CENTRE DE RECHERCHE
SUR LES STRATÉGIES ÉCONOMIQUES

How is Digital Evidence Used in the International Criminal Court? A Theoretical and Empirical Approach

Yann Lecorps, Khaoula Naili, Marie Obidzinski,
Yves Oytana, Téa Toutounji

July 2025

Working paper No. 2025 – 09

CRESE
30, avenue de l'Observatoire
25009 Besançon
France
http://crese.univ-fcomte.fr/

UNIVERSITÉ MARIE & LOUIS PASTEUR

How is Digital Evidence Used in the International Criminal Court? A Theoretical and Empirical Approach*

Yann Lecorps† Khaoula Naili‡ Marie Obidzinski§ Yves Oytana¶
Téa Toutounji‖

July 1, 2025

Abstract

We examine the use of digital evidence in cases handled by the International Criminal Court (ICC) at different stages of the proceedings, both theoretically and empirically, and how the parties use it. Our theoretical findings indicate that the extent to which digital evidence is used *versus* classical evidence may increase or decrease with the stringency of the standard of proof. This variation depends on the cost of gathering evidence and the degree of complementarity between digital and classical evidence. Our main empirical findings are as follows: i) the intensity of references to classical evidence increases more than that of references to digital evidence between the pretrial and trial phases; ii) the prosecution appears to rely more on digital evidence than the defense; iii) there is a positive correlation between the emotional tone of the prosecution and the defense, but no correlation between their emotional tone and the reference to either types of evidence.

Keywords : International Criminal Court, digital evidence, standard of proof, social media, textual analysis, sentiment analysis

JEL classification code : K4

*Research carried out with the support of the Institut des Etudes et de la Recherche sur le Droit et la Justice. We thank Sarah Jamal and Olivier Schaub for their insightful comments on the paper. All remaining errors are ours.

†Paris Panthéon Assas Université, CRED, 75014 Paris, France. e-mail: yann.lecorps@assas-universite.fr

‡Université Marie et Louis Pasteur, CRESE (UR 3190), F-25000 Besançon, France. e-mail: khaoula.naili@univ-fcomte.fr

§Paris Panthéon Assas Université, CRED, 75014 Paris, France. e-mail: marie.obidzinski@assas-universite.fr

¶Université Marie et Louis Pasteur, CRESE (UR 3190), F-25000 Besançon, France. e-mail: yves.oytana@univ-fcomte.fr

‖Paris Panthéon Assas Université, CRED, 75014 Paris, France. e-mail: tea.toutounji@u-paris2.fr

1 Introduction

In international crime investigations, the use of digital evidence can be crucial, especially when investigators cannot access to the territory where the crimes allegedly are took place. Indeed, individuals, whether they are witnesses, victims, or even perpetrators, can post pictures or videos on the internet (Aronson, 2018). The situation in Syria illustrates this point particularly well. Although Syria closed its borders to United Nations investigators in March 2011, the Independent International Commission of Inquiry on the Syrian Arab Republic was able to establish facts of human rights violations. This was only possible because the Commission received numerous videos posted on social networks.[1] However, the probative value of such information collected on the internet is often questioned in criminal courts (Laux, 2018). Indeed, using digital evidence poses specific challenges, such as the difficulty of formally identifying and authenticating the source of the information and the risk that the information is incorrect (Hellwig, 2021).

In this article, we use a positive approach to examine, both theoretically and empirically, the use of digital evidence *and* its interaction with classical evidence (*e.g.*, witness testimonies) in cases handled by the International Criminal Court (hereafter, ICC). Specifically, we aim to shed light on the following: **i)** the intensity of the references to digital evidence at different stage of proceedings; **ii)** the link between the standard of proof and reference to digital vs classical evidence; and finally **iii)** whether parties (in particular, prosecution and defense) involved in the trial proceedings use and present differently each types of evidence.

Our theoretical model primarily relies on the economics of judicial proof literature, issuing from Posner (1999). Specifically, the model relates to the literature on evidence production (*e.g.*, Froeb and Kobayashi, 2001, Emons and Fluet, 2009), as well as to the literature on the standard of proof (*e.g.*, Demougin and Fluet, 2008, Mungan, 2011, Rizzolli and Saraceno, 2013, Obidzinski and Oytana, 2019). However, rather than focusing on what the procedural rules should be, we describe how prosecutors choose the amount and type of evidence depending on the level of the standard of proof. The usefulness of digital evidence hinges on its admissibility in court and its probative value, keeping in mind that digital evidence is only corroborative (*i.e.*, this type of evidence is only used to strengthen or confirm existing evidence). Our theoretical results show that: **i)** as the standard of proof becomes more stringent, the amount of both digital and classical evidence

[1] See, for example, the "human rights channel" on YouTube. Source: `https://www.youtube.com/humanrights`.

produced by the prosecutor increases; **ii)** the extent to which each type of evidence increases depends on its relative marginal cost of collection and the extent to which the two types of evidence are complementary. Specifically, if classical evidence is more expensive (or harder to obtain) than digital evidence, and if the degree of complementarity between the two types of evidence is high enough, then, as the standard becomes more stringent, the additional amount of digital evidence collected by the prosecutor will tend to increase relatively more than the amount of classical evidence (and *vice versa*).

