

Burdens and Standards of Proof for Inference to the Best Explanation

Floris BEX^{a,1} and Douglas WALTON^b

^a*Argumentation Research Group, University of Dundee, United Kingdom*

^b*Centre for Research in Reasoning, Argumentation and Rhetoric (CRRAR), University of Windsor, Canada*

Abstract. In this paper, we provide a formal logical account of the burden of proof and proof standards in legal reasoning. As opposed to the usual argument-based model we use a hybrid model for *Inference to the Best Explanation*, which uses stories or explanations as well as arguments. We use examples of real cases to show that our hybrid reasoning model allows for a natural modeling of burdens and standards of proof.

Keywords. burden of proof, inference to the best explanation, arguments, stories

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 [6], [12]. 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, [2] have argued that argumentative approaches such as [6],[12] 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. [9] have shown that using explanatory stories is closest to how legal decision makers actually think about a case and [8] have argued that reasoning in trials involves the relative plausibility of the various explanations of the evidence. It is argued that this is not only true for criminal but also for civil trials, in which the parties have to provide competing "theories" about the facts of the case.²

In sum, the reasoning performed in legal trials involves arguments as well as explanations, at least when reasoning about the facts of the case. [2] propose a hybrid theory of *Inference to the Best Explanation* (IBE), which consists of a combination of

¹ Corresponding Author.

² [8] 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]).

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. Ultimately, the alternative explanations in a case should be compared. In this comparison, the burden of proof and standards of proof play an important role. 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. [8] 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 entirely informal and not specified in exact detail. [7] gives 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.

In this paper we make a first tentative step in logically modeling reasoning with the burden of proof and proof standards in IBE, using [2]'s hybrid theory of IBE as our model of reasoning. Thus, we want to explore how the different types of burdens and standards of proof may be modeled for reasoning that is not purely argument-based. The rest of this paper is organized as follows. Section 1 contains a summary of [2]'s hybrid theory. Section 2 briefly discusses different types of legal burdens and standards of proof. In section 3, we model these burdens and standards of proof in the hybrid theory and provide examples of a civil case (section 3.1) and a criminal case (section 3.2). Section 4 concludes the paper and makes some proposals for further research.

1. A Hybrid Theory of Inference to the Best Explanation

In [2], the authors propose a hybrid theory for reasoning with arguments, stories and criminal evidence. The basic idea of this hybrid theory is that the propositions 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 arguments. For example, arguments can be used to support the story with evidence or to reason about the relative plausibility of the stories.

The formal hybrid theory $HT = (ET, CT)$ is 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: $r_i: p_1 \wedge \dots \wedge p_n \Rightarrow q$. Here r_i is the name of the rule, p_1, \dots, p_n and q are literals. The type of rule is indicated with a subscript: \Rightarrow_E denotes an evidential rule used in arguments and \Rightarrow_C denotes a causal rule used in explanations.

In the abductive theory $CT = (H, T, F)$, T is a set of causal rules, H is the set of hypotheticals, ground literals occurring in the antecedent of some rule in T and F is the set of explananda. The basic idea of abductive inference is that if we have a rule $cause \rightarrow effect$ in T and we observe *effect*, we are allowed to infer *cause* as a hypothetical explanation of the effect. Such an explanation can be a single proposition but it can

also be a causally connected story consisting of chains of causal rules. Thus, given F and T we need to abductively infer some hypotheticals such that together with T they explain F , that is, the explananda follow from the hypotheticals and the causal theory. More precisely, $S = H_i \cup T_i$, where $H_i \subseteq H$ and $T_i \subseteq T$, is an explanation for a set of explananda F iff for each $f \in F$: $S \vdash f$. 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 logic [10], which integrates ideas on rule-based argumentation and structured arguments (e.g. [12]) within [3]’s 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: 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.

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 $\{a_1, \dots, a_n\} \subseteq K_A$). Here, arguments about the plausibility of explanations are based on assumptions from K_A , as reasoning about plausibility is done using commonsense knowledge 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. In principle, the higher the evidential support and the lower evidential contradiction and implausibility, the better the explanation. This comparison of explanations can be further specified if desired: [2] define more

criteria for judging the quality of explanations, such as completeness (whether the story has all its required elements in order to be coherent), and [9] also provide other criteria such as the exclusivity and specificity of explanations. Furthermore, we could attach different weights to evidence so that particular pieces of evidence give a higher degree of support to explanations than others. In addition, we could also distinguish between explanations supported by defensible arguments and explanations supported by justified arguments. 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.

2. 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. [4] 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 ([4], 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, [12] define three kinds of burden of proof in terms of their framework for defeasible argumentation.³ 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 examined ([12] define it as the task of making sure there is a winning 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 at any point by one side does not meet this burden, the issue can be decided as a matter of law against this side, and that is the end of the trial.

