Burdens and Standards of Proof for Inference to the Best Explanation: Three Case Studies

Floris Bex[1] and Douglas Walton[2]

Abstract.
In this paper, we provide a formal logical model of evidential reasoning with proof standards and burdens of proof that enables us to evaluate evidential reasoning by comparing stories on either side of a case. It is based on a hybrid inference model that combines argumentation and explanation, using inference to the best explanation as the central form of argument. The model applied to one civil case and two criminal cases. It is shown to have some striking implications for modeling and using traditional proof standards like preponderance of the evidence and beyond reasonable doubt.

Keywords.

 Inference to the best explanation, beyond reasonable doubt, evidence supporting a story, argumentation, shifting of burden of proof, critically questioning a story, plausible reasoning.

1. Introduction

In legal trials, the burden of proof and its associated standards of proof determine how strong a party's position needs to be in order to prevail. In AI and Law, various ways of logically modeling the legal burden of proof and proof standards have been proposed (Gordon and Walton, 2009; Prakken and Sartor, 2009). These approaches both presuppose some type of frameworks for defeasible argumentation, in which arguments (or argument graphs) are constructed by performing consecutive reasoning steps from the evidence to the facts in issue. When talking about the *facts* of a criminal case, however, Bex and colleagues (Bex 2011, Bex et al. 2010) have argued that argumentative approaches such as the ones mentioned above need to be expanded to include reasoning with stories or explanations, that is, alternative accounts about what (might have) happened in the case. In a purely argument-based approach the conclusions of arguments are individual facts in issue. In a real case, however, these facts will be related to each other in various ways (e.g. causally, temporally, motivationally) and these relations may also be the subject of argumentative reasoning. Pennington and Hastie (1993) have shown that using explanatory stories is closest to how legal decision makers actually think about a case. Pardo and Allen (2007) have persuasively argued that reasoning in trials involves evaluating the relative plausibility of the various explanations of the evidence. We build a formal and computational argumentation model that is used to implement Pardo and Allen's contention that inference to the best explanation can be applied to civil as well as criminal trials in which the parties provide competing "theories" about the facts of the case.[3]

---

[1] Argumentation Research Group, School of Computing, University of Dundee, UK.
[2] Centre for Research in Reasoning, Argumentation and Rhetoric, University of Windsor, Canada.
[3] Pardo and Allen (2007) cite a number of civil cases that involve competing explanations of the events. Examples are Anderson v. Griffin, 397 F.3d 515 [7th Cir. 2005], which is discussed in section 3.1, and Los Angeles v. Alameda Books, Inc., 535 U.S. 425, 437–38 [2002]).

Bex and colleagues proposed a hybrid theory of inference to the best explanation (IBE) that consists of a combination of abductive causal reasoning with explanations and defeasible evidential argumentation. In this hybrid theory, hypothetical explanations are constructed through abductive reasoning and these explanations can then be supported and attacked using arguments based on evidence. As the basis for weighing the strength or weakness of evidence, the alternative explanations in a case are compared. In this comparison, burdens and standards of proof play an important role (Bex and Walton 2010). For example, even if the prosecution's explanation of guilt in a criminal case is the "best" explanation (according to some rational standard) it may still not meet the legal "beyond a reasonable doubt" standard, resulting in the acquittal of the defendant. Similarly, when the parties' alternative explanations are equally good, the burden of persuasion influences which of these alternatives should be chosen.

At the moment, the hybrid theory does not include a notion of burden of proof nor does it say how various proof standards may be met (apart from Bex and Walton 2010, of which the current paper is an extension). Pardo and Allen (2007) discuss how the burden of proof influences the process of IBE in legal trials and they also provide ideas on how standards of proof may be met by an explanation. However, their theory of IBE is informal and not specified in the level of exact detail that is needed for a formal computational model. Building on our previous work (Bex and Walton 2010), in this paper we logically model reasoning with the burden of proof in inference to the best explanation, using Bex and colleagues' (2011, 2010) hybrid theory of IBE as our start point. The aim is to explore how the different types of burdens of proof (e.g. burden of persuasion, burden of production) may be modeled. We also show in a distinctive way how proof standards should be modeled in the hybrid theory. We provide extensive examples of both criminal cases and a civil case.

The rest of this paper is organized as follows. Section 2 contains a summary of the hybrid theory. Section 3 briefly discusses different types of legal burden of proof and proof standards. In section 4, we model these burdens of proof and proof standards in the hybrid theory. In the remaining sections we provide three examples of how to analyze and evaluate evidential reasoning in a civil case and two criminal cases. The third case in particular raises some interesting problems and has some surprising implications concerning the beyond reasonable doubt standard.

2. A Hybrid Theory of Inference to the Best Explanation

As its basic structure, our formal account of evidential reasoning will represent evidence as something presented to a trier of fact as data, often perceptually or by testimony, and then explained one way or another, leading to a conclusion of some sort. According to Josephson and Josephson (1994, 14), an abductive inference has the following general form.

D is a collection of data.
H explains *D*.
No other hypothesis can explain *D* as well as *H* does.
Therefore *H* is probably true.

Here, the hypothesis *H* may be a single proposition but it can also be an ordered sequence of events, a story. Various critical questions can then be asked (see below). This suggest that an explanatory story in IBE is defeasible, that is, it holds tentatively as an instance of plausible reasoning by placing a burden on an opponent to critically question it. Josephson (2001) has

offered a formal model of abductive inference and an indication as to how the standard of beyond a reasonable doubt may be modeled, but does not digress any further on the subject. Part of the problem with analyzing the notion of an explanation in a precise logical model suitable to support its implementation in fields like artificial intelligence was that, until Bex et al.'s hybrid theory (Bex 2011, Bex et al. 2010), structures such as argument diagrams were lacking a basic notion of an ordered sequence of events and actions. By showing how to build such a structured notion of a connected story that makes sense as a script underlying an explanation, our theory moves IBE ahead.

Bex and colleagues (Bex 2011, Bex et al. 2010) propose a formal hybrid theory for IBE, combining arguments and stories. In this hybrid theory, the data that is to be explained, the *explananda*, are causally explained by different stories, alternative accounts of what happened in the case. These stories can then be reasoned about using evidential arguments. The formal hybrid theory $HT = (ET, CT)$ is thus a combination of a causal-abductive theory CT and an evidential argumentation theory ET. The logic L of this theory is a combination of the inference rules of classical logic and a modus ponens inference rule for the connective $\Rightarrow$ (defeasible implication). Object-level rules in CT and ET are formalized using this connective: r_i: $p_1 \wedge ... \wedge p_n \Rightarrow q$. Here r_i is the name of the rule, $p_1,...,p_n$ and q are literals. The type of rule is indicated with a subscript: $\Rightarrow_E$ denotes an evidential rule used in arguments (e.g. $p \Rightarrow_E q$, which stands for 'p is evidence for q') and $\Rightarrow_C$ denotes a causal rule used in explanations (e.g. $p \Rightarrow_C q$, which stands for 'p causes q').

In the abductive theory $CT = (H, T, F)$, T is a set of *causal rules*, H is the set of hypothetical events or *hypotheticals* and F is the set of explananda. As in traditional models of abductive model-based reasoning (Console and Torasso 1991), the *explains* relation between a story and the explananda can be defined through a notion of logical consequence: the explananda should follow from a combination of hypothesized events and causal rules (expressing causal relations between events). Thus, given F we need to abductively infer some hypotheticals and some causal rules such that they explain F, that is, the explananda follow from the hypotheticals and the causal theory. More precisely, a story $S = H_i \cup T_i$, where $H_i \subseteq H$ and $T_i \subseteq T$, is an explanation for a set of explananda F iff for each $f \in F$: $S \mid\sim f$ (where $\mid\sim$ denotes defeasible entailment). Furthermore, we require that S is consistent and that S is minimal w.r.t. set-inclusion.

In the *argumentation theory $ET = (R, K)$*, R is a set of evidential rules and $K = K_E \cup K_A$ is a knowledge base, where K_E is a consistent set *evidence* and K_A is a set of *commonsense assumptions*. The logic for ET is similar to the ASPIC+ framework (Prakken 2010), which integrates ideas on rule-based argumentation and structured arguments (e.g. Prakken and Sartor 1997) within Dung's (1995) abstract approach. Evidential arguments can be built by taking evidence or assumptions from K and rules from R as premises and chaining applications of defeasible modus ponens into tree-structured arguments, where each node in the tree is thus an element of K, a rule from R or the result of an application of the defeasible modus ponens to one or more other nodes.