Our empirical analysis uses natural language processing (NLP) methods, and more specifically textual analysis, to examine 521 hearing transcripts from 8 cases involving crimes committed since 2010 that were handled by the ICC. We draw on recent literature applying NLP methods to extract quantitative information from legal texts (see Ash et al. (2022) for an application of sentiment analysis to court decisions, and Frankenreiter and Livermore (2020) for a review of the literature). First, we test the predictions of our theoretical model regarding the intensity of the use of digital evidence according to the standard of proof. Next, we focus on 148 hearing transcripts from the Al Hassan case to measure the intensity with which the parties (prosecution, defense, victims, and judges) used digital evidence in the trial, as well as whether this intensity correlates with a particular sentiment. Our main empirical findings from all 521 hearing transcripts reveal that: **i)** During trials, references to classical evidence are more frequent than references to digital evidence (the latter almost only occurs when there is also a reference to classical evidence, suggesting that digital evidence is only corroborative); **ii)** the intensity of references to both digital and classical evidence increases between the pretrial and trial phases as the standard of proof increases; **iii)** the intensity of references to classical evidence increases more than that of references to digital evidence between the pretrial and trial phases. Our main empirical findings from the 148 hearing transcripts in the Al Hassan case are as follows: **i)** the prosecution appears to rely more on digital evidence than the defense does; **ii)** sentiment analysis shows that the emotional variability of victims' representatives is higher than that of the other parties (judges, prosecution and defense); **iii)** there is no correlation between the emotional tone of the defense or prosecution and the use of each type of evidence, but there is a positive correlation between references to classical evidence and judges, and also between references to digital evidence and victims' representatives.

The rest of the paper is organized as follows. Section 2 presents our database construction method and some descriptive statistics. Section 3 theoretically investigates how the use of digital *versus*

classical evidence is affected by the stringency of the standard of proof. Section 4 examines the frequency of references to digital and classical evidence at the pretrial and trial stages. Section 5 conducts a textual analysis of transcripts from the Al Hassan case to examine how each party (prosecution, victims' representatives, defense, and judges) uses both types of evidence. Section 6 concludes.

2 Database Construction and Descriptive Statistics

Established in 2002, the International Criminal Court is a court of last resort that aims to deter international crimes, including crimes against humanity, genocide, war crimes, and crimes of aggression.[2] A case may be brought before the International Court in three main ways: i) at the request of a State Party to the Rome Statute; ii) at the request of the United Nations Security Council; iii) the Prosecutor may open an investigation on his or her own initiative. Criminal proceedings are then organized into four main stages, in addition to a possible appeal stage and enforcement of the sentence:

1. **Preliminary examination.** The Office of the Prosecutor must determine if there is sufficient evidence that a crime has been committed and if it is serious enough to fall within the ICC's jurisdiction.

2. **Investigations.** The Office of the Prosecutor collects evidence and identifies a suspect.

3. **Pretrial.** During the confirmation of charges hearing, which occurs after the initial appearance, three pretrial judges hear from the prosecution, the defense, and the legal representatives of the victims. Then, they decide if there is *enough evidence* for the case to go to trial.

4. **Trial.** In front of three trial judges, the Office of the Prosecutor must prove the guilt of the accused *beyond reasonable doubt*.

Note that the standard of proof[3] used in the trial phase (*beyond reasonable doubt*) is more stringent than that used in the pretrial phase (*enough evidence*). In the following, we will focus on the pretrial

[2]The Court's jurisdiction only comes into play when the state responsible for prosecuting the perpetrator fails to do so, that is, when national courts are unable or unwilling to do so. The ICC has jurisdiction in the 124 countries of the world that have ratified the Rome Statute, its founding statute. For a list of States Parties to the Rome Statute, see `https://asp.icc-cpi.int/states-parties`. The Rome Statute is available here: `https://www.icc-cpi.int/sites/default/files/RS-Eng.pdf`.

[3]The standard of proof is "the level of certainty the adjudicative authority has to reach in order to establish guilt

and trial phases.

We collected a database of transcripts from the ICC pretrial and trial hearings.[4] These documents can be found online in the ICC legal tools database. Information was added for each document, such as the name of the accused, the country where the crime was committed, and the year the crime began. The sample is restricted to ICC-managed cases involving crimes committed since 2010.[5] Using these criteria, we obtain a sample of 521 hearing transcripts.

Table 1: Transcripts by stage of proceedings

	n	%
Pretrial	32	6.1
Trial	489	93.9
Total	521	100

Table 1 shows that 93.9% of the hearing transcripts are from the trial phase. This is consistent with the trial phase being significantly longer than the pretrial phase and involving a higher standard of proof.

in a criminal proceeding" (Rizzolli, 2016). For example, the *beyond reasonable doubt* standard can be interpreted as a probability that the defendant is guilty that is greater than 95%. The requirement of *enough evidence* is closer to the *preponderance of evidence* or the *clear and convincing evidence* standards, which are roughly equivalent to the 50% and 75% probability standards, respectively.

[4]Document collection ended on July 15, 2023. Transcripts of hearings posted online by the ICC after this date are not included in the sample.

[5]In 2010, according to the World Bank, individuals using the internet represents 28% of the population. In 2024, 68% Source: World Telecommunication/ICT Indicators Database.

