A FORMAL ARGUMENTATION EXERCISE ON THE KARADŽIĆ TRIAL JUDGMENT

FEDERICO CERUTTI
University of Brescia, Italy
`federico.cerutti@unibs.it`

YVONNE MCDERMOTT
Swansea University
`yvonne.mcdermottrees@swansea.ac.uk`

Abstract

We present the methodology and the results of an application of argumentation theory to map the evidence and arguments as to whether Radovan Karadžić, President of the Serb Republic, possessed the requisite *mens rea*—the knowledge of wrongdoing that constitutes part of a crime—for genocide in Srebrenica. To evaluate the strengths and weaknesses of Trial Chamber's findings in the publicly available judgment, we used argumentation-based techniques available in the CISpaces.org tool. The results of our analysis were submitted to the Appeals Chamber in the same case as an *amicus curiæ* brief, to assist the Appeals Chamber in its consideration of whether the Trial Chamber erred in finding that Karadžić possessed the requisite *mens rea*.

1 Introduction

In this paper—which is an extended version of [9]—we present the methodology and the results of an application of argumentation theory to map the evidence and arguments as to whether Karadžić possessed *mens rea*[1] for genocide in relation to the Srebrenica mass killing. The results of our analysis were submitted to the Mechanism for International Criminal Tribunals as an *amicus curiæ*[2] brief [19] pursuant to Rule 83 of the MICT Rules

---

[1]*Mens rea*: the intention or knowledge of wrongdoing that constitutes part of a crime. For the crime of genocide, it must be shown that the perpetrator intended to destroy, in whole or in part, a national, ethnic, racial, or religious group.

[2]*Amicus curiæ*: a non-party in a lawsuit who argues or presents information relevant to the lawsuit.

of Procedure and Evidence. We based our analysis only on the judgment of Prosecutor v. Radovan Karadžić [15].[3]

On 24th March 2016, Radovan Karadžić was convicted for genocide in Srebrenica by the International Criminal Tribunal for the former Yugoslavia (ICTY). As reported in [15], at least 5,115 men were killed by members of the Bosnian Serb Forces in July 1995 in Srebrenica (Section 3).

The Trial Chamber's finding that the accused possessed the *mens rea*—*i.e.*, the intention and knowledge of wrongdoing that constitutes part of a crime—for genocide in relation to the Srebrenica joint criminal enterprise (JCE) was the subject of academic critique at the time of the Trial judgment, *e.g.*, [26].

Using the argumentation-based techniques available in the CISpaces.org[4] tool [7], reviewed in Section 2, we manually analysed a sub-set of the 2615 pages of [15] to highlight the three reasoning lines that are present in the judgment and that lead to the conclusion that Karadžić possessed the requisite *mens rea*. Of those, two of them might merit further discussions, and the last one relies on a single witness.

Our main contribution is to show that the methodology we propose in Section 4 can be used to show the weakness and strengths of a case—cf. Section 6. This can be of use for the plaintiff, the defendant, but also judges and jurors, as it helps clarifying which elements are proven beyond any reasonable doubt, and which ones are not. This is currently a live issue in international criminal law: one of the authors of this paper argues that "each piece of evidence should be evaluated on its own merits, in light of the other evidence on the record, to determine whether a point has been proven beyond reasonable doubt," [18] as also supported by several judgments. The opposite is often argued, namely that the Trial Chambers should find their decision on the basis of the the accumulation of all the evidence in the case, but without the need to link factual and legal findings to the final decisions.

The submission of our *amicus curiæ* triggered reactions from the academic community interested in international criminal justice, practitioners at the United Nations courts of law, and media. We critically analyse our research and comment on its impact and related work in Section 7.

---

[3]In the following, we will heavily rely upon the judgment [15] as the only source of information for our analysis and paper.

[4]Although the project's name is CISpaces.org, it is still a research-grade prototype not yet stable enough to be released to the general public, hence **it is not accessible** at `https://cispaces.org`. However, the source **is available** at `https://github.com/cispaces` and a best-effort service is provided at `https://tiresia.unibs.it/cispaces/`.

2 Background

For this analysis, we used the tools available from the CISpaces project [28] and then further developed in its CISpaces.org version, introduced in [7], that rely on argumentation schemes and computational models of argumentation.

A fundamental concept in computational models of argumentation is the one of defeasible inference *rule*,[5] where a statement (*antecedent*) becomes a (*prima facie*) reason to believe another statement (*consequent*). For instance, "Mary, a witness, says that John committed the fraud" (antecedent) can be seen as a *prima facie* reason to believe that "John committed the fraud" (consequent).

Rules provide the building blocks for the notion of *argument*, that — borrowing from the ASPIC literature [24] — is iterative in the chaining of rules. Statements that are tentatively assumed to hold provide the base case for such an iteration, and thus they are defined as arguments having the statement itself both as premise and as conclusion, where *premises* and *conclusion* are two attributes of the notion of argument. The premises of arguments constructed using this base case also take the name of *ordinary premises* in our approach. Iteratively, an argument requires the existence of a rule whose antecedents are the conclusions of other arguments (*sub-arguments*), and, as a consequent, a statement that becomes the conclusion of this new argument, while its premises are the union of all the premises of its subarguments.

A statement is the contrary of another one when they cannot be both true, albeit they can both be false. A flexible way of using such a notion of *contrariness* [24] is by allowing for a statement to be the contrary of another one. By requiring the vice versa, the two statements would become *contradictory*. We will make use of such a flexibility in the following of our analysis.

The notion of *contrariness* between statements leads to the concept of *defeat* between arguments: an argument defeats another argument if the former *rebuts* or *undermines* the latter. When the conclusion of an argument contradicts the conclusion of another argument, it is the case that the first *rebuts* the second, as well as all the other arguments that have such a second argument as sub-argument. If, instead, the conclusion of an argument contradicts one of the ordinary premises of another one, then the former *undermines* the latter.

Given a set of arguments and defeats between them, we need criteria to assess which arguments collectively survive the defeats and thus can provide a reasonable viewpoint (or *extension*) based on the statements and the rules that we were considering. Such criteria usually consider *conflict-freeness*, *i.e.*, the absence of defeats within the viewpoint; *admissibility*, *i.e.*, if an argument in the viewpoint is defeated by a second argument, the latter must in turn be defeated by a third argument also in the viewpoint; and *maximality*, *i.e.*, a

---

[5]We will not make use of strict rules in this work.

viewpoint cannot be a strict subset of another viewpoint. Multiple viewpoints can exist for the same set of arguments and the defeats between them: two equally reliable witnesses, each providing one reason for contradictory conclusions, lead to the situation that each of the two arguments per se is a reasonable viewpoint, hence there are two of them. In this case, if an argument belongs to at least one viewpoint, it is said to be *credulously accepted*. If, instead, an argument belongs to all the viewpoints, it is said to be *sceptically accepted*. In the following we will be making use of this notion of sceptically acceptance in connection with the principle of *proof beyond a reasonable doubt*.