An argument AR_1 can defeat another argument AR_2 in various ways. AR_1 and AR_2 *rebut* each other if they have an opposite (intermediate) conclusion. AR_1 *undercuts* AR_2 if there is a conclusion $\neg r_i$ in AR_1 and an application of defeasible modus ponens to r_i in AR_2. Finally, AR_1 *undermines* AR_2 if AR_1 has a conclusion that is the opposite of some assumption (from K_A) in AR_2. For a collection of arguments and their binary defeat relations, the dialectical status of the arguments can be determined: informally, arguments can be either *justified*, which means that they are not defeated by other justified arguments, *overruled*, which means that they are defeated

by other justified arguments, or *defensible*, which means that they are neither justified nor overruled.[4]

Stories and arguments can now be combined in various ways: arguments based on evidence can be used to support or attack stories or attack stories and arguments can also attack each other. Figure 1 visualizes the various ways in which stories and arguments interact as diagrams akin to the ones used in argumentation (cf. Freeman 1991, Reed and Rowe 2004). Here, ⇢ arrows denote causal links, → arrows denote evidential (argumentative) links and the —● arrow denotes an attack of an argument (i.e. contradiction). White boxes are part of the abductive theory *CT* and gray boxes are part of the argumentation theory *ET*.
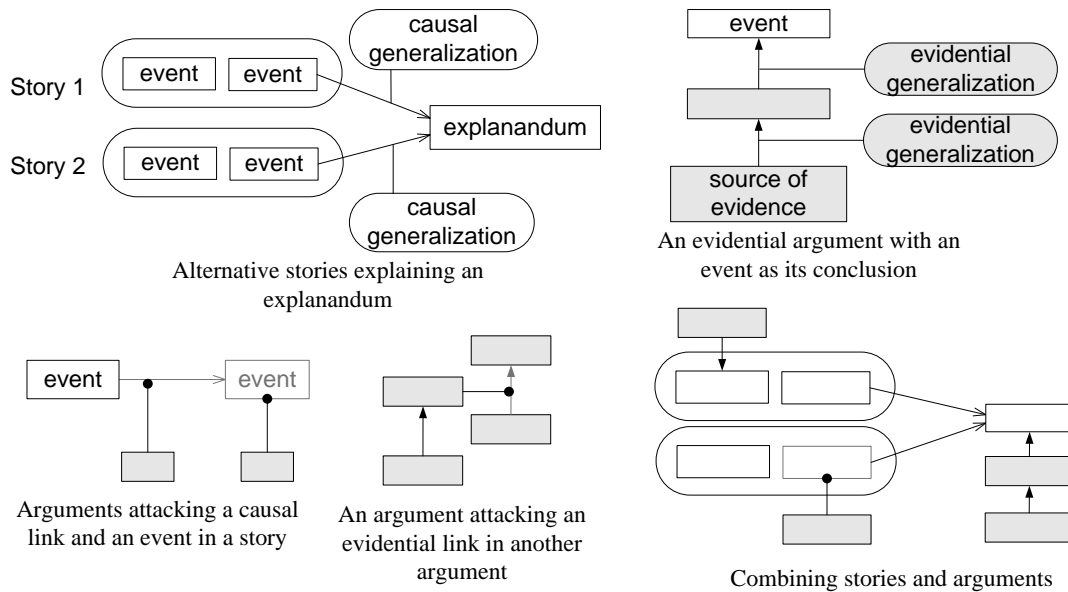

Figure 1: Visualizing the hybrid theory

In IBE, multiple explanations should be generated and compared according to criteria that express the degree to which they conform to the evidence and their plausibility. These criteria are defined using the argumentation theory. Arguments based on evidence can be used to show that an explanation is consistent or inconsistent with the evidence. More specifically, *evidential support* for an explanation S (denoted as $es(S)$) is the number of sources of evidence from K_E that support S through an argument A (i.e. $e \in K_E$ is a premise of A and $s \in S$ is a conclusion of A) and *evidential contradiction* for an explanation S (denoted as $ec(S)$) is the number of sources of evidence from K_E that contradict S through an argument A (i.e. $e \in K_E$ is a premise of A and $\neg s$ is a conclusion of A, where $s \in S$). Arguments may also be used to reason about the plausibility of an explanation, as the validity and applicability of causal rules can become the subject of an argumentation process. Thus, the third criterion for judging an explanation S is *implausibility* (denoted as $impl(S)$), which stands for number of elements in S explicitly contradicted by an argument A which is *not* based on evidence (i.e. $s \in S$ and $\neg s$ is a conclusion of A, where A's premises $\{a_1,\ldots,a_n\} \subseteq K_A$). Here, arguments about the plausibility of explanations are based on assumptions from K_A, as reasoning about plausibility is done using commonsense knowledge

---

[4] Note that these informal definitions have been formalized in various ways, see e.g. (Jakobovits and Vermeir 1999).

about how the world generally works. Note that for the criteria, only arguments which are not overruled support or contradict an explanation: if an argument based on evidence is itself defeated, the evidence does not support the explanation.

The above criteria can be used to judge the quality of explanations and thus they allow for the comparison of explanations. According to Josephson and Josephson (1994, 14), the evaluation of an abductive inference should be carried out by asking one or more of the following critical questions.

1. How decisively does H surpass the alternatives?
2. How good is H by itself as an explanation, independently of the alternatives?
3. How much confidence there is that all plausible explanations have been considered?
4. How good are the original data?
5. How can the cost of being wrong be weighed against the benefit of being right?
6. How strong is the need is to come to a conclusion at all, as opposed to collecting more information?

Now, question 1, 2 and 4 can be directly modeled in the formal theory. Question 2 can be answered by analyzing an explanation's evidential support and plausibility: the higher the evidential support and the lower evidential contradiction and implausibility, the better the explanation. Plausibility can be further specified if desired: Bex (2011) defines more criteria for judging the quality of explanations using story schemes and Pennington and Hastie (1993) also provide other criteria such specificity. Furthermore, we could attach different weights to evidence so that particular pieces of evidence give a higher degree of support to explanations than others. Thus we can answer question 1: if one explanation is supported by "better" evidence, it surpasses the other explanation. Finally, how "good" the evidence in itself is (e.g. how reliable a witness is) can be tested separately using the critical questions for the arguments based on the evidence (question 4). For current purposes, however, the three criteria are sufficient. Because the comparison of explanations is influenced by the standard of proof, comparing explanations according to these standards is discussed in detail in section 3.

3. Burdens of Proof and Proof Standards

Allocation of burden of proof tells each side in a dispute how strong its argumentation needs to be in order to be successful in prevailing over contention of the other side. Farley and Freeman (1995) presented a computational model of dialectical argumentation that has the notion of burden of proof as its key element. They defined it as the level of support that must be achieved by one side to win an argument. On their account, burden of proof has two functions (Farley and Freeman, 1995, p. 156): to act as a move filter at local moves in a dialogue and to act as a termination criterion that determines the winner at the end of the dialogue.

In their logical account, Prakken and Sartor (2009) define three kinds of burden of proof well-known in common law systems in terms of their framework for defeasible argumentation.[5] The first type of burden of proof is the burden of persuasion. It is set by law at the opening stage of a trial, and determines which party has to prove and what proof standard has to be met, that is, which side has won or lost the case at the end of the trial once all the arguments have been

---

[5] The burdens of claiming and contesting, which are also discussed by Prakken and Sartor (1997), assume an explicit dialogical context and will hence not be discussed here.

examined. Prakken and Sartor (2009) define it as the task of making sure there is a justified argument for one's claim. The burden of persuasion does not shift from the one side to the other during the trial. The second type of burden is called the burden of production, or sometimes burden of producing evidence. The burden of production, which is like the burden of persuasion assigned by law, specifies which party has to offer an argument based on evidence on some specific issue during the trial. If the evidence offered does not meet the standard set for this burden, the issue can be decided as a matter of law against this side.

The third type of burden is called the tactical burden of proof. It is a hypothetical assessment made at a given point by each party to try to determine whether they will win or lose the case if no further arguments are put forward at that point. A tactical burden can shift back and forth between the parties any number of times during the trial, because it depends on who has the winning argument at a particular point in the trial. Not all sources in law agree, but in general the tactical burden is the only one of the three of the three burdens that can be properly said to shift during the course of a trial. (Prakken and Sartor, 2009, 227) argue that the distinction between the burden of production and tactical burden of proof is usually not clearly made in common law, and is usually not explicitly considered in civil law countries, but is relevant for both systems of law, because it is induced by the logic of the reasoning process.

Burden of proof rests on the prior notion that there can be different standards of proof. In the domain of common law, there are four main proof standards for factual issues called scintilla of evidence, preponderance of evidence, clear and convincing evidence, and beyond reasonable doubt (Gordon and Walton, 2009, 241). The *scintilla of evidence* proof standard is met if "even the slightest amount of relevant evidence exists on an issue" (Garner, 1990, 1464). The *preponderance of evidence* standard is met by "evidence that has the most convincing force, superior evidentiary weight that […] is still sufficient to incline a fair and impartial mind to one side of the issue rather than the other" (Garner, 1990, 1301). The preponderance of evidence standard is often compared to a balance, where the evidence on one side has greater probative weight than the evidence on the other side. *Clear and convincing evidence* is "evidence indicating that the thing to be proved is highly probable or reasonably certain" (Garner, 1990, 636). This standard is supposed to be higher than that of preponderance of the evidence, but not as high as the highest standard in law, that of evidence beyond a reasonable doubt. Finally, the *beyond reasonable doubt* standard is used to determine guilt in criminal cases, and is often equated with the presumption that the defendant is innocent.

Law defines the standards using cognitive terminology, for example by using criteria of whether an attempt at proof is credible, or convincing to the mind examining it. These cognitive descriptions, although they are useful in law for a judge to instruct the jury on what the burden of proof is in the case, are not precise enough to serve the purposes of providing computational models of the standards useful in AI and law. A definition expressed in, for example, terms of the conviction of the jury that the charge against the defendant is true (Garner, 1990, 1380) is not very helpful for getting an idea of how this standard should be represented in a normative model of rational argumentation.

It is debatable whether such a precise definition of proof standards can be given. Take, for example, the standard of beyond a reasonable doubt. According to McCormick on Evidence (Strong, 1992, 447), "The term reasonable doubt is almost incapable of any definition which will add much to what the words themselves imply". Courts have held that the legal concept of reasonable doubt itself needs no definition (Strong, 1992, 447), the reason being that any definition might have to be so subtle and technical that there would be dangers of

misunderstandings if judges were to instruct juries with it. This judicial climate of opinion poses an apparently insurmountable challenge for any attempt to provide a computational model of the proof standard. However, as Tillers and Gottfried (2006) have shown, even though there is a well-settled maxim supported by judicial wisdom that the beyond reasonable standard is not quantifiable by assigning probability values to it, it does not follow that this standard of doubt is not open to precise analysis based on a computational argumentation model.