Table 2: Transcripts by case

Accused	Country	n	%
Al Hassan	Mali	145	27.8
Al Mahdi	Mali	7	1.3
Gaddafi Saif	Libya	3	0.6
Gbagbo et Blé Goudé	Ivory Coast	221	42.4
Gbagbo Laurent	Ivory Coast	4	0.8
Mokom	Central African Republic	4	0.8
Said	Central African Republic	8	1.5
Yekatom	Central African Republic	129	24.8
Total		521	100

As shown in Table 2, the transcripts come from a limited number of cases, and almost all of them relate to crimes that began in 2010, 2012 and 2013.[6] The Al Hassan (Mali), Gbagbo and Blé Goudé (Ivory Coast), and Yekatom (Central African Republic) cases account for more than 90% of the transcripts. There are at least two reasons for this small number of cases. First, the ICC is a court of last resort, and its jurisdiction to intervene is limited by the Rome Statute. Second, the ICC has a limited budget, whereas international criminal justice is costly to implement (Wippman, 2006).

Regarding the use of classical and digital evidence in hearing transcripts, we defined a list of keywords, including their lexical variants identified using the Thesaurus API.[7] Note that classical evidence refers here exclusively to witness testimonies while digital evidence refers to information obtained through the internet and the use of digital devices.[8] In a first step, all transcripts have been converted to lower case and cleaned to remove extraneous characters that could interfere with

[6] The number of transcripts by year of crime is shown in Table 8, in the Appendix.

[7] The list of keywords related to classical (digital) evidence can be found in Table 9 (Table 10) in the Appendix. Lexical variants and synonyms were identified using the Thesaurus API from API Ninjas: `https://api-ninjas.com/api/thesaurus`.

[8] Fact-finding missions, such as the International, Impartial and Independent Mechanism (the IIIM, see `https://iiim.un.org/`), aim to gather all information, including from the internet. They emphasize the distinction between open source, internet sources, and information coming from social media.

Figure 1: Digital evidence wordcloud

word frequency calculations. Once cleaned, the total word count for each transcription was computed, providing a baseline for assessing the relative frequency of specific keywords. Finally, we counted the occurrence of keywords related to digital and classical evidence allowing comparisons of their frequency between stages of the proceedings (section 4) and between parties involved in the case (section 5).

With respect to classical evidence, the most common word is "witness", with an average of 29.4 occurrences per hearing transcript, while the word "testimony" appears 6.7 times per transcript on average. Overall, 97% of the hearing transcripts contain at least one of the three keywords related to classical evidence.[9] With respect to digital evidence, we used a list of keywords that correspond to a broad view of the estimated frequency of references to evidence from online social networks and the internet. Figure 1 is a word cloud generated from these keywords, where the size of the words increases with their frequency.[10] The word "video" is the most common, with an average of 16.3 references per document, while some of the words, particularly those related to online social networking, are not cited in any hearing transcript.[11] Not surprisingly, we find that the three online social networks with the highest average number of citations are Facebook, YouTube, and Twitter.

[9]See Table 11 in the Appendix for the number and percentage of references to classical evidence in the transcripts.

[10]The words "video" and "footage" appear twice because these terms are counted separately in addition to the term "video footage".

[11]See Table 10 in the Appendix.

Table 3: Frequency of digital and classical evidence references

	Ref digital evidence	No digital evidence	Sum
Ref classical evidence	80.2%	16.5%	96.7%
No ref classical evidence	0.6%	2.7%	3.3%
Sum	80.8%	19.2%	100%

Comparing the references to both categories of keywords, we find that the proportion of transcripts containing references to digital evidence(80.8%) is slightly lower than for classical evidence (96.7%).[12] According to Table 3, 80.2% of pretrial and trial transcripts contain at least one reference to both digital and classical evidence. Almost no transcripts (0.6%) contain at least one reference to digital evidence and none to classical evidence. Conversely, 16.5% of transcripts have at least one reference to classical evidence and none to digital evidence.[13] These observations may be explained by the corroborative nature of digital evidence.

3 A Model of Evidence Collection

In this section, we provide a simple model to study how the prosecutor's office decision to allocate its time and effort between the collection of classical *versus* digital evidence is affected by the stringency of the standard of proof. Our main assumption is that digital evidence has only a corroborative value (*i.e.*, digital evidence has not, by itself, a strong enough probative value to satisfy the standard of proof).

3.1 The Basic Framework

There are two types of evidence: classical (in quantity C) and digital (in quantity D). Let us denote $f(C, D)$ the probative value of a level C (D) of classical (digital) evidence. We assume $f_C(C, D) > 0$, $f_D(C, D) > 0$, $f_{CC}(C, D) < 0$, $f_{DD}(C, D) < 0$: the probative value of evidence increases with the

[12] See Tables 11 and 12.

[13] A chi-squared test of independence applied to the contingency Table (3) unsurprisingly yields a p-value very close to zero: the hypothesis of independence between references to digital and classical evidence can be rejected.

quantities of both types of evidence, but at a decreasing rate.[14] Additionally, we make the following assumptions:

i. $f_C(C,D) > f_D(C,D)$ if $C = D$.

ii. $f(0,D) = 0$ and $f(C,0) \geq 0$.

iii. $f_{CD}(C,D) > 0$.

Assumption (i) encapsulates the idea that classical evidence has a higher marginal probative value than digital evidence. Assumption (ii) means that, in the absence of classical evidence, digital evidence has no probative value (while classical evidence, even when it is used on its own, may have some). Finally, assumption (iii) means that the two types of evidence are complements, in the sense that increasing the quantity of one type of evidence increases the marginal probative value of the other type of evidence. Taken together, assumptions (ii) and (iii) refer to the corroborative nature of digital evidence.