CISpaces.org provides a convenient visual language and an effective Human-Machine Interface for argumentation mapping. It builds on top of the Argument Interchange Format AIF [11] that specifies a graph structure composed of two types of nodes connected by links. The nodes can be either information (in the following identified by squared boxes) or scheme nodes (in the following identified by round boxes). Information nodes define the antecedents and consequents that we will be making use in the generation of arguments. Scheme nodes can be either rule of inference applications or conflict applications. A rule of inference application provides the connection between antecedents and a consequent: if one or more information nodes are linked to an inference node, and the latter is in turn linked to another information node, we will interpret this sub-graph as an inference rule. Conflict nodes, instead, express the contrariness relationship between two inference nodes: once again, links here are directed too.

2.1 Argumentation Schemes

Argumentation schemes [34] are abstract reasoning patterns commonly used in everyday conversational argumentation, legal, scientific argumentation, etc. Schemes have been derived from empirical studies of human arguments and debate. Each scheme has a set of critical questions that represent standard ways of critically probing into an argument to find aspects of it that are open to criticism. For instance, the following is the scheme for arguments from *evidence to hypothesis* [34]:

Major Premise: If A (a hypothesis) is true, then B (a proposition reporting an event) will be observed to be true.

Minor Premise: B has been observed to be true in a given instance.

Conclusion: Therefore, A is true.

CQ1: Is it the case that if A is true, then B is true?

CQ2: Has B been observed to be true?

CQ3: Could there be some reason why *B* is true, other than its being true because of *A* being true?

The other argumentation schemes used in this analysis are: the *abductive argumentation scheme*; the *argumentation from cause to effect*; the *argumentation from witness testimony*; and the *argumentation from (popular) opinion* [34].

An abductive argument aims at identifying a chain of inferences to fill in the gaps in the line of reasoning towards a given conclusion. It often involves identifying reasonable causes for a given outcome. It can be criticised on the basis of discussing alternative causes or on the actual explanatory power of the identified probable cause.

Connected to the previous scheme, an argument from cause to effect link two phenomena, *A* and *B*, in a possible causal link, hence stating that if *A* occurs, then *B* will (might) occur. This is also the main element for critique, namely how strong is such a causal generalisation?

Moving towards schemes widely used in trials, witness testimony is a strong argument when there is no direct access to the facts. In this case, to evaluate it one needs to rely upon comparison to other available evidence and evaluation of its consistency, both internal and external. It is worth mentioning that for this work we did not have access to the original set of testimonies as they are not available *verbatim* in the judgement.

An appeal to (popular) opinion may refer to just a majority in the cited reference group — *i.e.*, the court — . In general, the argument from popular opinion may be undermined under three aspects: the actual agreement of the majority with the proposition; the weakness of the argument itself when used to prove the truth of a proposition; and the link with the true opinion.

2.2 Charting Arguments, Mapping into ASPIC+, and Evaluating them

CISpaces.org [28, 7] enables a user to draw a directed graph representing an argument map, which can then be compiled into an ASPIC+ theory for automatic reasoning. In particular, an argument map is a directed graph (WDG $= \langle N, \rightsquigarrow \rangle$) based on the AIF format [11], thus with two distinct types of nodes: information nodes (or I-nodes) and scheme nodes (or S-nodes). S-nodes can be either rule of inference application (RA-nodes), or conflict application (CA-nodes), respectively represented as *Pro* and *Con* nodes. Pro links can be further labelled with the argumentation scheme they instantiate. In a WDG, nodes are connected by edges whose semantics are implicitly defined by their use [11].

Similarly to [24, 17], a WDG can be mapped into an ASPIC+ system [20]. Assume a logical language $\mathscr{L}$, and a set of *strict* or *defeasible* inference rules — resp. $\varphi_1, \ldots, \varphi_n \longrightarrow \varphi$ and $\varphi_1, \ldots, \varphi_n \Longrightarrow \varphi$. A strict rule inference always holds — *i.e.*, if the *antecedents*

$\varphi_1, \ldots, \varphi_n$ hold, the *consequent* φ holds as well — while a defeasible inference "usually" holds.

An *argumentation system* is as tuple $AS = \langle \mathcal{L}, \mathcal{R}, {}^-, v \rangle$ where:

- $^- : \mathcal{L} \mapsto 2^{\mathcal{L}}$ is a contrariness function s.t. if $\varphi \in \overline{\psi}$ and: $\psi \notin \overline{\varphi}$, then φ is a *contrary* of ψ; $\psi \in \overline{\varphi}$, then φ is a *contradictory* of ψ ($\varphi = -\psi$);

- $\mathcal{R} = \mathcal{R}_d \cup \mathcal{R}_s$ is a set of strict ($\mathcal{R}_s$) and defeasible ($\mathcal{R}_d$) inference rules such that $\mathcal{R}_d \cap \mathcal{R}_s = \emptyset$;

- $v : \mathcal{R}_d \mapsto \mathcal{L}$, is a partial function.[6]

A *knowledge base* $\mathcal{K}$ in an *AS* is a set of *axioms* $\mathcal{K}_n$ that cannot be attacked, and *ordinary premises* $\mathcal{K}_p$ that can be attacked, *i.e.*, $\mathcal{K}_n \cup \mathcal{K}_p = \mathcal{K} \subseteq \mathcal{L}$.

Building upon the notion of an argumentation system and of a knowledge base, an *argumentation theory* is a pair $AT = \langle AS, \mathcal{K} \rangle$.

To map a WDG into an ASPIC+ system, let us assume that:

- $P \subseteq N$ is the set of I-nodes, where each I-node in the graph is written p_i;

- ℓ_{type}, with $type = \{Pro, Con\}$, refers to a S-node;

- $[p_1, \ldots, p_n \rightsquigarrow \ell_{Pro} \rightsquigarrow p_\phi]$ indicates an inference rule, where $p_1, \ldots, p_q$ are parent nodes of the S-node ℓ_{Pro}, and p_ϕ is a child of ℓ_{Pro};

- conflict schemes can be either $[p_1 \rightsquigarrow \ell_{Con} \rightsquigarrow p_2]$ or $[p_1, \ldots, p_n \rightsquigarrow \ell_{Con} \rightsquigarrow p_\phi]$.

For this work, we make use of a subset of the ASPIC+ system: in particular, we will use neither strict rules nor preferences.