Gordon and Walton (2009) show how proof standards can be analyzed in the formal Carneades system as follows. For there to be a scintilla of evidence (SE), there should be at least one applicable argument[6] for a claim. For the preponderance of evidence (PE) standard, SE should be satisfied and the maximum weight assigned to an applicable *pro* argument (for the claim) is greater than the maximum weight of an applicable *con* argument (against the claim). For clear and convincing evidence (CCE), PE should be satisfied, the maximum weight of applicable *pro* arguments exceeds some threshold α, and the difference between the maximum weight of the applicable *pro* arguments and the maximum weight of the applicable *con* arguments exceeds some threshold β. Finally, for beyond reasonable doubt (BRD), CCE is satisfied and the maximum weight of the applicable con arguments is less than some threshold γ. Notice that here the thresholds α, β and γ are left open, and not given fixed numerical values. Doing this would involve quantifying the proof standards, which, as was argued above, is not easily done.

4. Burden of Proof and Inference to the Best Explanation

In this section, we will give an indication of how the burden of proof and proof standards might influence the process of IBE in legal trials. For a large part, we draw inspiration from 0, who provide an informal overview of IBE in trials. Note that we mainly concern ourselves with the *factual* part of trials, that is, the evidence and the facts which we might infer from this through evidential reasoning. The legal reasoning, which is intertwined with this reasoning about the facts,[7] is not shown in detail and we simply assume that the legal conclusion follows from the established facts in some way.

In the hybrid theory, having the burden of persuasion for an explanation *S* means that at the end of the trial *S* should be accepted as the correct explanation of what happened in the case. Note that just having the best explanation is not always enough: in order to satisfy, for example, a BRD proof standard *S* should be much better than the other explanations, so good that the other explanations do not even raise a (reasonable) doubt whether *S* happened. This will be further discussed below. The burden of production may be met by providing evidence that ultimately supports the explanation on which the burden rests. So, for example, one can give evidential arguments for elements of one's own explanation or the negation of elements of the opponent's explanation. Finally, the tactical burden means that one should gauge if one's explanation is currently the best and if it trumps the other explanations by a particular margin, dependent on the standard of proof.

---

[6] Roughly, an argument is considered applicable if its premises are not defeated and there is no exception to the inference.

[7] For example, once we have established that it was in fact John who killed Harry, we still need to determine whether the killing was manslaughter or murder.

Standards of proof may also be formalized in the hybrid theory. Definitions of standards of proof require us to indicate not only when one explanation is stronger or better than another but also by which margin they are better and how good they are in themselves. Following (Gordon and Walton, 2009), such margins or standards will not be given fixed values but rather they will be left open. Now, an explanation S meets the scintilla of evidence (SE) standard if there is a justified supporting argument based on evidence ($es(S) \geq 1$). An explanation S meets the preponderance of evidence (PE) standard if it meets the SE standard and it is better than each alternative explanation S'. That is, all else being equal S is either supported by more evidence ($es(S) > es(S')$) or contradicted by less evidence ($ec(S) < ec(S')$). This way of comparing explanations is similar to that in (Bex et al. 2007). However, a key difference is that in (Bex et al. 2007), if two explanations have the same evidential support and contradiction, the explanation with the lowest implausibility is best. While comparing explanations on their relative plausibility is in principle perfectly rational, we cannot say that S meets a formal proof standard if it is just more plausible than S'.

For clear and convincing evidence (CCE), an explanation S should be good in itself as well as much better than each competing explanation S'. In order to be good, S should have a high evidential support ($es(S) > \alpha$, where α is some threshold) and low evidential contradiction ($ec(S) < \beta$, where β is some threshold). In order to be much better than any alternative S', S should have either significantly higher evidential support ($es(S) - es(S') > \gamma$, where γ is some threshold) or significantly lower evidential contradiction ($ec(S') - ec(S) > \delta$, where δ is some threshold). Finally, an explanation meets the beyond a reasonable doubt (BRD) standard if it is strong and much stronger than its competing explanations (i.e. it meets the CCE standard with high threshold α and a low threshold β), and each competing explanation S' is very weak, so weak as to be highly implausible. As Pardo and Allen (2007) argue, a plausible explanation consistent with innocence creates a reasonable doubt, so for each competing explanation S', either the evidential contradiction or the implausibility should be high ($ec(S') > \alpha$ or $impl(S') > \beta$, where α and β are thresholds) if the "guilt" explanation is to meet the standard of proof. Note that the evidential support of the competing explanations is not tied to extra requirements (beyond those set for the CCE standard), as these explanations only have to be *consistent* with the evidence.

5. The Case of Anderson v. Griffin

In this section we give an example of IBE in a civil trial. In the case, which is based on the Summary of Case of Anderson v. Griffin (397 F.3d 515), the driveshaft suddenly broke on a tractor-trailer truck proceeding down an interstate highway, severing the connection between the brake pedal and the brakes. Debris kicked up from the surface of the highway (road junk) struck a pickup truck behind the tractor-trailer. The pickup truck crashed into a part of the tractor-trailer and a car following the pickup truck struck the wreckage from the collision between the two trucks, injuring the two people in the car. Plaintiffs, the two people in the car, sued the truck dealer, who (supposedly) was responsible for the technical maintenance of the trailer. Now, the plaintiffs should propose a coherent explanation from which it follows that the dealer had been negligent. Three weeks earlier, the trucking company who owned the tractor-trailer had noticed a looseness in the driveshaft and had asked the truck dealer to tighten the driveshaft. The dealer tightened all the joints except for the middle one, which broke. This first explanation is supported by the truck dealer's records about the repairs on the truck (they state that the repairmen did not repair the joint). Figure 2 shows this explanation (and defendant's alternative,

see below). In the figure, *p* and *d* denote whether an event is part of plaintiff's explanation, defendant's explanation or both.
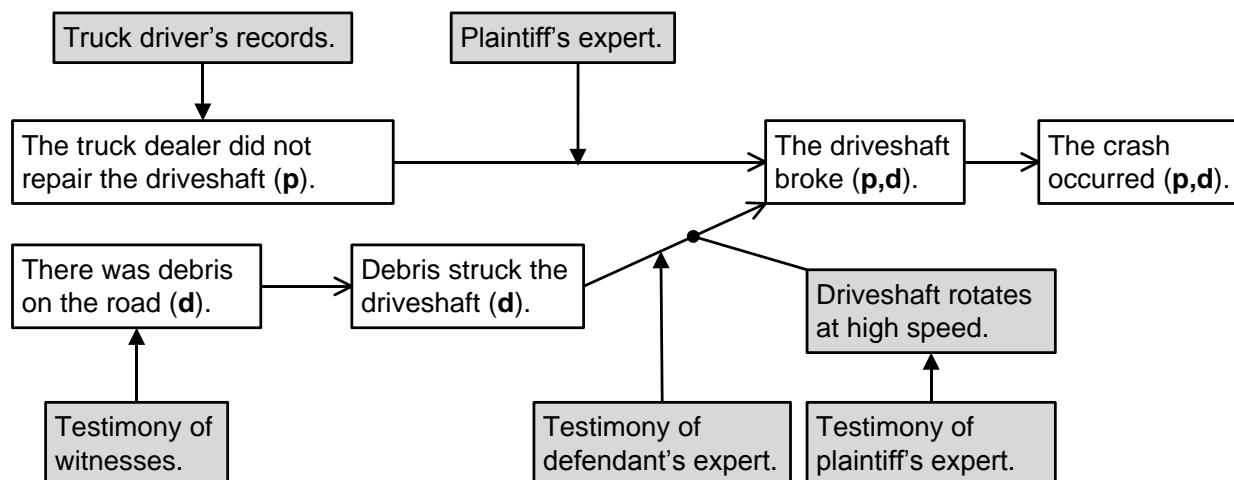

Figure 2: Two explanatory stories in the Anderson v. Griffin case

Defendant, the truck dealer, now has the tactical burden of proof: if he does not question plaintiffs' explanation or provide a better explanation for the crash from which it does *not* follow that he had been negligent, the jury will rule for plaintiffs. Defendant gives such an alternative explanation, claiming that debris struck the driveshaft properly. That there was debris on the road follows from statements made by witnesses. Defendant could also have denied the fact that he did not repair the driveshaft (i.e. attacked the plaintiff's explanation), but as his own records state he did not repair the slip yoke, this might not be a strong argument.

Plaintiffs now have the burdens of persuasion and production for their explanation whilst defendant only needs to cast sufficient doubt on this explanation, which he has done by providing a reasonable alternative which is at least as good as plaintiffs' explanation. If a verdict were to be given now, the judgment would go against the party with the burden of persuasion, in this case the plaintiffs, because they have failed to meet the burden of production, i.e. produce evidence so that a fact-finder can differentiate between explanations. Thus plaintiffs now have the tactical burden of proof and they produce evidence to improve their explanation: an expert witness who states that the crash was caused by the fact that defendant did not repair the driveshaft.

The tactical burden now shifts to defendant, as plaintiffs' explanation is slightly better (because it is supported by more evidence). Defendant can now, for example, have his own expert deny the causal link between the dealer's failure to repair the driveshaft and the crash or he may want to question the plaintiffs' expert's veracity, undercutting the support the testimony gives to plaintiffs' explanation. However, defendant chooses to support his own explanation with an expert testimony. It was proposed that the accident had been caused by debris on the highway that might have been yanked up and against the driveline by chains hanging from the truck. This explanation is also shown in Figure 2 (bottom sequence of events).