Collecting evidence is costly. We denote by α (β) the marginal cost of gathering a unit of classical (digital) evidence. We assume that the prosecutor aims at minimizing the total cost of producing evidence, while still satisfying the standard of proof, with $\bar{x}$ the level (or stringency) of that standard. Therefore, the problem faced by the prosecutor is:

$$\begin{cases} \min\limits_{C,D} & \alpha C + \beta D \\ \text{subject to} & f(C,D) \geq \overline{x} \end{cases} \tag{1}$$

Solving (1) shows that the optimal levels of classical evidence (C^*) and digital evidence (D^*) are characterized by:[15]

$$\begin{cases} \frac{\frac{\partial f}{\partial C}(C^*,D^*)}{\frac{\partial f}{\partial D}(C^*,D^*)} & = \frac{\alpha}{\beta} \\ f(C^*,D^*) & = \overline{x} \end{cases} \tag{2}$$

The first equation of (2), together with assumption (i), imply that *if the cost of gathering each type of evidence is the same (i.e. $\alpha = \beta$), then the optimal level of classical evidence is higher than the optimal level of digital evidence.* The second equation, together with assumption (ii), imply that *the prosecutor cannot satisfy the standard of proof by using only digital evidence: it is never optimal to obtain digital evidence in the absence of classical evidence.*

[14] $f(C,D)$ can be interpreted as a production function, with the inputs being C and D.

[15] The details to solve this program and for the results given in subsection 3.1 are given in the appendix.

3.2 The Effect of a More Stringent Standard of Proof on Evidence Collection

In the following, we study the effect of a change in the level of the standard of proof ($\overline{x}$) on C^* and D^* (indeed, C^* and D^* are functions of $\overline{x}$, although we do not make that explicit in order to streamline the notation).

Result 1. *When the standard of proof becomes more stringent, the quantities of both digital and classical evidence that the prosecutor gathers increase.*

Note that this result implies that the total cost of gathering evidence ($\alpha C^* + \beta D^*$) increases with the standard.

Does an increase in the standard of proof increases relatively more the gathering of classical evidence, or conversely?

Result 2. *For a small increase in the standard, the quantity of classical evidence increases more than the quantity of digital evidence if:*

$$\beta \frac{\partial^2 f}{\partial C^2}(C^*, D^*) - \alpha \frac{\partial^2 f}{\partial D^2}(C^*, D^*) + (\beta - \alpha) \frac{\partial^2 f}{\partial C \partial D}(C^*, D^*) > 0 \tag{3}$$

Condition (3) may or may not be satisfied. Thus, *when the standard of proof rises, the quantity of classical evidence may increase more than the quantity of digital evidence, or conversely.* More specifically, whether that condition holds depends on three effects (corresponding to the three terms on the left-hand side of (3)).

The first effect (the first term in (3)) is negative, in the sense that the stronger is this effect, the more the prosecutor tends to favor the collect of digital rather than classical evidence. It is related to the decreasing rate with which the probative value of classical evidence increases with its quantity (*i.e.* the assumption $f_{CC}(C, D) < 0$). The second effect (the second term in (3)) is similar to the first one, except that it is positive and related to the decrease in the marginal probative value of digital evidence. Taken together, these two effects mean that, *when the standard of proof increases, if the probative value of an additional unit of classical evidence decreases sharply with its quantity (when compared to the decrease in the probative value of an additional unit of digital evidence), then*

the prosecutor may want to intensify relatively more his effort to gather digital rather than classical evidence (and conversely).

The third effect (the third term in (3)) relates to the degree of complementarity between the two types of evidence. To illustrate the intuition behind this effect, let us assume that the prosecutor wants to increase the probative value originating from classical evidence to reach a more stringent standard of proof,[16] and that the marginal cost of gathering classical evidence is *higher* than the cost for gathering digital evidence ($\alpha > \beta$). In this specific case, the prosecutor has basically two choices: (i) increasing the quantity of classical evidence to *directly* increase its probative value, or (ii) increasing the quantity of digital evidence, thus increasing *indirectly* the probative value of classical evidence due to the complementarity of the two types of evidence (increasing the quantity of digital evidence reinforce the probative value of each unit of classical evidence already gathered). Because the cost of a unit of classical evidence is greater than the cost of a unit of digital evidence, the prosecutor may lean toward option (ii) (that is, increasing the quantity of digital evidence), because it is cheaper to do so rather than choosing option (i) (investing directly in gathering more classical evidence). This effect is stronger when the degree of complementarity increases (*ceteris paribus*). Moreover, note that when the marginal costs are the same for each type of evidence ($\alpha = \beta$), that third effect cancels out, while it works in the opposite direction when the marginal cost to produce evidence is relatively *lower* for classical evidence ($\alpha < \beta$). To summarize, *when the standard of proof increases, if the marginal cost of classical evidence is higher (lower) than the marginal cost of digital evidence, the more the two types of evidence complement each other, the more (the less) the prosecutor will intensify his effort to gather digital evidence (when compared to his effort to gather classical evidence).*

4 Use of Digital and Classical Evidence Between the Pretrial and Trial Stages

In this section, we examine the frequency of references to digital and classical evidence at the pretrial and trial stages. During the pretrial phase, judges hear from the prosecution, the defense, and the legal representatives of the victims. Then, they decide whether there is *enough evidence* for the case

[16] This is a simplification for illustration purpose, since in our model, the probative values of each type of evidence are always linked through the "production" function f.

to go to trial. During the trial, the prosecution must prove the defendant's guilt *beyond reasonable doubt*.