Given a WDG $= \langle N, \rightsquigarrow \rangle$, its corresponding ASPIC+ system $AS = \langle \mathcal{L}, \mathcal{R}, {}^-, v \rangle$ is such that:

- $\forall p \in P \subseteq N, p \in \mathcal{L}$;

- $\mathcal{R} = \mathcal{R}_s \cup \mathcal{R}_d$ with $\mathcal{R}_s = \emptyset$ and $\forall [p_1, \ldots, p_n \rightsquigarrow \ell_{Pro} \rightsquigarrow p_\phi]$, $p_1, \ldots, p_n \Rightarrow p_\phi \in \mathcal{R}_d$;

- $\forall [p_1 \rightsquigarrow \ell_{Con} \rightsquigarrow p_2]$, $p_1 \in \overline{p_2}$;

- $\forall [p_1, \ldots, p_n \rightsquigarrow \ell_{Con} \rightsquigarrow p_\phi]$, is mapped as $p_1, \ldots, p_n \Rightarrow p_h \in \mathcal{R}_d$ and $p_h \in \overline{p_\phi}$;

---

[6]Informally, $v(r)$ is a wff in $\mathcal{L}$ which says that the defeasible rule r is applicable. However, we will not make use of this feature in the following.

and the knowledge base $\mathscr{K}_n \cup \mathscr{K}_p = \mathscr{K} \subseteq \mathscr{L}$ is such that, given $[p_1, \ldots, p_n \rightsquigarrow \ell_{Pro} \rightsquigarrow p_\phi]$, $\forall p_i \in \{p_1, \ldots, p_n\}$, if p_i is not a conclusion of any inference rule $\nexists [\ell_{Pro} \rightsquigarrow p_i] \in \rightsquigarrow$, $p_i \in \mathscr{K}_p$. In addition, assume $\mathrm{WDG}' = \langle N', \rightsquigarrow' \rangle$ a subset of WDG—*i.e.*, such that $N' \subseteq N$ and $\rightsquigarrow' \subseteq \rightsquigarrow$—containing only a single cycle of inference schemes—*i.e.*, analogous to the case $p_i \Rightarrow p_i$—then $\forall p_i \in P' \subset N'$, if $[\ell_{Pro} \rightsquigarrow p_i], [p_i \rightsquigarrow \ell_{Pro}] \in \rightsquigarrow'$, then $p_i \in \mathscr{K}_p$ is an ordinary premise.

Following [20], an *argument* **a** on the basis of a $AT = \langle AS, \mathscr{K} \rangle$, $AS = \langle \mathscr{L}, \mathscr{R}, ^-, v \rangle$ is:

1. φ if $\varphi \in \mathscr{K}$ with: $\mathrm{Prem}(\mathbf{a}) = \{\varphi\}$; $\mathrm{Conc}(\mathbf{a}) = \varphi$; $\mathrm{Sub}(\mathbf{a}) = \{\varphi\}$; $\mathrm{Rules}(\mathbf{a}) = \mathrm{DefRules}(\mathbf{a}) = \emptyset$; $\mathrm{TopRule}(\mathbf{a}) = $ undefined.

2. $\mathbf{a}_1, \ldots, \mathbf{a}_n \longrightarrow / \Longrightarrow \psi$ if $\mathbf{a}_1, \ldots, \mathbf{a}_n$, with $n \geq 0$, are arguments such that there exists a strict/defeasible rule $r = \mathrm{Conc}(\mathbf{a}_1), \ldots, \mathrm{Conc}(\mathbf{a}_n) \longrightarrow / \Longrightarrow \psi \in \mathscr{R}_s / \mathscr{R}_d$. $\mathrm{Prem}(\mathbf{a}) = \bigcup_{i=1}^n \mathrm{Prem}(\mathbf{a}_i)$; $\mathrm{Conc}(\mathbf{a}) = \psi$; $\mathrm{Sub}(\mathbf{a}) = \bigcup_{i=1}^n \mathrm{Sub}(\mathbf{a}_i) \cup \{\mathbf{a}\}$; $\mathrm{Rules}(\mathbf{a}) = \bigcup_{i=1}^n \mathrm{Rules}(\mathbf{a}_i) \cup \{r\}$; $\mathrm{DefRules}(\mathbf{a}) = \{d \mid d \in \mathrm{Rules}(\mathbf{a}) \cap \mathscr{R}_d\}$; $\mathrm{TopRule}(\mathbf{a}) = r$

An argument can be attacked in its premises (*undermining*) or its conclusion (*rebuttal*). Since we will not use the preference ordering between arguments, we will omit it from the definition. Similarly for the notion of *undercut* on the inference rule (cf. [20]).

Given **a** and **b** arguments, **a** *defeats* **b** iff **a** *successfully rebuts* or *successfully undermines* **b**, where: **a** *successfully rebuts* **b** (on **b'**) iff $\mathrm{Conc}(\mathbf{a}) \notin \overline{\varphi}$ for some $\mathbf{b}' \in \mathrm{Sub}(\mathbf{b})$ of the form $\mathbf{b}''_1, \ldots, \mathbf{b}''_n \Longrightarrow -\varphi$; **a** *successfully undermines* **b** (on φ) iff $\mathrm{Conc}(\mathbf{a}) \notin \overline{\varphi}$, and $\varphi \in \mathrm{Prem}(\mathbf{b}) \cap \mathscr{K}_p$.

An *argumentation framework* (AF) [13] is a pair $\Delta = \langle \mathscr{A}, \rightarrow \rangle$ where $\mathscr{A}$ is a set of arguments[7] and $\rightarrow \subseteq \mathscr{A} \times \mathscr{A}$ is an attack relation. We denote with $\mathbf{a}_2 \rightarrow \mathbf{a}_1$ when $\langle \mathbf{a}_2, \mathbf{a}_1 \rangle \in \rightarrow$.

An AF $\langle \mathscr{A}, \rightarrow \rangle$ is the *abstract argumentation framework defined by* $AT = \langle AS, \mathscr{K} \rangle$, $AS = \langle \mathscr{L}, \mathscr{R}, ^-, v \rangle$ if $\mathscr{A}$ is the set of all finite arguments constructed from $\mathscr{K}$ (as above); and $\rightarrow$ is the defeat relation on $\mathscr{A}$.

Given an AF $\Delta = \langle \mathscr{A}, \rightarrow \rangle$: a set $S \subseteq \mathscr{A}$ is a *conflict-free* set of Δ if $\nexists \mathbf{a}_1, \mathbf{a}_2 \in S$ s.t. $\mathbf{a}_1 \rightarrow \mathbf{a}_2$; an argument $\mathbf{a}_1 \in \mathscr{A}$ is *acceptable* with respect to a set $S \subseteq \mathscr{A}$ of Δ if $\forall \mathbf{a}_2 \in \mathscr{A}$ s.t. $\mathbf{a}_2 \rightarrow \mathbf{a}_1$, $\exists \mathbf{a}_3 \in S$ s.t. $\mathbf{a}_3 \rightarrow \mathbf{a}_2$; a set $S \subseteq \mathscr{A}$ is an *admissible* set of Δ if S is a conflict-free set of Δ and every element of S is acceptable with respect to S.

A set of argument $S \subseteq \mathscr{A}$ is a *preferred extension* if and only if it is a maximal (with respect to set inclusion) admissible set.

An argument is *skeptically accepted* with regards to preferred semantics if and only if it belongs to *each* preferred extension. Checking this is a problem that lies at the second level of the polynomial hierarchy [14], hence the need—in general—for efficient implementations [10].

---

[7]In this paper we consider only *finite* sets of arguments: see [3] for a discussion on infinite sets of arguments.