Both explanations now have equal support, so the tactical burden again shifts back to plaintiffs, who decide to attack defendant's explanation. The plaintiffs argued that a piece of road junk would be highly unlikely to strike the driveshaft with enough force to break it, because of the speed at which the driveshaft rotates (27 times a second). This argument is shown in the

sequence of grey boxes at the bottom of Figure 2. Now plaintiffs' explanation is slightly better than defendant's so the preponderance of evidence standard seems to have been met. However, in the case the jury ruled for the defendant. For some reason, they must have found that the attacking argument based on plaintiffs' expert was not convincing enough.

The sequence of dialogue in which the burdens of proof shift from one side to the other is summarized in figure 3. The shifts are indicated by the darkened boxes.
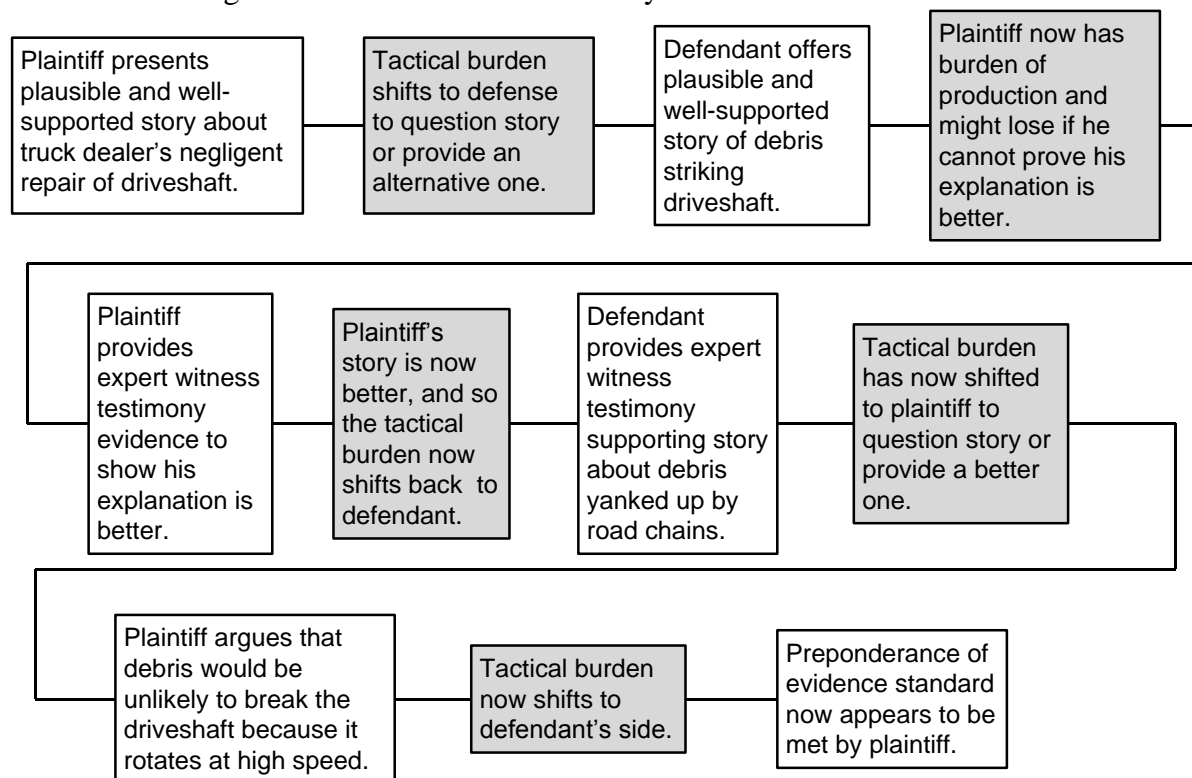
Plaintiff presents plausible and well-supported story about truck dealer's negligent repair of driveshaft.	Tactical burden shifts to defense to question story or provide an alternative one.	Defendant offers plausible and well-supported story of debris striking driveshaft.	Plaintiff now has burden of production and might lose if he cannot prove his explanation is better.
Plaintiff provides expert witness testimony evidence to show his explanation is better.	Plaintiff's story is now better, and so the tactical burden now shifts back to defendant.	Defendant provides expert witness testimony supporting story about debris yanked up by road chains.	Tactical burden has now shifted to plaintiff to question story or provide a better one.
Plaintiff argues that debris would be unlikely to break the driveshaft because it rotates at high speed.	Tactical burden now shifts to defendant's side.	Preponderance of evidence standard now appears to be met by plaintiff.	

Figure 3: Sequence of Shifting Burdens in the Anderson v. Griffin case

6. The Case of Jackson v. Virginia

As an example of a criminal case we discuss Jackson v. Virginia: 443 U.S. 307. The case concerns the death of Mary Cole, who had been a member of staff at the county jail where she had befriended James Jackson, an inmate. After his release, Cole and Jackson stayed in contact. Witnesses testified that on the day the crime was committed, Jackson had been drinking while shooting at targets with his revolver. Later that day, Cole and Jackson drove to a diner where they were seen drinking by two police officers. As the two were preparing to leave the diner in Cole's car the sheriff testified that he had offered to keep Jackson's revolver until he sobered up, but that Jackson had said this would be unnecessary since he and Cole were about to engage in sexual activity. The same evening, Jackson drove from Virginia to North Carolina. A day and a half later, Cole's body was found in a secluded parking lot, naked from the waist down, her slacks beneath her body; Jackson was arrested a few days later.

Jackson was convicted of first-degree murder by the Virginia court and sentenced to 30 years in the penitentiary. Then he appealed to the Supreme Court arguing that the beyond reasonable doubt standard had not been met. The case went to the Supreme Court (Jackson v. Virginia 443

U.S. 307 (1979)) because of the raising of the question whether there was sufficient evidence to justify a rational trier of fact to find guilt beyond a reasonable doubt. His argument was that the evidence was insufficient to support the finding that he had intended to kill Mary Cole. Jackson had appealed to a District Court in Virginia, claiming that there was no evidence of premeditation. This Court found that there was some evidence that he had intended to kill Cole, and listed this evidence as follows (page 312). He reloaded his gun after firing warning shots. He had time to reload his gun. He shot the victim not once but twice. The court also supported the conclusion of the state trial judge that he was not so intoxicated that he was incapable of premeditation. A problem posed, as expressed in the Supreme Court notes, is whether a small amount of evidence, called a "mere modicum of evidence", is by itself adequate to rationally support a conviction to the standard of beyond a reasonable doubt (page 320).

The prosecution in the Supreme Court case proposed an explanation S_P from which it followed that Jackson murdered Cole. The explanation first recounted the events at the diner and then, without mentioning a specific motive, argued that Jackson intended to kill Cole, that he shot her with his revolver, killing her, and that he then drove to North Carolina. The story was supported by witness statements (on Jackson shooting and drinking) and police officers' statements, as well as expert medical evidence that Cole had been shot twice at close range with Jackson's revolver. Because Jackson admitted he had shot Cole, the main factual dispute at the trial was whether there was sufficient evidence to prove Jackson's *intention* to kill Cole. Evidence that he had so intended was that Jackson had fired two shots at Cole at close range, shots that were predictably fatal given that he was a person experienced in the use of firearms. Figure 4 shows this argument supporting the prosecution's story S_p.
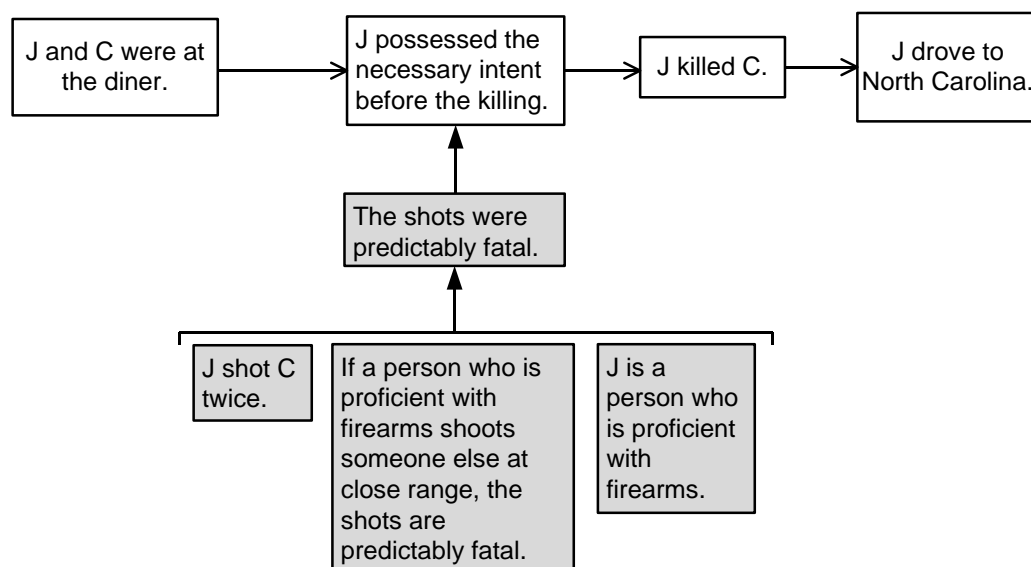

Figure 4: The Argument that Jackson Intended to Shoot Cole

Thus, the prosecution gave an argument A_1 that Jackson knowingly (and hence intentionally) performed his act.