Our goal is to determine whether the theoretical results obtained in the preceding section regarding the effect of the standard of proof on evidence collection are empirically supported. Given the strengthening of the standard of proof between the two trial phases, result 1 suggests that references to both types of evidence should increase. However, according to result 2, the effect of strengthening the standard of proof on the relative increase of each type of evidence is less straightforward. This depends on the marginal cost of collecting each type of evidence and the extent to which the two types of evidence complement each other.

Table 4: Proportion of transcripts with (any) reference to digital/classical evidence, by stage of proceedings

	Classical evidence	Digital evidence	Ratio classic/digital
Pretrial	78.1%	56.2%	1.39
Trial	98%	82.4%	1.19
Pretrial and Trial	96.7%	80.8%	1.2

Table 5: Average reference rate per 1000 words by proceeding stage

	Classical evidence	Digital evidence	Ratio classic/digital
Pretrial stage	1.12	0.66	1.69
Trial stage	3.48	2.04	1.71
Pretrial and trial	3.34	1.95	1,71

There are more references to both types of evidence at the trial stage. As shown in Table 4, the proportion of hearing transcripts containing references to both types of evidence increases between

the pretrial and trial phases. We have a similar result when we look at the average reference rate per 1000 words to classical and digital evidence in Table 5: the average reference rate increases for both digital and classical evidence. This is consistent with our theoretical result 1 that as the standard of proof becomes more stringent (as is the case from pretrial to trial in ICC cases), the amount of both types of evidence increases.

References to classical evidence outnumber references to digital evidence at both the pretrial and trial stages. Specifically, Table 4 shows that the number of transcripts with at least one reference to classical evidence is 39.0% higher than the number of transcripts with at least one reference to digital evidence at pretrial and 19.1% higher at trial. In addition, Table 5 shows that the average reference rate per 1000 words to classical evidence is 69.7% higher than the average reference rate to digital evidence at pretrial and 71% higher at trial.[17] This observation provides (indirect) support for the assumption that classical evidence has a higher marginal probative value and that the cost of collecting classical evidence should not be too high compared to the cost of collecting digital evidence.

Between the pretrial and trial stages, the comparison of growth rates of references to digital versus classical evidence is ambiguous. Starting with the number of transcripts with at least one reference to digital and classical evidence (Table 4), we find that between pretrial and trial, the number of transcripts with at least one reference to digital evidence increases faster than the number of transcripts with at least one reference to classical evidence (47% increase for transcripts with references to digital evidence, *versus* 25% for classical evidence). Although this result suggests that the growth rate of references to digital evidence is higher, this conclusion is mitigated when we consider the average reference rates per 1000 words to digital and classical evidence (Table 5). In fact, from Table 5, we see that between pretrial and trial, there is a comparable increase by a factor 3 in the average reference rate to digital and classical evidence. Therefore, it is unclear whether the growth rate of references to digital evidence is higher or similar to that of classical evidence.

One limitation of the previous analysis is that it does not distinguish between parties. The following section attempts to distinguish between the words of each party in the transcripts (specifically,

[17] Of course, we should be cautious about interpreting the number of references in our keyword list as a proxy for the intensity with which each type of evidence is used in hearings. Among other limitations, we may have missed some keywords related to digital or classical evidence. In addition, we deliberately excluded some keywords because they were too ambiguous (their use could in some cases be unrelated to references to evidence, or could be used for both digital and classical evidence).

members of the prosecution and defense) in order to highlight each party's strategy regarding the use of digital and classical evidence.

5 Use of Digital and Classical Evidence by Parties

This section examines whether members of the prosecution and defense attorneys use and present digital and classical evidence differently. Indeed, the two parties have different roles during the trial. As the party bearing the burden of proof (Kaplow, 2012, Talley, 2013), the prosecutor must provide the court with compelling evidence that the accused committed a crime. Specifically, the prosecutor must convince the judges "beyond any reasonable doubt". Consequently, the prosecutor may be open to more diverse evidence than the defense in order to meet this burden of proof.[18] Second, members of the prosecution can be viewed as "repeat players", while the defense is closer to a "one-shot player". These parties are thus expected to "play the litigation game differently" (Galanter, 1974). In our case, this suggests that the prosecution and defense may have different strategies when using digital evidence. Based on these considerations, we hypothesize that prosecutors are more likely to rely on digital evidence than defense attorneys. We explore this issue in subsection 5.1.

Furthermore, the interaction between judges, victims representatives, defense attorneys, and prosecutors during a trial involves an exchange that can elicit some emotional states. The analysis presented in subsection (5.2) aims to highlight the emotional tone of the parties and whether emotions are more or less present when referring to digital and classical evidence. Following the distinction made by Jouannet (2007), the presence of emotions, if proven, may be a sign that the evidence is being used to "persuade" (*i.e.*, as part of an argumentative, deliberative logic) rather than to "demonstrate" (*i.e.*, as part of a formal, deductive legal logic). If the first possibility is true (*i.e.*, evidence is used to persuade), we may expect stronger emotional reactions. According to Jouannet (2007), "this does not mean that we no longer resort to reasoned observation (...) to establish proof of facts, but that this logic of establishment passes through a confrontation with others and through a logic of argumentation-persuasion that predominates."

