIX 2013)* (Kevin D. Ashley, ed.): 29-38.
- Biagioli, C., E. Francesconi, A. Passerini, S. Montemagni, and C. Soria. 2005. Automatic semantics extraction in law documents. *Proceedings of tenth*



- international conference on artificial intelligence and law* (ICAAIL-05): 133-40. ACM: New York.
- Brüninghaus, Stefanie, and Kevin D. Ashley. 2005. Generating Legal Arguments and Predictions from Case Texts. *Proceedings of tenth international conference on artificial intelligence and law* (ICAAIL-05): 65-74. ACM: New York.
- Casey v. Secretary of Health and Human Services, No. 97-612V (Office of Special Masters, United States Court of Federal Claims, December 12, 2005).
- Cetrulo, Lawrence G. 2013. Case Reports. *Toxic Torts Litigation Guide* 1: §5:33. Westlaw, Thomson Reuters.
- Chorley, Alison, and Trevor Bench-Capon. 2005. An empirical investigation of reasoning with legal cases through theory construction and application. *Artificial Intelligence and Law* 13:323-371.
- Copi, Irving M., and Carl Cohen. 1998. *Introduction to Logic* (Tenth Edition). Prentice Hall, Upper Saddle River, New Jersey.
- Cross, Frank B., James F. Spriggs II, Timothy R. Johnson, and Paul J. Wahlbeck. 2010. Citations in the U.S. Supreme Court: An Empirical Study of Their Use and Significance. *University of Illinois Law Review* 2010: 489-575.
- Cusati v. Secretary of Health and Human Services, No. 99-0492V (Office of Special Masters, United States Court of Federal Claims, September 22, 2005).
- Daubert v. Merrell Dow Pharmaceuticals, Inc., 509 U.S. 579 (1993).
- Grant v. Secretary of the Department of Health and Human Services, 956 F.2d 1144 (Fed. Cir. 1992).
- Kaye, David H., David E. Bernstein, and Jennifer L. Mnookin. 2014. Types of Causal Studies. *The New Wigmore: A Treatise on Evidence: Expert Evidence* §12.5.1. CCH Incorporated, Aspen Publishers.
- Knudsen v. Secretary of the Department of Health and Human Services, 35 F.3d 543 (Fed. Cir. 1994).
- Levi, Edward H. 1949. *An Introduction to Legal Reasoning*. The University of Chicago Press, Chicago, Illinois.
- Meyers v. Secretary of the Department of Health and Human Services, No. 04-1771V (Office of Special Masters, United States Court of Federal Claims, May 22, 2006).
- Mochales, Raquel, and Marie-Francine Moens. 2008. Study on the Structure of Argumentation in Case Law. *Legal Knowledge and Information Systems* (Jurix 2008) (E. Francesconi, G. Sartor, D. Tiscornia, eds.): 11-20. IOS Press.
- Mochales, Raquel, and Marie-Francine Moens. 2011. Argumentation mining. *Artificial Intelligence and Law* 19:1-22.
- Roper v. Secretary of Health and Human Services, No. 00-407V (Office of Special Masters, United States Court of Federal Claims, December 9, 2005).
- Saravanan, M., and B. Ravindran. 2010. Identification of Rhetorical Roles for Segmentation and Summarization of a Legal Judgment. *Artificial Intelligence and Law* 18: 45-76.
- Sawyer v. Secretary of the Department of Health and Human Services, No. 03-2524V (Office of Special Masters, United States Court of Federal Claims, June 22, 2006).
- Stewart v. Secretary of the Department of Health and Human Services, No. 06-287V (Office of Special Masters, United States Court of Federal Claims, March 19, 2007).
- Thomas v. Secretary of the Department of Health and Human Services, No. 01-645V (Office of Special Masters, United States Court of Federal Claims, January 23, 2007).
- Walker, Vern R. 1996. Preponderance, Probability and Warranted Factfinding. *Brooklyn Law Review* 62: 1075-1136.
- Walker, Vern R. 2003. Epistemic and Non-epistemic Aspects of the Factfinding Process in Law. *American Philosophical Association Newsletter on Law and Philosophy* 3(1): 132-136.
- Walker, Vern R. 2007. A default-logic paradigm for legal fact-finding. *Jurimetrics* 47: 193-243.
- Walker, Vern R. 2009. Designing factfinding for cross-border healthcare. *Opinio Juris in Comparatione* 3, paper n. 1, 1-40.
- Walker, Vern R., Nathaniel Carie, Courtney C. DeWitt, and Eric Lesh. 2011. A framework for the extraction and modeling of fact-finding reasoning from legal decisions: lessons from the Vaccine/Injury Project Corpus. *Artificial Intelligence and Law* 19:291-331.
- Walker, Vern R., Chan Hee Park, Philip H. Hwang, Arthur John, Evgeny I. Krasnov and Keith Langlais. 2013. A process approach to inferences of causation: empirical research from vaccine cases in the USA. *Law, Probability and Risk* 12(3-4): 189-205.
- Walton v. Secretary of the Department of Health and Human Services, No. 04-503V (Office of Special Masters, United States Court of Federal Claims, April 30, 2007).

- Walton, Douglas N. 1996. *Argumentation Schemes for Presumptive Reasoning*. Lawrence Erlbaum Associates, Publishers, Mahwah, New Jersey.
- Walton, Douglas, Chris Reed, and Fabrizio Macagno. 2008. *Argumentation Schemes*. Cambridge University Press, Cambridge, UK.
- Werderitsh v. Secretary of the Department of Health and Human Services, No. 99-319V (Office of Special Masters, United States Court of Federal Claims, May 26, 2006).
- Wolfe v. Secretary of Health and Human Services, No. 05-0878V (Office of Special Masters, United States Court of Federal Claims, November 9, 2006).
- Wyner, Adam, and Wim Peters. 2010. Lexical Semantics and Expert Legal Knowledge towards the Identification of Legal Case Factors. *Legal Knowledge and Information Systems (Jurix 2010)* (Radboud G.F. Winkels, ed.): 127-36. IOS Press.
- Wyner, Adam, Wim Peters, and Daniel Katz. 2013. A Case Study on Legal Case Annotation. *Proceedings of the 26<sup>th</sup> International Conference on Legal Knowledge and Information Systems (JURIX 2013)* (Kevin D. Ashley, ed.): 165-74.