Teleological reasoning can be used to establish intent by drawing an inference from premises concerning facts of a case to a conclusion that an intention exists (Walton and Schafer, 2006). A sequence of teleological reasoning leads from a set of circumstances in a case to a hypothesis

that postulates the existence of a motive. To see how this was done in (Walton and Schafer, 2006), we have to put this scheme into a broader argumentation framework by introducing the argumentation scheme for practical reasoning. This scheme can be used to show how an agent reasons from a goal (or intention) to action, and also how an agent can reason abductively from an action to the postulation of an intention. How this scheme can be so used will be illustrated n the next case (the Beard case) below.

The prosecution met tactical burden for their explanation: S_P had sufficient evidential coverage and no evidential contradiction and there was no competing explanation. Jackson had the tactical burden which he could meet by casting doubt on S_P. Under the BRD standard, this doubt may be created by proposing a sufficiently plausible alternative explanation consistent with the evidence. Jackson presented the following story S_J. After they left the diner, Cole had made sexual advances towards Jackson. When he had resisted her sexual advances, she had attacked him with a knife. He defended himself by firing warning shots into the ground, and then reloaded the weapon. When Cole attempted to take the gun away from him, it went off during the ensuing struggle. He claimed that he had fled without seeking help for Cole because he was afraid. Later, during the trial he claimed that he had acted in self-defence. He also offered the argument A_2 that, as the State's own evidence (i.e. the police officers' testimonies) showed, he had been too intoxicated to form the specific intent necessary to make him guilty of the crime of first-degree murder. In other words, he attacked S_P.

The tactical burden then shifted to the prosecution, who had to defend the argument A_1 (for Jackson's intentions) by giving a counterargument to A_2 (that he was intoxicated) as well as show that Jackson's explanation S_J is implausible. First, the fact that Jackson drove without mishap from Virginia to North Carolina was taken to be at odds with his argument of extreme intoxication at the time of the killing. This counterargument, together with A_1 and A_2 are shown in figure 5.
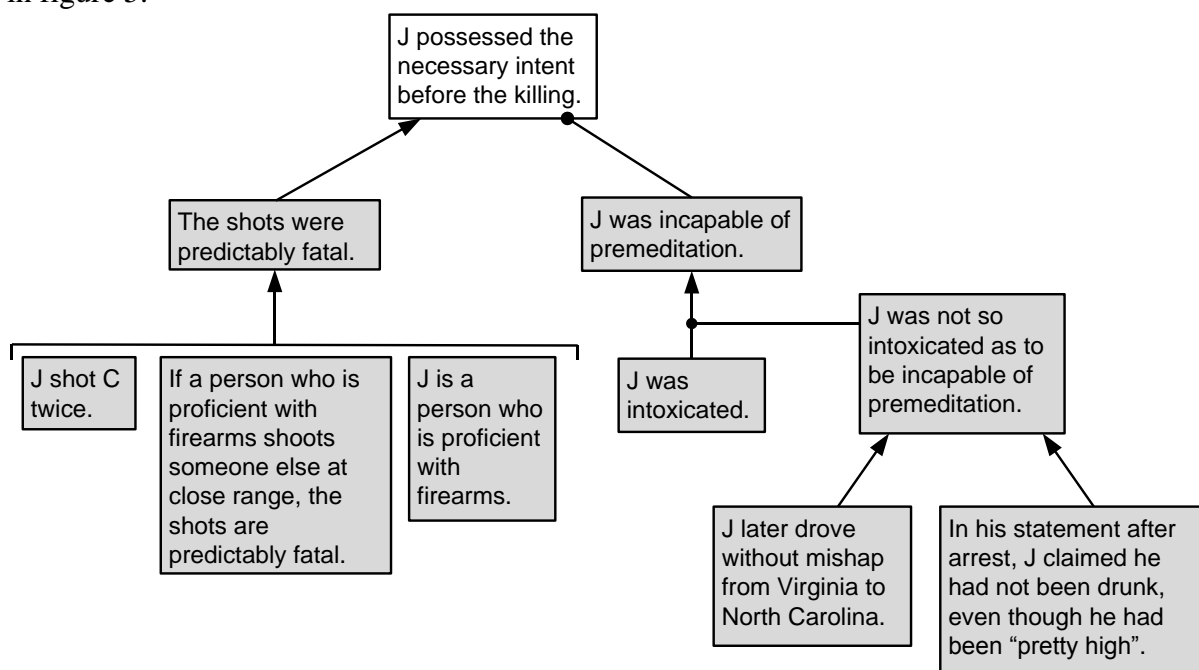

Figure 5: Prosecution Arguments Attacking Jackson's Story about No Premeditation

Thus, A_2 was considered overruled and A_1 again justified.

It was further argued that S_J contained implausible elements. First, Jackson said to the police officer's he was going engage in sexual activity with Cole but later supposedly resisted Cole's sexual advances. Furthermore, it was found implausible that Cole first willingly removed part of her clothing and then attacked him with a knife when he resisted her advances, even though he was armed with a loaded revolver that he had just demonstrated he knew how to use. The whole story in figure 6 is implausible, and many aspects of it could be questioned.
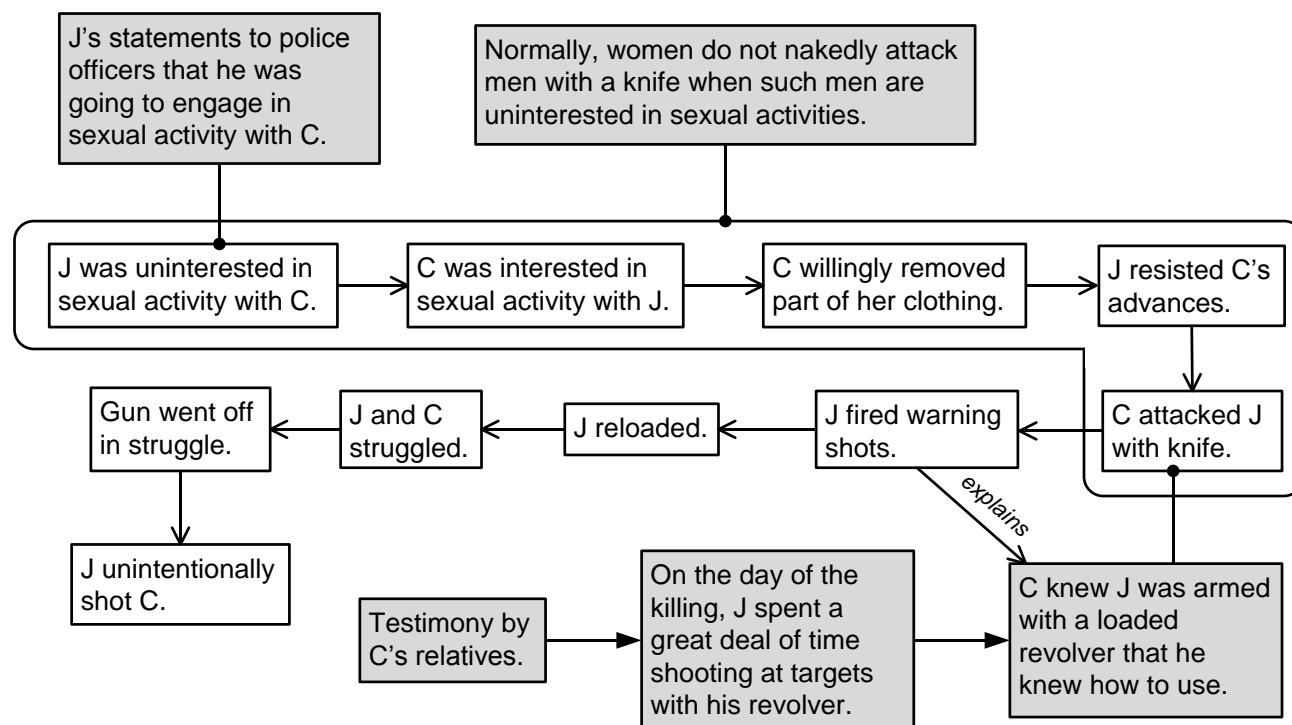
Figure 6: The Full Sequence of Jackson's Story Questioned by an Argument

First, consider the argument based on the testimonies of the relatives. From this evidence we can conclude that Cole knew Jackson was armed with a loaded revolver. This attacks the story: if she knew he had a gun, it would be foolish to attack him with a knife so this event is not very plausible. Furthermore, the conclusion also shows that Jackson's case as a whole is contradictory: the fact that Cole knew he was armed follows not only from the relatives' testimonies but is also explained by the event `Jackson fired warning shots' in Jackson's story.

The second case of implausibility concerns not single causal links or events, but rather the story as a whole. Jackson claims he was uninterested in Cole and that hence she took off her clothes and attacked him with a knife. This is implausible because it does not fit any plausible *story scheme* (Bex 2011)[8] or script (Schank and Abelson 1977), a general sequence of events as we expect them to normally happen. As an example of such a story scheme, consider the well-known restaurant script, which describes a pattern of events that a typical story about a visit to a restaurant contains (e.g. ordering, eating, paying). Now, if a story deviates from such a script we usually consider it to be implausible (for example, a man enters a restaurant, tells the waiter he needs a hamburger as soon as possible because he is very hungry, receives his hamburger from

---

[8] Story schemes are formalised in (Bex 2011).

the waiter, stands up from his seat and removes his pants, dances the tango as he moves around the floor, offers the waiter his pants in exchange for the hamburger and starts reciting the Declaration of Independence from memory when confronting a policeman called to the restaurant deal with the disturbance). In the case of Jackson, it is the genders of the actors that are the wrong way around (women do not usually try to force men to have sex, at least not physically) and the sequence of events also makes no sense (first Cole undresses, then she grabs a knife; from where, one might wonder).