For this analysis, we will focus on a recent ICC case, the Al Hassan case, which is named after

[18] At the ICC, there is no rule of evidence that sets out *ex ante* the methods of proof and their probative value. The judges are free to assess the probative value of the evidence.

the accused.[19] The alleged crimes occurred between January 2012 and January 2013.[20] The trial started in July 2020 and ended in May 2023. Al Hassan was found guilty of most of the crimes he was charged with on June 26, 2024.[21] We have collected 148 transcripts of the pretrial and trial hearings in this case.[22]

An ICC trial consists of three successive phases: (i) a first phase during which the office of the prosecutor presents its evidence; (ii) a second, very short phase during which the victims' "views and concerns" are presented; (iii) and finally a third phase for the presentation of the defense's evidence. During phase (i), the office of the prosecutor will refer to the incriminating digital or classical evidence he has collected. However, the defense is not passive and may cross-examine the prosecution's evidence in accordance with the adversarial principle. The same applies to phase (iii): the defense will present the exculpatory evidence it has, and the prosecution will then have the opportunity to cross-examine the exculpatory evidence to challenge it. With respect to phase (ii), victims' representatives may present evidence only with the express agreement of the Trial Chamber. In such cases, the evidence in question may be cross-examined by both the defense and the prosecution.

5.1 References to Digital Versus Classical Evidence by Parties

In this subsection, we conduct a textual analysis of the transcripts of the Al Hassan case to determine the frequency of words related to digital and classical evidence in relation to each party to the trial. In order to do this, we compiled a list of all the speakers present at each hearing, along with their respective roles within the court proceedings. This list allows us to distinguish between

[19] The other case on which we could have focused on, given the large number of transcripts, is that of Gbagbo and Blé Goudé. However, the Gbagbo and Blé Goudé cases were initially separate cases, with separate transcripts of the preliminary phases.

[20] Al Hassan Ag Abdoul Aziz was suspected of having committed crimes against humanity and war crimes in Timbuktu, Mali, between April 1, 2012 and January 28, 2013, in the context of an attack by armed groups Ansar Eddine/Al-Qaeda in the Islamic Maghreb. On March 16, 2013, the ICC Prosecutor concluded that there are reasonable grounds to believe that crimes within the jurisdiction of the ICC have been committed in Mali and decided to open an investigation. The arrest warrant was issued in March 2018 and the pretrial phase took place in July 2019.

[21] On June 26, 2024, Al Hassan was convicted of the crimes against humanity of torture, persecution, and other inhumane acts; and the war crimes of torture, outrages upon personal dignity, and mutilation. However, he was acquitted of the crimes against humanity of rape, sexual slavery and other inhumane acts in the form of forced marriage; and the war crimes of rape, sexual slavery and attacking protected objects. (Source: ICC)

[22] All pretrial and trial transcripts have been collected. However, two transcripts were excluded from our database later on because they were in French.

members of the prosecution team, the defense team, and the victims' representatives. In our analysis, witnesses called by the prosecution (or defense, victims' representatives) are considered part of the prosecution team (or defense, victims' representatives). We then used NLP tools to search for predefined keywords (see Tables 9 and 10) for each party. Further methodological details are provided in Appendix D.

Table 6: Total number of words in pretrial and Trial by phase - Al Hassan case

Pretrial	Trial			Total
	Phase 1 Office of the prosecutor	Phase 2 Victims representatives	Phase 3 Defense	
87,742 (12%)	440,503 (59.95%)	22,215 (3.02%)	184,377 (25.09%)	734,837 (100%)