6th	Shelling of Srebrenica began
11th, Afternoon	Srebrenica has fallen Karadžić appoints Deronjić as Civilian Commissioner for Srebrenica
11th, Night	A column of Bosnian Muslim men tried to escape by walking in a northwesterly direction towards the safe haven of Tuzla
12th, Morning	Shelling of the column began
12th, Afternoon	Large numbers of the members of the columns surrendered
13th, Morning	Groups of detainees from the column marched towards the Kravica Warehouse
13th, 1630h–Night Kravica Warehouse	One of the Bosnian Muslim detainees took away the rifle of a soldier and shot him dead: other soldiers started shooting at the detainees in response. Others shot at the detainees with machine guns and automatic rifles. Hand-grenades were thrown in the warehouse through the windows. By nightfall, between 755 and 1,016 Bosnian Muslim men were killed.
13th, 1700h–1840h Pale	Karadžić had an hour-long conversation on the phone during which he was briefed by General Maladić that Srebrenica "[wa]s done."
13th, 2010h	Intercepted call between Deronjić and Karadžić through an intermediary : Deronjić, the President is asking how many thousands? **D:** About two for the time being. […] : Deronjić, the President says: "All the goods must be placed inside the warehouses before twelve tomorrow." **D:** Right. : Deronjić, not in the warehouses over there, but somewhere else.
13th, around 2100h Bratunac SDS Office	Deronjić ordered to bury the detainees that had been killed at the Kravica Warehouse in a bauxite mine in Milići.
14th Just after midnight	Momir Nikolić drove Beara to the Bratunac SDS office, where Beara met with Deronjić and Vasić. Beara and Deronjić argued about where the Bosnian Muslim men were to be executed, as Beara insisted that he had instructions from his "boss" that the detainees were to remain in Bratunac, and Deronjić countered that the Accused had instructed him that all detainees in Bratunac should be transferred to Zvornik. Eventually, Beara and Deronjić agreed that the detainees would indeed be transferred to Zvornik. Detainees began to be transferred to the first of four detention sites in Zvornik.
14th, 1240h–1310h	Karadžić met with Deronjić alone.
14th, afternoon, after 1310h	Karadžić and Deronjić met with Srebrenica represetatives for about four hours.
14th, 2245h–2310h	Kovač met with Karadžić after touring Srebrenica, and the Bratunac and Zvornik areas on 13 and 14 July.
15th, 0035h–0125h	Bajagić—who has a substantive knowledge of the events in Srebrenica being the technical service procurement clerk—met with Karadžić.
16th	By the end of 16th July 1995, at least 3365 Bosnian Muslims men were killed.

Figure 1: Timeline of some of the most relevant events related to the Srebrenica mass killing. All dates refer to the month of July 1995.
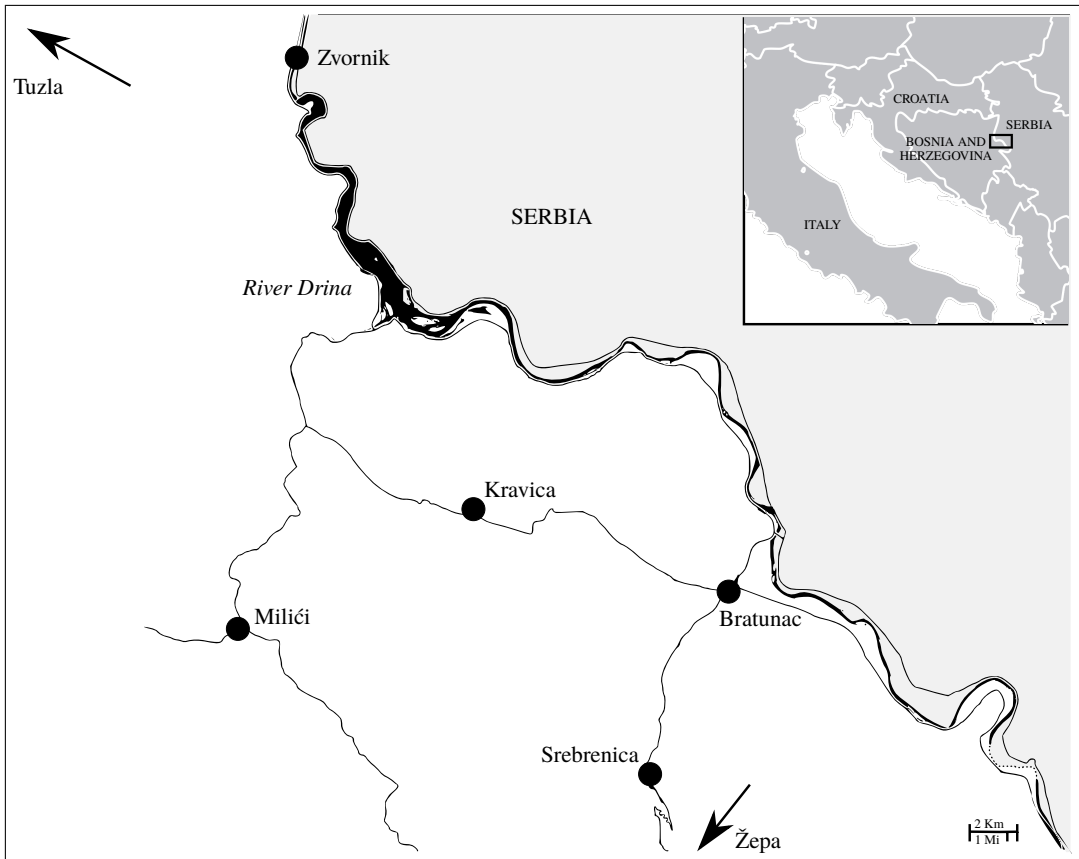
Figure 2: Relevant locations next to the Drina river. In white in the main chart the Socialist Republic of Bosnia and Herzegovina.

3 Karadžić and Srebrenica

What follows is a short historical summary of the events that lead to the Srebrenica massacre as reported in [15]. Figure 1 summarises the timeline of the most relevant events for our analysis starting from 6 July 1995.

The Socialist Republic of Bosnia and Herzegovina (SRBiH) was one of the six republics that once constituted the Socialist Federal Republic of Yugoslavia (SFRY): unlike the other republics, it possessed no single majority ethnic grouping. One of its political parties, the Serbian Democratice Party or SDS — led by Radovan Karadžić, campaigned to establish separate Serbian institutions. Following a plebiscite held on 9 and 10 November 1991, an autonomous Serb Republic (Republika Srpska) was proclaimed in 1992.

Among other key personnel within the Serb Republic, Radovan Karadžić served as President and Supreme Commander of the Bosnian Serb Army (VRS). Tomislav Kovač was the Assistant Minister of the Ministry of Intern (MUP), and the acting Ministry from September 1993 until January 1994. Ratko Mladić served as Commander of Main Staff, the highest operative body of the VRS. His assistant commander for Security Administration was Ljubiša Beara, with duties of management of the main staff of the Military Police, as well as co-ordinating with the bodies of the Ministry of the Interior. Momir Nikolić was Chief of the Security and Intelligence Organ, which was responsible for issues of security in the corps composing the VRS, including the arrest and detention of prisoners of war and other persons.

When in 1992 the population of Republic of Bosnia and Herzegovina voted for independence from the Socialist Federal Republic of Yugoslavia in a referendum, forces of the Serb Republic attacked different parts of the Republic of Bosnia and Herzegovina, whose state administration effectively ceased to function having lost control over the entire territory. The Assembly of the Serb Republic adopted the strategic goal to eliminate the border with Serbia: Srebrenica—a town with a majority of Bosnian Muslims—was close to that border (Figure 2).