   Now S_P's quality is restored (as A_1 again supports it) and S_J's implausibility is demonstrated and thus the beyond reasonable standard is met, as is evident from the trial record.

7. The Beard Case

This case (US v. Beard, 354 F. 3d 691 - Court of Appeals, 7th Circuit 2004) was an appeal of an earlier conviction in which John Beard was found guilty of carrying a gun in relation to a drug offense, a violation of the law. Police were watching a parking lot in which two cars were parked side-by-side. Beard got out of one of the cars, entered the other car for a few minutes, and then returned to his car. Both cars left the scene and were later stopped by the police. In Beard's car police found drugs in a secret compartment behind the rear seat and a loaded derringer in the closed center console of the front seat, concealed under some papers. In the other car they found cash. Beard did not own the car. He had borrowed it eight months before from its owner, and had been seen driving it from time to time during this period. However, he was not the only person who used the car during that period. The papers concealing the derringer did not seem to have belonged to Beard either. They did not pertain to him.

   Let's look at the arguments on both sides. The argument given on the prosecution side can be set up as follows. It depends on premises that express some generally known characteristics of activities during drug dealing cases.

   Premise 1: The gun was being carried during a drug offense.
   Premise 2: Drug dealers bring guns to a deal to instil fear in their business associates, or if they feel the need for protection.
   Conclusion: It is reasonably clear that the gun was Beard's.

This argument is a fairly weak one, but it carries some weight, because of the general knowledge from previous cases about how drug dealers generally conduct their operations. It tells us that drug dealers have reasons to carry a gun during a drug offense, and it can be inferred from this as an interim conclusion that drug dealers commonly bring a gun along during a drug deal. These assumptions suggest the conclusion that Beard took the derringer along during this drug deal and concealed it in the console of the car.

   The issue was whether the derringer belonged to Beard. Let's look at the argument that it belonged to Beard using a standard argument mapping representation in figure 7. The ultimate conclusion, represented at the top left, is the statement that it is reasonably clear that the gun belonged to Beard. To the right of the diagram at the top we have a linked argument with two premises. One premise is the statement that the gun was being carried during a drug offense. The other is the statement that drug dealers commonly bring a gun along during a drug deal. This latter statement has been inserted as an implicit premise, as indicated by the fill in the text box being shaded (in yellow). Another implicit premise, the statement that drug dealers have reasons

to carry a gun along during a drug offense, supports the implicit premise above it. This statement is in turn supported by the arguments below it. Note that this way of representing the argument makes it depends on a number of assumptions in such a way that the argument is not very strong. Not only does it depend on two implicit premises, but it also depends on generally accepted knowledge about how the way things go in a drug deal. It depends on your knowledge about how things can be generally expected to go in a script that, at least in general outline, follows a stereotypical pattern. Two parties exchange something for money. The transaction is illegal, and large amounts of money are involved. Hence the situation is a dangerous one, and as we all know from Hollywood movies, it can easily erupt into gunfire.
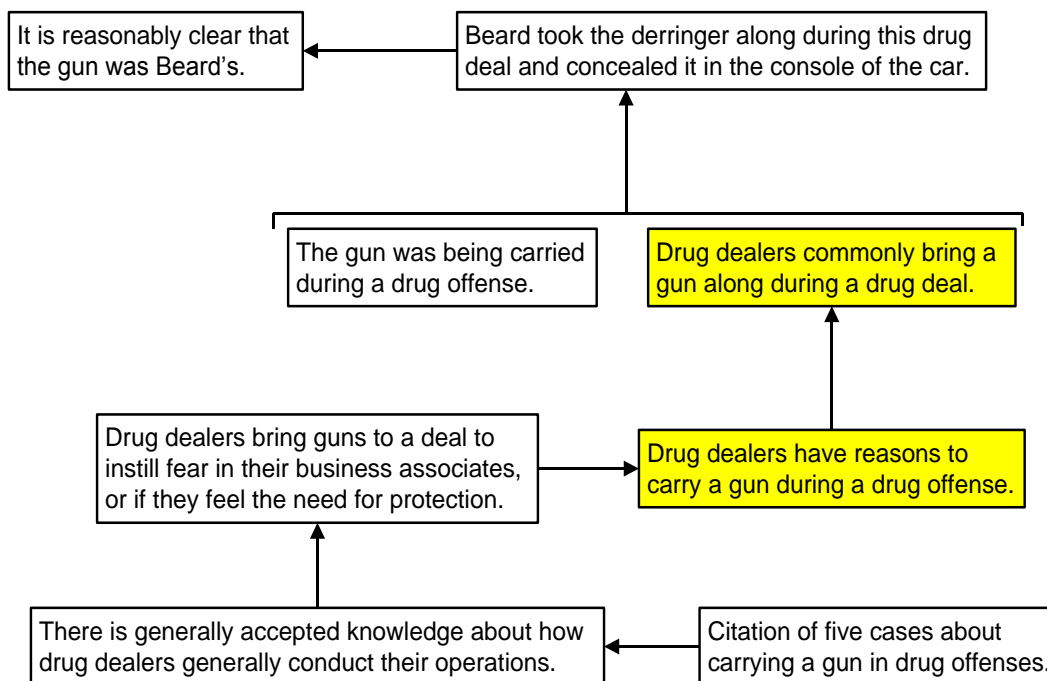

Figure 7: Argument Map of the Argumentation Concluding the Gun was Beard's

Adopting this way of modeling the argumentation in the case, we could also examine the argument on the other side. Conceivably one of the others who used the car during the eight month period placed the gun there. However, there doesn't seem to be any evidence supporting the conclusion that some other person left a gun in the console. The argument on the other side appears to be nonexistent. So looking at the case from a standard argumentation point of view, we have a weak argument on one side and no argument on the other. But there is a different way to analyze and evaluate the argument by using inference to the best explanation.

   Here we put forward another hypothesis giving an analysis of the logical structure in the decision. The hypothesis is that it was decided by inference to the best explanation of the facts. The choice had to be made between three explanations of the facts of the case. What needed to be explained was how the gun got into the console between the seats in the car. One of the three possible explanations is that the derringer was the property of the car's owner, and that she had the loaded gun in the console when she lent Beard the car. This explanation was ruled out by the court. We are not told why, but possibly it was because the car owner had no previous criminal convictions and was not related to the drug trade. The second explanation is that Beard put the

derringer in the console because he had the goal of instilling fear in a business associate or the goal of protection during a dangerous drug deal. This seems to be a fairly plausible explanation, because it had been decided by the previous court that Beard was guilty of conducting the sale of drugs, and it is known from previous cases that there is common knowledge about how drug dealers conduct these exchanges. It is known these exchanges can be highly dangerous, and that therefore the drug dealers have reasons for being armed.

Next we show how this explanation is based on the use of abductive reasoning to draw inferences to agents presumed goals and intentions based on evidence of actions and speech attributed to the agent. In forward practical reasoning (see e.g. Walton, Reed and Macagno, 2008, 323) an agent has a goal, sees that there is an action that is a means of fulfilling to or contributing to the goal, and concludes that he ought to perform the action. Backward practical reasoning is an abductive use of practical reasoning in which we observe the actions and speech of an agent, an attempt to explain them by hypothesizing that the agent carried them out in order to realize some goal or intention (Bex et al. 2009).

In figure 8 it is shown how an explanation can be given to support the hypothesis that Beard put the derringer in the console. The explanation is based on practical reasoning applied to an explanation (Bex et al., 2009) and it postulates that Beard had a goal in mind, and that having an available gun was a means required to fulfill this goal. The explanation is based on the argument that drug dealers can use a gun to achieve their goal of instilling fear in their associates. The argument thus provides evidence for the causal link between Beard wanting to instil fear and him bringing a gun, and in turn this provides an explanation for Beard's supposed action of putting the derringer in the console of the car.
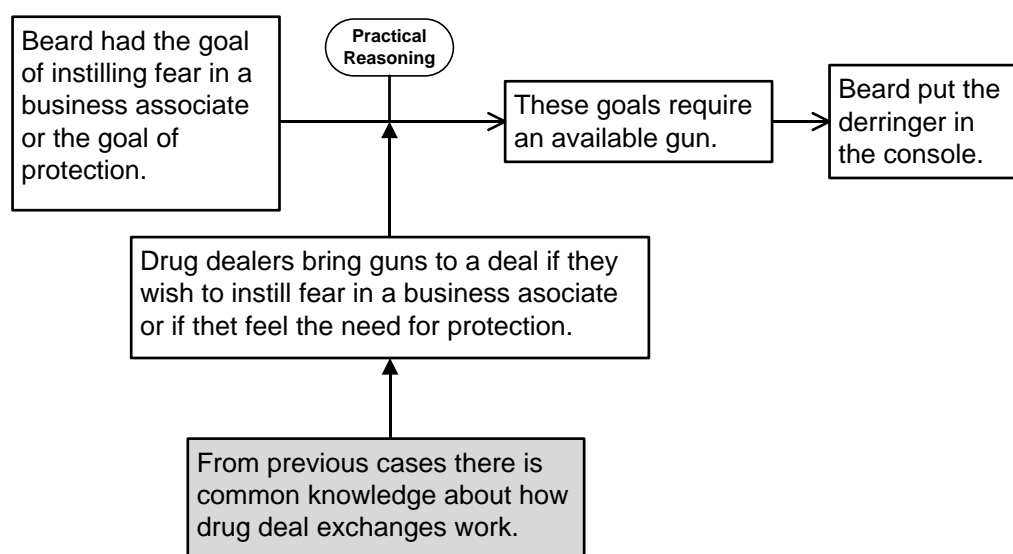

Figure 8: The Story that Beard Put the Gun in the Console

The third explanation is that one of the other people who used car during the eight-month period placed the derringer in the console. This explanation is possible, but as pointed out in the court notes, it is not very plausible that someone would place a loaded gun in the console, cover it with papers to conceal it, and then forget about it. If the defence had investigated further to collect evidence about who borrowed the car during that period, and what reasons they might have had

for placing a loaded gun in the console, and then later not removing it, such an investigation could have given at least some reason to think that this explanation has some slight plausibility. However, the defence did not present any such evidence, nor did they put forward an argument giving reasons to suppose that this third explanation could be even slightly plausible. Therefore, according to the interpretation of the argumentation in the case put forward here, the decision was arrived at on the basis of evaluating and comparing these three explanations. The first explanation was ruled out as implausible. The second explanation was shown to be plausible, given what is known about the general practices of drug dealers during an exchange, even though there was no evidence to support the explanation other than that given by this general knowledge.