Figure 2: Distribution of words by participant for each phase

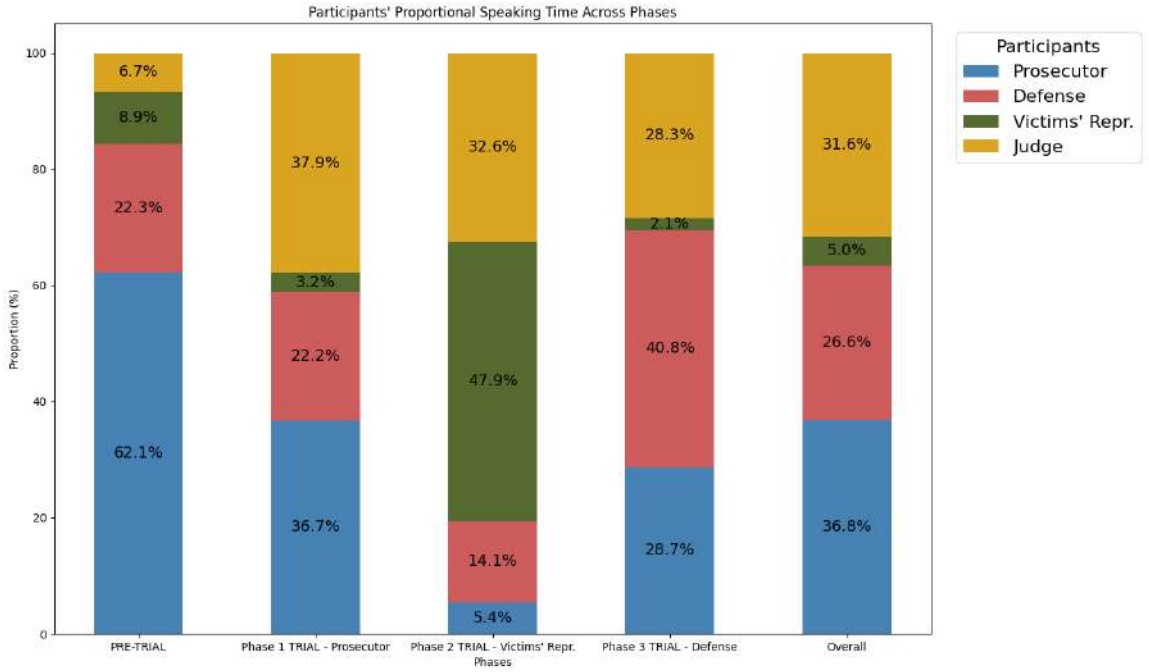
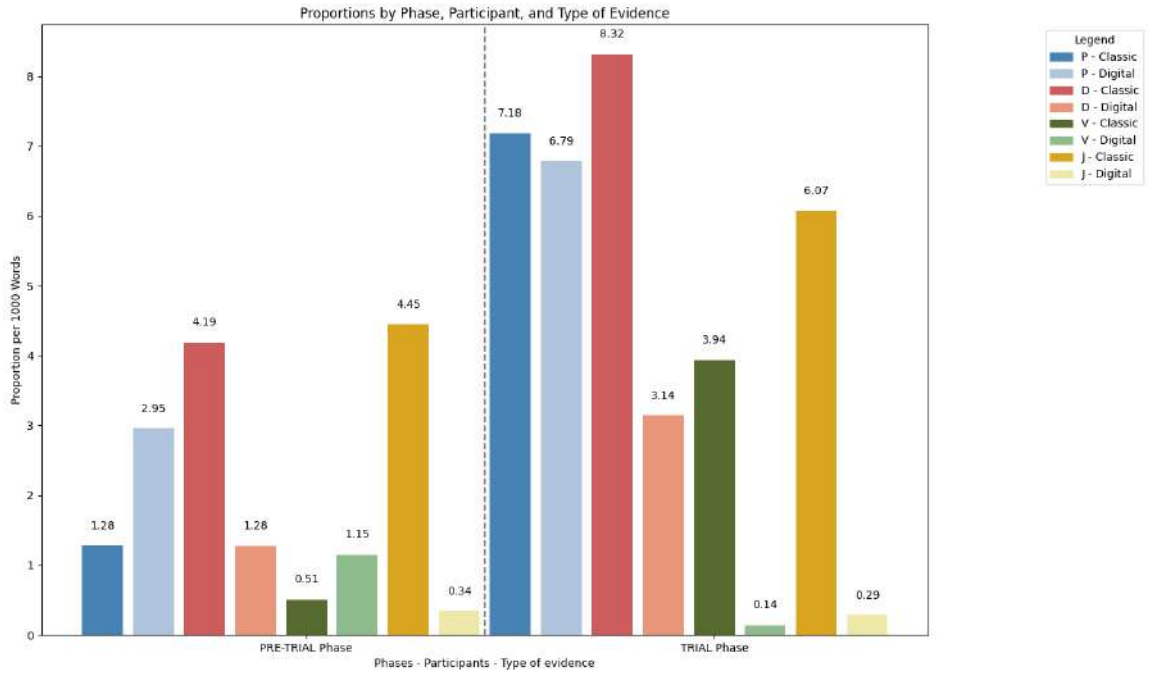


Table 6 provides an overview of the relative size (in terms of word counts) of the different phases.

Figure 2 shows the distribution of words by each participant and phase. Overall, the members of the prosecution speak the most (36.8% of the total number of words), followed by the judge (31.6% of the words) and the defense (26.6% of the words). Victims' representatives account for only 5% of the words. In the pretrial phase, the prosecutor dominates with 62.1% of the words. During the trial, the most prominent speakers vary depending on the phase, reflecting the organization of speech during a trial as described above.[23] The judge, who plays a relatively minor role in the pretrial phase, becomes more prominent during the trial.

Figure 3: References (per 1000 words) to digital and classical evidence by participant (pretrial and trial)



[23]In Phase 1, the members of the prosecution are the most prolific speakers with 36.7% of the words, while it is the victims' representatives in Phase 2 (47.9% of the words) and the defense in Phase 3 (40.8% of the words). The details of the number of words of each participant in the Al Hassan case for each phase are given in Appendix C.

Figure 4: References (per 1000 words) to digital and classical evidence by participant (trial phases)