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. [12] (p. 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

³ The burdens of claiming and contesting, which are also discussed by [12], assume an explicit dialogical context and will hence not be discussed here.

issues called scintilla of evidence, preponderance of evidence, clear and convincing evidence, and beyond reasonable doubt ([6], p. 241). The *scintilla of evidence* proof standard is met if “even the slightest amount of relevant evidence exists on an issue” ([5], p. 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” ([5], p. 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” ([5], p. 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 ([5], p. 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 ([13], p. 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 ([13], p. 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 [14] 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.

[6] 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 argument⁴ 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

⁴ Roughly, an argument is considered applicable if its premises are not defeated and there is no exception to the inference.

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.

3. Burden of Proof and IBE

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 [8], 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,⁵ 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 an evidential argument for the claim on which the burden rests. This claim may be, for example, an element of one's own explanation or the negation of an element of the opponent's explanation. The burden of production is essentially the same as in [12], because of the similarity between the hybrid theory's (evidential) argumentation theory and [12]'s framework. 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, dependant on the standard of proof.

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 [6], 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 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 [1]. However, a key difference is that in [1], 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

⁵ 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.

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 very 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 [8] 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.

3.1. Inference to the best explanation in civil trials

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 1 shows this explanation (and defendant’s alternative, see below). Here, \rightarrow arrows denote causal links, \dashrightarrow arrows denote evidential (argumentative) links and the $\dashrightarrow\bullet$ arrow denotes evidential contradiction. White boxes are part of the abductive theory CT (p or d denotes whether events are part of plaintiffs’ or defendant’s explanation) and gray boxes are part of the argumentation theory ET . Premises with italic text denote evidence in K_E .

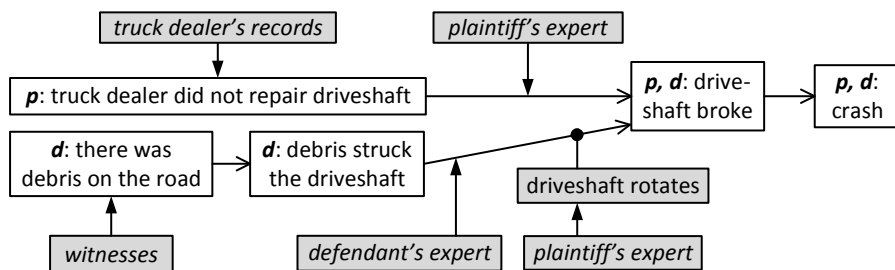


Figure 1: two explanations in the *Anderson v. Griffin* case.

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 seeing 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.

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). 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.

3.2. A Criminal Case with Two Competing Stories

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.

The prosecution 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. Thus, the prosecution gave an argument A_1 that Jackson knowingly (and hence intentionally) performed his act, using the generalization "if a person who is proficient with firearms shoots someone else at close range, the shooter intended to kill the victim" (more formally r_1 : x is proficient with firearms \wedge x shoots y at close range $\Rightarrow_E x$ intended to kill y).

The prosecution met the burden of persuasion and production 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-defense. 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 undercut r_1 , thus taking away the support this argument gave to 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. 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. Now S_p 's quality is restored (as A_1 again supports it) and S_j 's implausibility is demonstrated and thus the BRD standard is met, as is evident from the trial record.

4. Conclusion

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 [2]'s 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. [6] and [12] 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 [2] and [9] 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 [11] 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, [9]'s 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.

References

- [1] 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.
- [2] 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).
- [3] 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.
- [4] 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.
- [5] B.A. Garner, *Black's Law Dictionary* (9th ed.), Thomson Reuters, St Paul, Minn., 1990.
- [6] 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.
- [7] J.R. Josephson, On the proof dynamics of inference to the best explanation. *Cardozo Law Review*, **22** (2001), 1621-1643.
- [8] M.S. Pardo and R.J. Allen, Juridical Proof and the Best Explanation, *Law and Philosophy* **27** (2007), 223-268.
- [9] N. Pennington and R. Hastie, Reasoning in explanation-based decision making. *Cognition* **49:1-2** (1993), 123-163.
- [10] H. Prakken, An abstract framework for argumentation with structured arguments. *Argument and Computation*, **1:2** (2010).
- [11] H. Prakken and G. Sartor, Argument-based extended logic programming with defeasible priorities. *Journal of Applied Non-classical Logics* **7** (1997) 25-75.
- [12] 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.
- [13] J.W. Strong, *McCormick on Evidence* (4th ed.), West Publishing Co., St. Paul, Minn., 1992.
- [14] 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.