In late June 1995, Karadžić gave a combat assignment that led to an offensive against Srebrenica and ultimately to the killing of at least 5,115 Bosnian Muslim men.

4   Methodology

The goal of our *amicus curiae* brief [19] was to identify the precise factual and inferential bases for the Trial Chamber's findings of Karadžić's genocidal intent in the Trial Chamber judgment, and to elucidate the forms of reasoning that led to these conclusions. We limited our analysis to the reasoning process that can be fathomed from the Trial Chamber's judgment. As such, we did not analyse issues such as the reliability of witnesses or evidence since they are the purview of the Trial Chamber alone, and also because the entire set of evidence used by the Trial Chamber is not publicly available.

In the present case, Karadžić's *mens rea* is an element of the offence of genocide in Srebrenica, as genocide requires each member of the joint criminal enterprise to be knowledgeable of the *dolus specialis* of the principal perpetrator. The *material facts* upon which proof of *mens rea* hinged were the Trial Chamber judgment's findings on: (1) Karadžić's knowledge of the expansion of the plan to remove Bosnian Muslims from Srebrenica to include the killing of men and boys, hence Karadžić sharing the intent to destroy the Bosnian Muslims in Srebrenica; and (2) his active involvement in the killings.

Following [21], we manually and in full agreement identified the arguments—and their general argumentation schemes when possible—that the Trial Chamber put forward in [15]

related to the two hypotheses (*Was Karadžićaware of the intent to kill the detainees?* and *Was Karadžićactively involved in the oversight of the killing of the men and boys?*, *cf.*, Section 5), together with (1) those instances of schemes for which not all critical questions have been satisfactorily addressed; (2) and particular facts that seem missing but necessary to expose the entire line of reasoning, labelling them with **Unstated**. In those cases, we did not include an analysis of critical questions for the inferences based on such unstated pieces of information, as it would be a detour from the purpose of this work.

It is worth noticing that the text proved resistant to attempts of automatic analysis. This is also evident in the graphical charting of our analysis (Figure 3), where we consider pieces of information spanning more than 210 pages (Para 5312 fn. 18025 [15, p. 2203] to Para 5808 [15, p. 2413]), in addition to historical information scattered around the entire document. For instance, the information that Srebrenica has fallen on 11 July 1995 has been presented in Para 5033 [15, p. 2079], 331 pages before being used in an argument to support the hypothesis.

5 Results

5.1 Was Karadžić aware of the intent to kill the detainees?

Figure 3 depicts our understanding of the reasoning lines that the Trial Chamber describes in its judgment [15] in support of the hypothesis that Karadžić was aware of the intent to kill the detainees. This is also the conclusion of the skeptically accepted arguments with regards to preferred semantics (cf. Section 2), quite unsurprisingly given the scarce number of conflicts: this is expected since the judgment does not record each exchange of arguments between the defence and the prosecution.

There are three main lines of reasoning in favour of this conclusion. The first one is based upon Nikolić's testimony that he overheard Deronjić saying that the accused had instructed Deronjić that all detainees should be transferred to Zvornik, cf. Figure 1, 14th July 1995, just after midnight. This testimony gives reasons to the chamber to refute the alternative explanation—highlighted by the defence and reported in the judgment—that Karadžić was referring to a place different from Zvornik in the intercepted conversation with Deronjić, cf. Figure 1, 13th July 1995, 2010h. In this line of reasoning, the Chamber decided also to link additional pieces of information (*Inference 3.cX* of Figure 3), as supporting the conclusion that Karadžić ordered that detainees should be transferred to Zvornik, such as a complaint to Beara by Deronjić about the presence of detainees in Bratunac. However, for those facts, it appeared that the Chamber did not consider some relevant critical question, *e.g.*, *Is there any other reasonable explanation for why Deronjić had previously complained to Beara about the detainees' presence in Bratunac, other than it being true because Karadžić conveyed to Deronjić the direction that the detainees should be transferred*
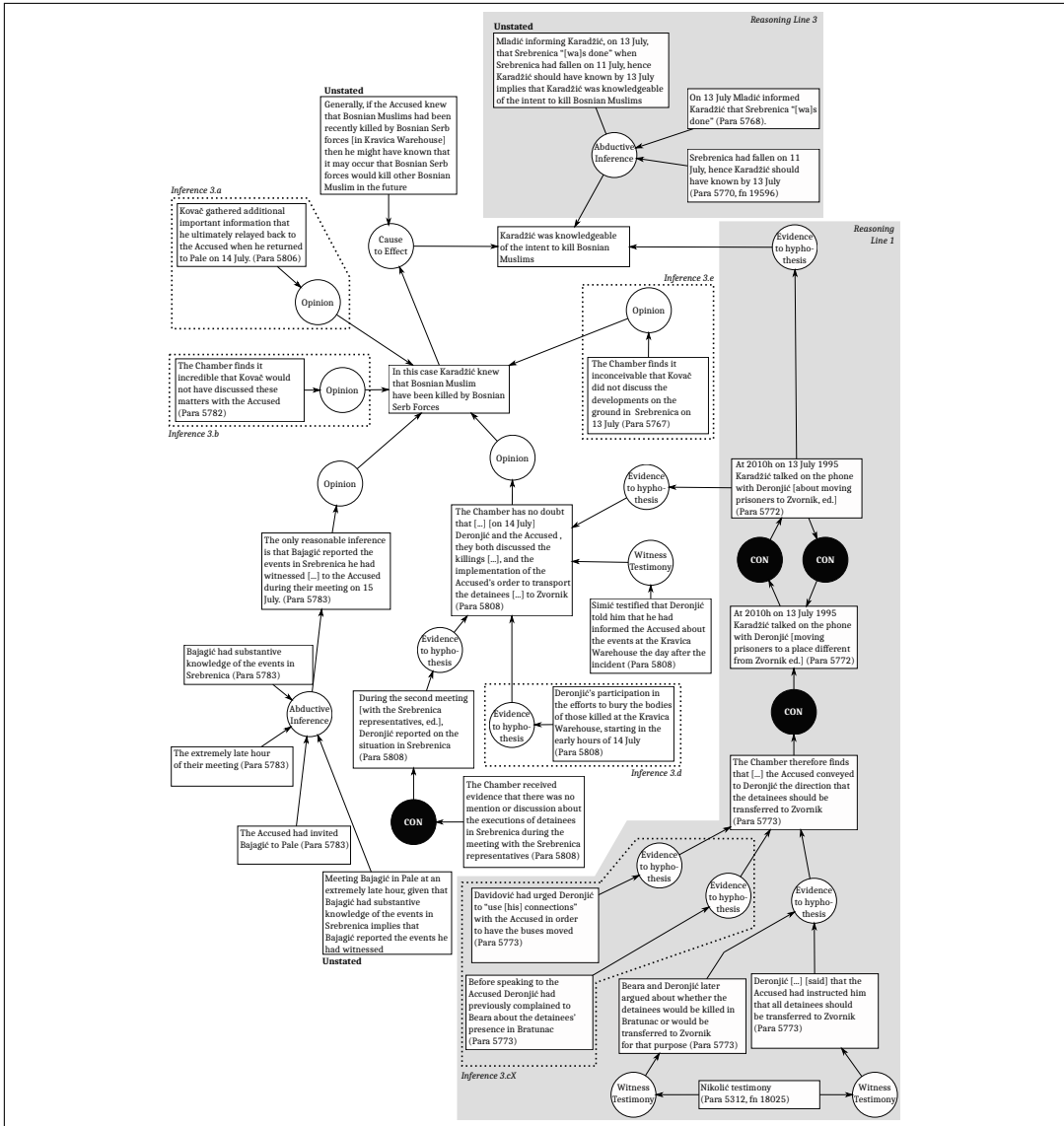
Figure 3: Analysis of arguments in [15] in favour of the hypothesis that Karadžić was knowledgeable of the intent of killing Bosnian Muslims men. Each "Para" reference refers to a paragraph of [15]. Names and events are introduced in Section 3, except for Milorad Davidović, who was a senior official in the MUP and, later on, a witness. Squared boxes are claims; white circles are *Pro* nodes, labelled with the argumentation schemes they refer to; while black circles are *Con* nodes. Dotted areas identify inferences for which there are critical questions that were not explicitly addressed in [15]. Three reasoning lines are highlighted as they are referred to in Section 5.1.