The remarks quoted above suggest that the hypothesis that someone who borrowed the car placed a loaded gun in the console and covered with papers to conceal it is not a very lively possibility. As will be suggested in some remarks quoted from the court transcript below, it seems implausible that someone would conceal a gun in the console of a car and then forget about it.

No one supposed that the derringer was the property of the car's owner - that she hid a loaded gun in the center console when she lent Beard the car. Hence there remain only two possible explanations of the facts. One is the explanation that Beard placed the gun in the console. The other is the explanation that someone who borrowed the car from Beard placed it in the console. The remarks in the court transcript compare the two explanations.

[Quote begins] Since others besides himself used the car during the eight months that he possessed it, conceivably the gun was left there by one of these users, but that is highly unlikely. It would mean that someone who borrowed the car from Beard placed a loaded gun in the console, covered it with papers to conceal it, and then — what? Forgot about it? That is possible, but it was not so lively a possibility as to compel a reasonable jury to acquit Beard [Quote ends].

The problem for the defence is that it provided no evidence to support the alternative story, and without such support it is inherently implausible. The court's remarks make this clear by asking a question. Is it plausible that someone borrowed the car from Beard, placed a loaded derringer in the console, covered with papers to conceal it, and then forgot about it? As these remarks suggest, this sequence of actions is just not very plausible as a story.
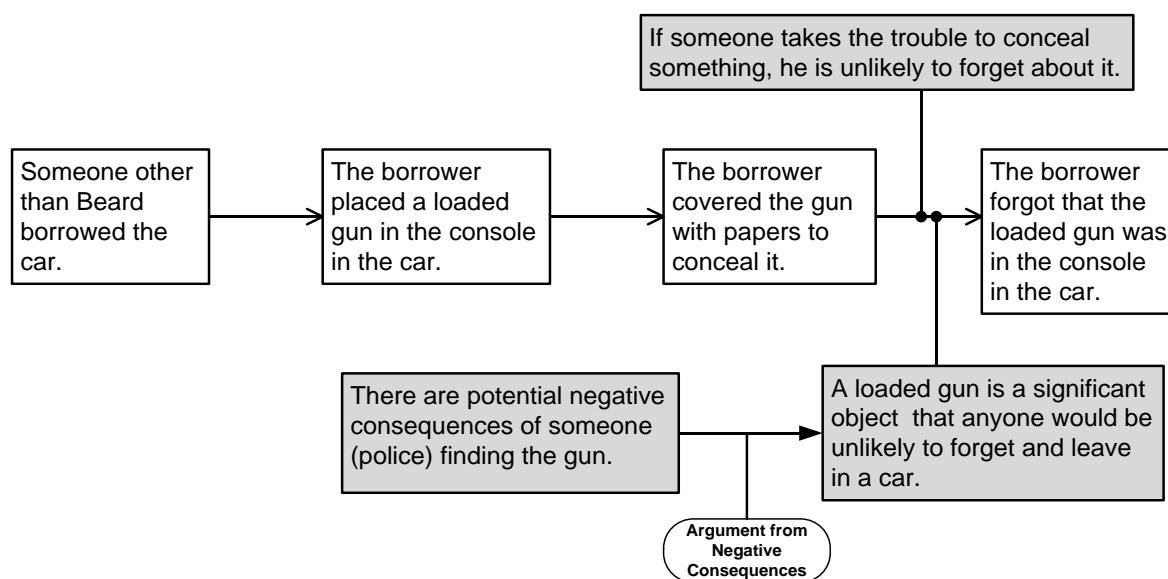

Figure 9: The Story that the Car Borrower Left the Gun in the Console

The third explanation could possibly have been given some plausibility by the collecting of additional evidence about the particulars of the case. However the defence did not present any evidence of this sort. Also, the argument posed in the question above, "Is it plausible that someone borrowed the car from Beard, placed a loaded derringer in the console, covered with papers to conceal it, then forgot about it?" could have been turned against Beard's side. The defence could have asked the question "Is it plausible that Beard placed a loaded derringer in the console, covered with papers to conceal it, then forgot about it?" The tactic would have turned the argument on its head. But the defence did not use this clever argument either. For these reasons, the best explanation of the three was the second one. Hence it can be conjectured that the process of reasoning that led to the conclusion that the gun belonged to Beard can only be adequately represented by using inference to the best explanation.

At the appeal, Beard's lawyer offered no explanation for the presence of the derringer in the car. The reason suggested for this failure by the trial notes (written by Judge Posner) is worth quoting.

[Quote begins] The lawyer seems to have thought that since the government had the burden of proof, and Beard was privileged not to testify (and he did not testify), it was irrelevant that the jury was given no alternative to the government explanation as to whose gun it was [Quote ends].

The explanation given by the court of the approach of the defence is also worth quoting, because it cites a number of precedents and writings of legal scholars showing that the argumentation in a case like this needs to be evaluated by looking at competing explanations, and depends on evaluating the plausibility of alternative explanations against the plausibility of an explanation that has been put forward.

[Quote begins] "[T]he plausibility of an explanation depends on the plausibility of the 693*693 alternative explanations." Spitz v. Commissioner, 954 F.2d 1382, 1384 (7th Cir. 1992). And so, "realistically, a jury called upon to decide guilt must compare the prosecution's version of the incident giving rise to the case with the defence version." Sandoval v. Acevedo, 996 F.2d 145, 150 (7th Cir.1993); see Ronald J. Allen, "Factual Ambiguity and a Theory of Evidence", 88 Nw. L.Rev. 604, 611 (1994); Allen, "The Nature of Juridical Proof," 13 Cardozo L.Rev. 373, 409-40 (1991); Allen, "A Reconceptualization of Civil Trials," 66 B.U.L.Rev. 401 (1986). Confidence in a proposition, such as Beard's guilt, is created by excluding alternatives and undermined by presenting plausible alternatives. See United States v. Tucker, 716 F.2d 576, 580 (9th Cir.1983); United States v. Reyes, 302 F.3d 48, 56 (2d Cir.2002) [Quote ends].

Thus the problem in the case is that the defence failed to carry out a reasonable search for evidence that would support its explanation of the case. As the trial notes put the point, this objective of raising sufficient doubt against the given explanation "is not satisfied just by looking for ways of poking holes in the government's case". In other words, asking critical questions about the government's case is not enough to refute their argument. What is being said is that since the government has provided an explanation, to mount an adequate defence the other side must put forward a competing explanation that is at least plausible enough to raise reasonable doubt about the government's explanation. Since the defence failed to do this, and indeed did not provide any evidence that supported any explanation at all, the jury was left with no alternative to the government's explanation of what happened in the case.

As the court's remarks make clear, the defence offered no alternative argument for the conclusion that the gun belonged to somebody else who had used the car. The defence lawyer might have thought that, since the prosecution had to prove beyond reasonable doubt, and since the prosecution's argument was relatively weak, no counterargument against it was needed. This case was an appeal, and it was noted that the jury hung at Beard's first trial. So it is interesting to take into account that the jury at Beard's first trial did not find that he was guilty of

carrying a gun in relation to his drug offense beyond a reasonable doubt. In the court notes it is stated that the jury could reasonably have acquitted Beard on this charge. Indeed, the basis of the appeal was that the evidence was so thin that no reasonable jury could have found Beard guilty beyond a reasonable doubt. And it was also emphasized that the standard applied in judging the appeal was not one of whether the defendant was guilty beyond a reasonable doubt, but one of "whether a reasonable jury could have found he was guilty beyond reasonable doubt". Several cases were cited to support this interpretation (United States v. Powell, 469 U.S. 57, 67, 105 S.Ct. 471, 83 L.Ed.2d 461 (1984); United States v. Morris, 349 F.3d 1009, 1013 (7th Cir.2003); United States v. Capozzi, 347 F.3d 327, 337 (1st Cir.2003)).