to Zvornik? Despite those additional pieces of information (*Inference 3.cX* of Figure 3), this line of reasoning does not rely on unstated findings or pieces of information for which critical questions have not explicitly been answered. It will be recalled that we methodologically chose not to assess the reliability of Nikolić's testimony as we did not have access to the entire trial records and besides, credibility and reliability are adjudged on a number of factors, including the witness's demeanour and/or evasiveness in the witness box, which would be difficult to determine from a transcript of proceedings [12].

A second line of reasoning justifying the hypothesis is based on Simić's testimony that Deronjić told him that he had informed Karadžić about the events at the Kravica Warehouse the day after the incident, in conjunction with the unstated assumption that if the accused knew that Bosnian Muslims had been recently killed by Bosnian Serb forces (in Kravica Warehouse), then he might have been known that it may occur that Bosnian Serb forces would kill other Bosnian Muslims in the future. It could be questioned whether all relevant critical questions find an answer in the judgment [15], with regard to *Inferences 3.a, 3.b, 3.d, 3.e* of Figure 3. For instance, *what evidence supported the finding that Kovač relayed back additional important information to Karadžić when he returned to Pale on 14 July?* (*Inference 3.a*, Figure 3); or *what evidence supported the finding that Kovač discussed these matters with Karadžić?* (*Inference 3.b*, Figure 3).

A third line of reasoning is based on an abductive inference with the unstated premise that Mladić informed Karadžić, on 13 July, that Srebrenica "[wa]s done." The Trial Chamber appears to have concluded that, given that Srebrenica had fallen on 11 July, Karadžić would have known this by 13 July. From that unstated inference, it drew a further inference that the conversation implied that Karadžić knew of the intent to kill the Bosnian Muslims of Srebrenica.

5.2 Was Karadžić actively involved in the oversight of the killing of the men and boys?

Figure 4 depicts our understanding of the Chamber's line of reasoning in concluding that the accused was actively involved in the oversight of the killing of the men and boys after the 13 July conversation. This is also the conclusion of the skeptically accepted arguments with regard to preferred semantics.

However, for each inference line supporting this conclusion, either necessary premises are unstated (hence left to the reader to assume), or at least one relevant critical question is not explicitly answered, namely:

- Regarding *Inference 4.a*: *Was there any other reasonable explanation for the statements that 'several thousand fighters did manage to get through' and 'we were not able to encircle the enemy and destroy them', other than that they were an illustration of regret that the corridor had been opened on 16 July?*
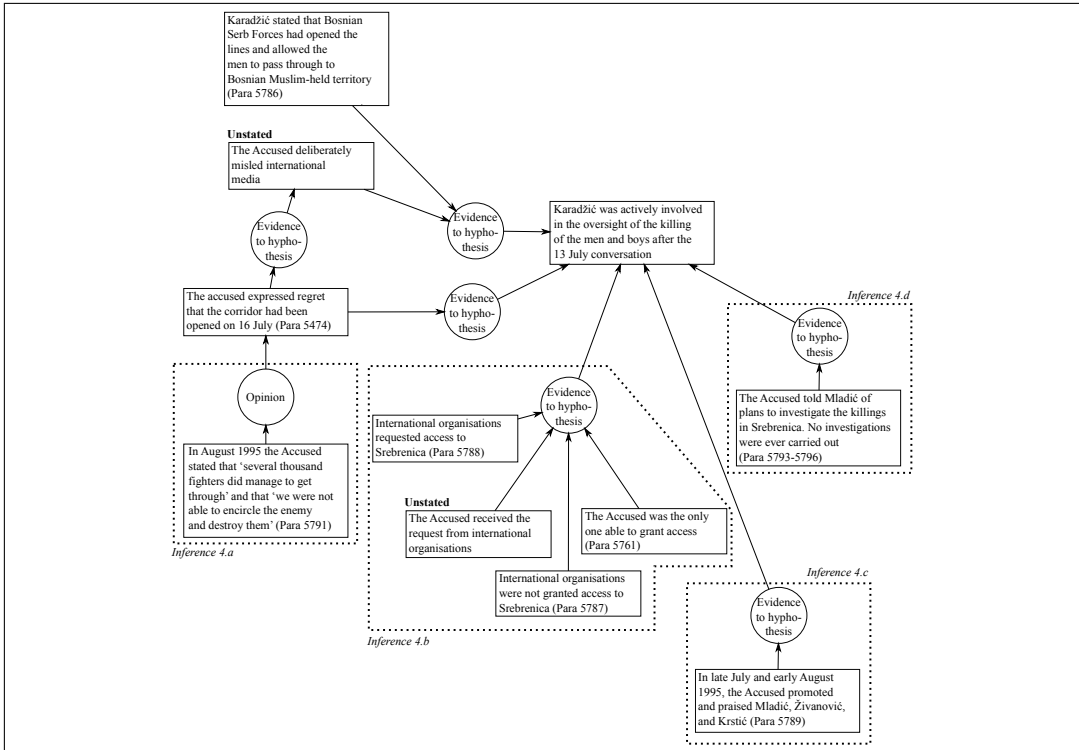
Figure 4: Analysis of arguments in [15] in favour of the hypothesis that Karadžić was actively involved in the oversight of the killing of the men and boys after the 13 July conversation. Each "Para" reference refers to a paragraph of [15]. Names and events are introduced in Section 3, except for Milorad Davidović, who was a senior official in the MUP and later on a witness. Squared boxes are claims; white circles are *Pro* nodes, labelled with the argumentation schemes they refer to; while black circles are *Con* nodes. Dotted areas identify inferences for which there are critical questions that were not explicitly addressed in [15]. Three reasoning lines are highlighted as they are referred to in Section 5.2.

- Regarding *Inference 4.b*:

    1. *Was it established as true that the accused received the request for access from international organisations?*

    2. *Was there any other reasonable explanation for why international organisations were not granted access to Srebrenica, other than this being true because the accused was actively involved in the oversight of the killings after the 13 July conversation?*

- Regarding *Inference 4.c*: *Was there any other reasonable explanation for why, in late July and early August 1995, the accused promoted and praised Mladić, Živanović, and Krstić, other than it being true because the accused was actively involved in the oversight of the killings after the 13 July conversation?*

- Regarding *Inference 4.d*: *Was there any other reasonable explanation for why no investigations were ever carried out, other than it being true because the accused was actively involved in the oversight of the killings after the 13 July conversation?*

6   Discussion

On 21 February 2018, we sent to the United Nations Mechanism for International Criminal Tribunals (MICT) a request for leave to make submissions as *amicus curiæ* pursuant to Rule 83 of the MICT Rules of Procedure and Evidence [19].

Given the overall results (cf. Section 5), we believed that our analysis was, on balance, probably more helpful to the prosecution than to the defence in the appeal, insofar as it illustrated that, while some inferential steps could have been explicated in greater detail, the Trial Chamber's reasoning was broadly sound.

This is clearly not the first attempt to apply formal argumentation to judicial findings. In [31], Verheij introduces the notion of *automated argument assistance* which is in spirit very close to our work here, as it explicitly aims at drafting and testing of court pleadings. Walton [35] provides an extensive account of argumentation in legal systems, and argues for a *new method* for legal argumentation which, among others, includes the use of "argument diagramming to map out the network of inference in a given case" [35, p. 323]. For completeness of discussion, authoritative colleagues criticise the use of argument diagramming, notably van Gelder in [30]. In reflecting on his experience, he notices that argument diagramming might not serve well the purpose of deliberation, possibly because deliberation is a dialectical activity rich in nuances. However, he did not consider deliberation activities where the incentives for a proper epistemic investigation are significant, such as writing a judgment for an international criminal case.

The work by Walton on legal argumentation [35] and in general on argumentation schemes — summarised in [34] — motivated researchers in deriving computational models, thus building on the tradition initiated by Verheij [31]. Bex et.al. [6] expanded on the idea of using argumentation schemes for providing a formal account of reasoning, and subsequently in [5] they also considered the advantages of merging it with storytelling. The latter also takes into consideration the different positions of the plaintiff and the defendant, which is also the case of [23] — where a formal dialogue system is used as a formalisation tool — and [22], where ASPIC+ is used for formalising legal case-based reasoning. In contrast to previous approaches, we considered explicitly the role of argumentative semantics

using skeptical acceptance according to preferred semantics as a proxy for the *beyond any reasonable doubt* standard of proof. This is clearly questionable, but it looks a reasonable approximation as it is a rather conservative choice, although it might be a little difficult to explain to non-experts. Further analysis using other semantics are already planned for future work, as well as a deeper comparison with the ANGELIC methodology [1], in particular after the recent paper [2] showing a correspondence with ASPIC+.

We also feel that there is very little we can add to van Gelder's observations in [29], where he analyses some legal arguments. His comments strongly resonate with us, as we also experienced "little use of verbal indicators of logical structure, and often use obscure or vague indicators" [29]. We also encountered incomplete arguments, with text scattered across the document, and possibly serving multiple purposes.[8] Although far from providing an off-the-shelf support tool, in retrospect our analysis would have significantly benefit from (1) entity-relations extractors, *e.g.*, [27] and (2) topic modelling system, *e.g.*, [36], which together might transform a static PDF document into a database that can be queried.

In contrast to previous approaches, we considered a case under discussion at ICTY offering the results of our analysis as an *amicus curiæ* brief to the Appeal Chambers. It unfortunately denied admissibility of our application on 28 March 2018, observing that "the issues regarding whether Karadžić possessed the *mens rea* for genocide in relation to the Srebrenica JCE were extensively litigated before the Trial Chamber and have been fully briefed by Karadžić and the Prosecution on appeal." The Appeals Chamber also seems to criticise the fact that "the *Amicus Curiæ* Observations seek to guide the Appeals Chamber's analysis of the Trial Judgment without consideration of or access to the entire record that is relevant to the Trial Chamber's conclusions." This however would raise the question: what is the purpose of having a 2615 page judgment, if the judgment does not actually fully reflect the grounds for the conclusion? Finally, confirmation that our analysis was trustworthy comes from the Appeal Chamber Judgment [16] that in its section D.2 provides a summary of the Trial's Chamber Judgment regarding whether Karadžić was knowledgeable of the intent to kill Bosnian Muslims which is almost entirely present in our resulting argumentation network depicted in Figure 3.

7   Conclusion

In this paper, we presented the methodology and the results of an application of argumentation theory to map the evidence and arguments as to whether Karadžić possessed *mens rea* for genocide in Srebrenica based on [15]. As discussed in Section 5, we summarised

---

[8]A reviewer of an earlier version of this paper commented that some of the instances of argument from opinion in Figure 3 seem more of instance of argument from ignorance. This is something that only a judge mindful of the purpose of their prose could clarify when writing the document.

the results of our analysis testing whether Karadžić was knowledgeable of the intent — of General Mladić and others — to kill Bosnian Muslims. This hypothesis is supported within the Trial Chamber's judgment [15] by three lines of reasoning, two of which might merit further discussion, and the last one relying on a single witness.

Although at first sight this paper seems to be similar to other attempts to analyse legal reasoning with formal argumentation, *e.g.*, [23, 22], it differs from them substantially as we did not try to capture the debate component, hence distinguishing between Prosecutor and Defence claims. Instead, our analysis is closer to works aimed at analysing arguments in a single document, like, for instance, [21] that analyses the role of argumentation in written financial communications.

Although the Appeals Chamber denied the admissibility of our application, the interest that applying formal argumentation theories triggered in the international criminal law community suggests that there is scope for future work in this area. We cannot claim that the methodology used in this analysis is beyond critique, but we can claim that it can help creating a better judgment that fully reflects the grounds for the overall conclusion.

This is the long-term aspiration of the ongoing research underpinning this paper, and we are fully aware that this will require to provide answers and innovative proposals both from a technical perspective as well as from the legal one. From a technical perspective, for instance, we still lack appropriate methodologies for adequately transforming statements of natural language into formal logic — a problem most students of logic encounter without being presented with satisfactory solutions, cf. among others [4] — thus inevitably exposing the subjectivity of each formalisation. In addition, following [8], we will also work in the direction of assessing the quality and the strengths of different argumentation reasoning lines, by taking into consideration quantitative measurements of uncertainty and trust, thus enriching the community proposals looking at probabilistic elements in legal reasoning, *e.g.*, [25, 33, 32].

Acknowledgments

of Defence or the U.K. Government. The U.S. and U.K. Governments are authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation hereon.

References

[1] Latifa Al-Abdulkarim, Katie Atkinson, and Trevor J. M. Bench-Capon. A methodology for designing systems to reason with legal cases using abstract dialectical frameworks. *Artificial Intelligence and Law*, 24(1):1–49, 2016.