This case is interesting from an argumentation point of view because it suggests that the argumentation in a criminal case is based on competing explanations, and on the comparative evaluation of the acceptability and plausibility of these explanations. It also tells us something about the meaning of the how the expression 'beyond reasonable doubt' is applied as a proof standard. This was a criminal case, and so, as is often emphasized, the defendant had to be convicted by evidence strong enough to meet the beyond reasonable doubt standard. As stated in Beard's first trial, the evidence the gun belonged to him was very slim. Indeed, there was no circumstantial evidence that the gun belonged to Beard. All the evidence was about how things normally go in drug deal. So how could it be decided by the jury on this basis that Beard was guilty of carrying a gun in relation to his drug offense beyond a reasonable doubt? It does not seem to be logically possible that the issue could have been decided on the basis of the prosecution providing an argument for this claim that was so strong that it met the standard of beyond a reasonable doubt. Some other interpretation of the logic of the argumentation in the decision is called for.

The really interesting problem posed by the case is if you look at the argumentation in the case in the standard way by analyzing an exclusively as an argument without recourse to explanation, or to inference to the best explanation, it appears that the evidence used to convict did not meet the standard of beyond reasonable doubt. Judge Posner's comment that the standard should not be seen as one of finding guilt beyond reasonable doubt, but rather one of whether a reasonable jury could have found guilt beyond a reasonable doubt, seems to soften the strangeness of this finding. However, it does not solve the problem altogether. We still need to ask how a reasonable jury could have found guilt beyond a reasonable doubt when the evidence was so slim. There is still a clash between the judge's instruction that the jury has to find guilt to the standard of beyond reasonable doubt and the capability for the jury to come to a decision that the defendant is guilty based on such slim evidence. If you look at the beyond reasonable doubt standard of proof as a criterion that has to be met by a chain of argument that mounts up so much evidence that the conclusion has to be proved without leaving any possibility for reasonable doubt, the standard is surely not met in a case like that of Beard. The only way to analyze the structure of the reasoning in such a case in a realistic way is to look at it, instead, as based on a comparing of explanations for the facts in the case.

8. Conclusions

In this paper we have shown how notions of burden of proof and proof standards can be incorporated in a formal model of IBE, namely Bex et al.'s (2010) hybrid theory of stories and arguments. We have also given the first example of a civil case in the hybrid theory, which shows that at least some civil cases lend themselves well to being analyzed in the theory. In

addition to adding to the research on the burden of proof, we have thus also looked at how the hybrid theory may be expanded. An interesting further expansion of the theory could allow one to perform legal reasoning (e.g. reasoning with legal rules and exceptions). It would then be interesting to see how this combination of (hybrid) reasoning about the facts and the law would influence the modeling of the burden of proof and proof standards.

Even though a precise, quantitative definition of proof standards cannot reasonably be expected, the hybrid theory is a good tool for analyzing and modeling these standards. Gordon and Walton (2009) and Prakken and Sartor (2009) argue that for a standard of proof to be met, one position has to be stronger (by a certain margin) than another and the criteria for the quality of explanations can be used to give a fine-grained analysis of why one explanation is stronger than another. In future research, the additional criteria defined in (Bex et a., 2010) and (Pennington and Hastie, 1993) may also be used to further analyze the standards.

Another possible avenue for future research in this respect is to expand the hybrid theory to make it possible to reason *about* the (relative) strength of explanations, in the same way as Prakken and Sartor (1997) allow one to give reasons for priorities between arguments. The criteria are not hard-and-fast rules for which explanation is the best; this often depends on the context of the actual case. In some cases, for instance, Pennington and Hastie's (1993) criterion of *specificity* might play an important role whilst in other cases, such as the Jackson case, it is sufficient that the best explanation outlines what happened in abstract terms. Being able to give reasons for priorities between explanations would allow using the criteria as they are intended, namely as *reasons* for why one explanation is better than another.

Perhaps the most controversial aspect of our paper is the way we have modeled the beyond reasonable doubt standard of proof using the hybrid theory. We can broadly sum up in outline the position we have taken with respect to modeling argumentation meeting the BRD standard as follows. An explanation of the facts of a case meets the beyond reasonable doubt standard if (a) it is a plausible explanation and (b) the competing explanations (if there are any) are implausible. Criterion (a) means that the explanation selected not only fits together into a coherent story, but is also supported by evidential arguments and meets other criteria of a good explanation. Criterion (b) means that each of the competing explanations is so weak that it fails to raise a reasonable doubt. The tactical burden of proof shifts back and forth during the trial as each side offers arguments to support its own explanation and arguments to weaken the explanation offered by the other side. At the end, the prosecution has to fulfill its burden of persuasion set according to the beyond reasonable standard, or otherwise the defence wins.

Our analyses of these cases throws new light on how the logic of reasoning with burden of proof in both civil and criminal cases can be modeled by an IBE framework that reaches an outcome by comparatively evaluating competing explanations. It also tells us something about the meaning of the how the expression 'beyond reasonable doubt' can be applied as a proof standard in the framework. In the Beard case it did not seem to be logically possible that the issue could have been decided in the way it was on the standard argumentation approach that requires the prosecution to provide an argument for the charge that was so strong that it met the standard of beyond a reasonable doubt. However, we have argued for a different model of the logic of the reasoning in this and comparable cases that may not justify the decision, but can at least give a more robust and realistic account of how the reasoning behind it works.

References

A. Amaya, Inference to the best legal explanation. In H. Kaptein, H. Prakken & B. Verheij (eds.) *Legal Evidence and Proof: Statistics, Stories, Logic*. Farnham: Ashgate Publishing, Applied Legal Philosophy Series, (2009), 135-159

F.J. Bex, *Arguments, Stories and Criminal Evidence – A Formal Hybrid Theory*. Law and Philosophy Library, Vol. 92. (2011) Springer, Dordrecht.

F. Bex, T. Bench-Capon and K. Atkinson, Did he jump or was he pushed? Abductive practical reasoning, *Artificial Intelligence and Law*, **17**, 2009, 79-99.

F.J. Bex, S.W. van den Braak, H. van Oostendorp, H. Prakken, B. Verheij and G. Vreeswijk, Sense–making software for crime investigation: how to combine stories and arguments? *Law, Probability and Risk* **6** (2007), 145-168.

F.J. Bex, P.J. van Koppen, H. Prakken and B. Verheij, A hybrid formal theory of arguments, stories and criminal evidence, *Artificial Intelligence and Law*, **18:2** (2010).

F.J. Bex and D. Walton, burdens and standards of proof for inference to the best explanation. In R. Winkels (ed.): Legal Knowledge and Information Systems. JURIX 2010: The Twenty Third Annual Conference. (2010), 37-46. IOS Press, Amsterdam.

L. Console and P. Torasso A spectrum of logical definitions of model–based diagnosis. *Computational Intelligence* **7** (1991) 133 – 141.

P.M. Dung, On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming, and n-person games. *Artificial Intelligence* **77** (1995), 321-357.

A.M. Farley and K. Freeman, burden of proof in legal argumentation. In T. Bench-Capon (ed.), *5th International Conference on Artificial Intelligence and Law* (1995) 156-164, ACM, New York.

J.B. Freeman, *Dialectics and the Macrostructure of Arguments: A Theory of Argument Structure* (1991) Berlin: Foris Publications.

B.A. Garner, *Black's Law Dictionary* (9th ed.), Thomson Reuters, St Paul, Minn., 1990.

T.F. Gordon and D. Walton, Proof burdens and standards. In I. Rahwan and G. Simari (eds.) *Argumentation and Artificial Intelligence*, 239-260, Springer, Berlin 2009.

H. Jakobovits and D. Vermeir, Robust Semantics for Argumentation Frameworks, *Journal of Logic and Computation*, **9:2**, (1999), 215-261.

J.R. Josephson, On the proof dynamics of inference to the best explanation. *Cardozo Law Review*, **22** (2001), 1621-1643.

J. R. Josephson and S. G. Josephson, *Abductive Inference: Computation, Philosophy, Technology*. New York: Cambridge University Press, 1994.

M.S. Pardo and R.J. Allen, Juridical proof and the best explanation, *Law and Philosophy* **27** (2007), 223-268.

N. Pennington and R. Hastie, Reasoning in explanation-based decision making. *Cognition* **49:1-2** (1993), 123-163.

H. Prakken, An abstract framework for argumentation with structured arguments. *Argument and Computation*, **1:2** (2010), 93–124.

H. Prakken and G. Sartor Argument-based extended logic programming with defeasible priorities. *Journal of Applied Non–classical Logics* **7** (1997), 25-75.

H. Prakken and G. Sartor, A logical analysis of burdens of proof. In H. Kaptein, H. Prakken & B. Verheij (eds.) *Legal Evidence and Proof: Statistics, Stories, Logic*, 223-253. Applied Legal Philosophy Series, Ashgate Publishing, Farnham, (2009).

C.A. Reed and G.W.A. Rowe, Araucaria: software for argument diagramming, analysis and representation. *International Journal of AI Tools* **13** (2004) 961 – 980.

R.C. Schank and R.P. Abelson *Scripts, Plans, Goals and Understanding: an Inquiry into Human Knowledge Structures* (1977) Lawrence Erlbaum, Hillsdale (New Jersey).

J.W. Strong, *McCormick on Evidence* (4th ed.), West Publishing Co., St. Paul, Minn., 1992.

P. Tillers and J. Gottfried, Case comment – United States v. Copeland, 369 F. Supp. 2d 275 (E.D.N.Y. 2005): a collateral attack on the legal Maxim that proof beyond a reasonable doubt is unquantifiable?.*Law, Probability and Risk* **5** (2006), 135−157.

D. Walton, C. Reed, and F. Macagno, F. *Argumentation Schemes*. Cambridge: Cambridge University Press (2008).

D. Walton and B. Schafer. Arthur, George and the Mystery of the Missing Motive: Towards a Theory of Evidentiary Reasoning about Motives, *International Commentary on Evidence*, **4:2** (2006), 1-47.