[2] Katie Atkinson and Trevor J. M. Bench-Capon. Relating the ANGELIC methodology and ASPIC+. In *COMMA 2018*, pages 109–116, 2018.

[3] Pietro Baroni, Federico Cerutti, Paul E. Dunne, and Massimiliano Giacomin. Automata for Infinite Argumentation Structures. *Artificial Intelligence*, 203(0):104–150, 2013.

[4] Michael Baumgartner and Timm Lampert. Adequate formalization. *Synthese*, 164(1):93–115, 2008.

[5] F. Bex, S. 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(1-4):145–168, oct 2007.

[6] Floris Bex, Henry Prakken, Chris Reed, and Douglas Walton. Towards a Formal Account of Reasoning about Evidence: Argumentation Schemes and Generalisations. *Artificial Intelligence and Law*, 11(2/3):125–165, 2003.

[7] Federico Cerutti, Timothy J. Norman, Alice Toniolo, and Stuart E. Middleton. Cispaces.org: from fact extraction to report generation. In *COMMA 2018*, pages 269 – 280, 2018.

[8] Federico Cerutti and Gavin Pearson. Supporting scientific enquiry with uncertain sources. In *21st International Conference on Information Fusion*, pages 404–411, 2018.

[9] Federico Cerutti and Yvonne McDermott Rees. Did karadzic possess the mens rea for genocide in srebrenica? In Marcello D'Agostino, Fabio Aurelio D'Asaro, and Costanza Larese, editors, *Proceedings of the 5th Workshop on Advances in Argumentation in Artificial Intelligence 2021 co-located with the 20th International Conference of the Italian Association for Artificial Intelligence (AIxIA 2021), Milan, Italy, November 29th, 2021*, volume 3086 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2021.

[10] Federico Cerutti, Mauro Vallati, and Massimiliano Giacomin. An Efficient Java-Based Solver for Abstract Argumentation Frameworks: jArgSemSAT. *International Journal on Artificial Intelligence Tools*, 26(02):1750002, 2017.

[11] Carlos Iván Chesnevar, Jarred McGinnis, Sanjay Modgil, Iyad Rahwan, Chris Reed, Guillermo R. Simari, Matthew South, Gerard A. W. Vreeswijk, and Steven Willmot. Towards an argument interchange format. *The Knowledge Engineering Review*, 21(04):293, 2006.

[12] Gabrielė Chlevickaitė, Barbora Holá, and Catrien Bijleveld. Judicial witness assessments at the icty, ictr and icc: Is there 'standard practice'in international criminal justice? *Journal of International Criminal Justice*, 18(1):185–210, 2020.

[13] Phan Minh Dung. On the Acceptability of Arguments and Its Fundamental Role in Nonmonotonic Reasoning, Logic Programming, and n-Person Games. *Artificial Intelligence*, 77(2):321–357, 1995.

[14] Paul E. Dunne and Michael Wooldridge. Complexity of abstract argumentation. In *Argumentation in AI*, chapter 5, pages 85–104. Springer-Verlag, 2009.

[15] ICTY. Prosecutor v. Karadžić. Judgement. `http://www.icty.org/x/cases/karadzic/tjug/en/160324_judgement.pdf`, 2016.

[16] ICTY. Prosecutor v. Karadžić: Appeal Judgement. MICT-13-55-0660/2, 2019.

[17] Mathilde Janier, John Lawrence, and Chris Reed. OVA+: An argument analysis interface. In *COMMA 2014*, pages 463–564, 2014.

[18] Yvonne McDermott. Inferential reasoning and proof in international criminal trials: The potentials of wigmorean analysis. *Journal of International Criminal Justice*, 13(3):507–533, 2015.

[19] Yvonne McDermott Rees and Federico Cerutti. Request for leave to make submissions as amicus curiae. `http://jrad.unmict.org/webdrawer/webdrawer.dll/webdrawer/rec/240941/view/`, 2018.

[20] Sanjay Modgil and Henry Prakken. A general account of argumentation with preferences. *Artificial Intelligence*, 195:361–397, 2013.

[21] Rudi Palmieri, Andrea Rocci, and Nadzeya Kudrautsava. Argumentation in earnings conference calls. corporate standpoints and analysts' challenges. *Studies in Communication Sciences*, 15(1):120 – 132, 2015.

[22] H. Prakken, A. Wyner, T. Bench-Capon, and K. Atkinson. A formalization of argumentation schemes for legal case-based reasoning in ASPIC+. *Journal of Logic and Computation*, 25(5):1141–1166, 2015.

[23] Henry Prakken. Formalising ordinary legal disputes: a case study. *Artificial Intelligence and Law*, 16(4):333–359, 2008.

[24] Henry Prakken. An abstract framework for argumentation with structured arguments. *Argument & Computation*, 1(2):93–124, June 2010.

[25] Henry Prakken. Argument schemes for discussing Bayesian modellings of complex criminal cases. In *JURIX 2017*, pages 69–78, 2017.

[26] Milena Sterio. The Karadžić Genocide Conviction: Inferences, Intent, and the Necessity to Redefine Genocide. *Emory International Law Review*, jan 2017.

[27] Mihai Surdeanu, David McClosky, Mason R. Smith, Andrey Gusev, and Christopher D. Manning. Customizing an information extraction system to a new domain. In *ACL2011*, pages 2–10, 2011.

[28] Alice Toniolo, Timothy J. Norman, Anthony Etuk, Federico Cerutti, Robin Wentao Ouyang, Mani Srivastava, Nir Oren, Timothy Dropps, John A. Allen, and Paul Sullivan. Agent Support to Reasoning with Different Types of Evidence in Intelligence Analysis. In *AAMAS 2015*, pages 781–789, 2015.

[29] Tim van Gelder. Why are legal arguments to hard to follow? `https://timvangelder.com/2009/08/13/why-are-legal-arguments-so-hard-to-follow/` (on 23

Aug 2018), 2009.

[30] Tim van Gelder. Cultivating Deliberation for Democracy. *Journal of Public Deliberation*, 8(1), apr 2012.

[31] Bart Verheij and Bart. Automated argument assistance for lawyers. In *ICAIL '99*, pages 43–52, New York, New York, USA, 1999.

[32] Bart Verheij, Floris Bex, Sjoerd T. Timmer, Charlotte S. Vlek, John-Jules Ch. Meyer, Silja Renooij, and Henry Prakken. Arguments, scenarios and probabilities: connections between three normative frameworks for evidential reasoning. *Law, Probability and Risk*, 15(1):35–70, mar 2016.

[33] Charlotte S. Vlek, Henry Prakken, Silja Renooij, and Bart Verheij. A method for explaining Bayesian networks for legal evidence with scenarios. *Artificial Intelligence and Law*, 24(3):285–324, sep 2016.

[34] D Walton, C Reed, and F Macagno. *Argumentation schemes*. Cambridge University Press, NY, 2008.

[35] Douglas N Walton. *Legal argumentation and evidence*. Pennsylvania State University Press, 2002.

[36] Hai Zhuge and Lei He. Automatic maintenance of category hierarchy. *Future Generation Computer Systems*, 67:1 – 12, 2017.