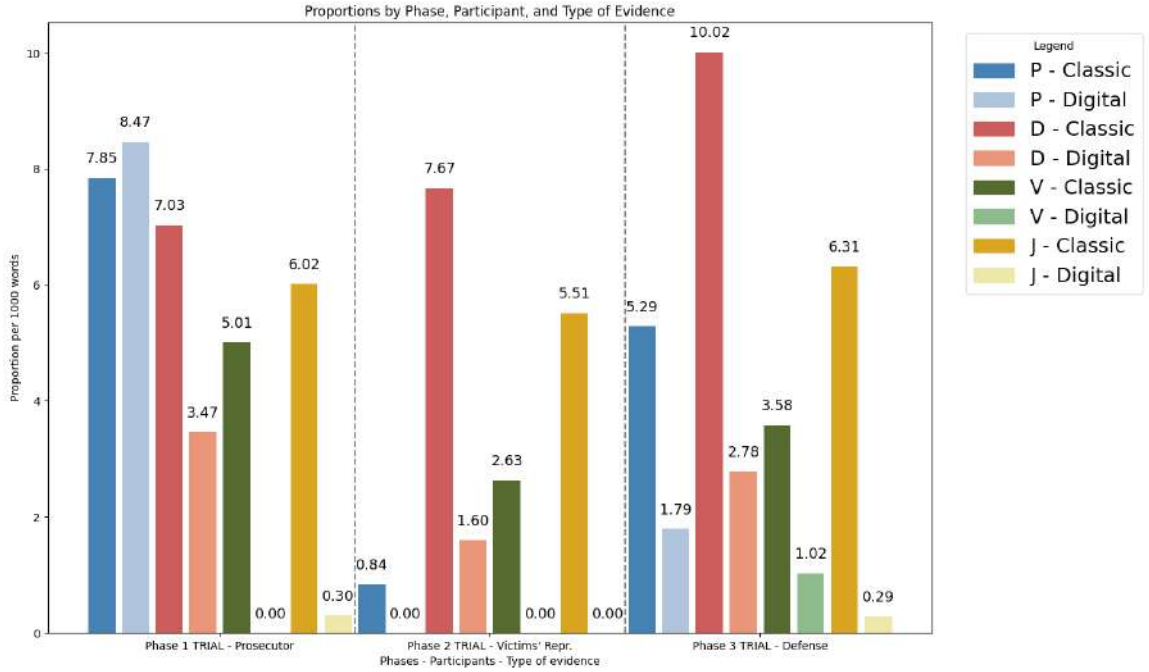


Figure 3 displays the proportion of references to digital and classical evidence made by each party during the pretrial and trial phases. Figure 4 shows the proportions for each subphase of the trial. In both figures, the proportions are calculated by dividing the number of references to digital or classical evidence made by a party in a given phase by the total number of words spoken by that party in the same phase, then multiplying by 1000.[24] Globally, these figures suggest that the defense and the prosecution adopt different strategies regarding the use of digital *versus* classical evidence.

Figure 3 shows that, at the pretrial stage (where judges must decide whether there is "enough evidence" for the case to go to trial), the proportion of references to classical evidence made by the prosecutor (1.28‰) is lower than the proportion of references to digital evidence (2.95‰). The opposite is true for defense attorneys. In comparison, at the trial stage (where judges must decide

[24] For example, the proportion of references to classical evidence by the prosecution in the pretrial phase is calculated as follows:

$$\frac{\text{number of references to classical evidence made by the prosecution in pretrial}}{\text{number of words used by the prosecution in pretrial}} \times 1000$$

according to a *beyond reasonable doubt* standard), the proportion of words referring to classical evidence is higher than the proportion of words referring to digital evidence for each party. However, even at this stage, the prosecution relies more heavily on digital evidence than the defense does. The defense's preference for classical evidence appears to be a recurring pattern, as the number of references to classical evidence per 1000 words is consistently higher than the number of references to digital evidence.

Figure 4 confirms this insight. For the prosecution, the proportion of words referring to classical evidence is lower than the proportion referring to digital evidence in Phase 1, when the prosecution presents evidence. In phase 3, when the defense presents evidence, the proportion of references to classical evidence by the prosecutor (5.29‰) becomes higher than that to digital evidence (1.79‰). However, this is likely due to the fact that the debates focus on the evidence presented by the defense.

The differences in the strategies of the defense and the prosecution may be related to their respective roles and experiences in the trial. First, the prosecution is often a repeat player in ICC trials, while the defense is less frequently involved in such trials. As repeat players, prosecutors may be more familiar with using digital evidence in the courtroom, while defense teams may find it easier to manipulate classical evidence. Second, the prosecution bears the burden of proof and is the first party to present evidence at trial. In this context, proving the defendant's guilt *beyond a reasonable doubt* may require gathering incriminating evidence from various sources, including social networks. In contrast, the exculpatory evidence used by the defense may differ in nature from that used by the prosecution and may therefore be less readily available on digital platforms. As pointed out by Hellwig (2021), "as the Prosecutor holds the burden of proof, the Prosecution must proof all elements of the crime [...]. Crimes prosecuted under [international criminal law (ICL)] present some peculiarities with evidentiary implications, as the prosecution must provide context-related evidence, crime-based evidence and linkage evidence."

5.2 Sentiment Analysis

In this section, we use natural language processing (NLP) and, more specifically, sentiment analysis, to detect the presence of emotions during the debates between the parties and to determine whether emotions are more or less present when referring to digital and classical evidence. Sentiment analysis, also known as opinion mining, is a subfield of NLP. It can be used to analyze the emotional tone

and subjectivity of a text or speech. Here, we derive a sentiment index for each party in each of the 146 English transcripts in the Al Hassan case.[25] Specifically, we use TextBlob as our sentiment analysis algorithm.[26] This algorithm enables us to calculate a polarity index for each transcript and each participant in the trial. A polarity index (hereafter sentiment index) ranges from -1 to 1, with higher values indicating more positive sentiments.

Evolution of the sentiment index for each participant during the case. Figure 5 shows the variations of the sentiment index over the course of the case for each participant. The blue line shows the sentiment index, and its fluctuations reflect changes in emotional tone. The red dashed lines represent transitions between different phases of the procedure (Pretrial, Phase 1 – Prosecutor, Phase 2 – Victim, and Phase 3 – Defense). The red crosses indicates sentiments "peaks". To be considered as a relevant peak, the sentiment index has to be outside the set $[-0.5; 0.5]$. Indeed, the threshold range of 0.5 to -0.5 is often recognized as the "neutrality margin".[27] Small deviations from 0 are seen as lacking strong sentiment polarity and thus are interpreted as neutral. Several significant positive peaks (exceeding the 0.5 threshold) are observed in the *victims*' sentiments during the trial. Only one significant positive peak is observed in the *prosecution*'s sentiment index. The same is true for the *defense*. The *judge*'s sentiment index shows no significant positive or negative peaks. This last result suggests that the judge remained largely neutral throughout the trail, reflecting his position.

Spread of the sentiment index of each party. The spread of the parties' sentiment indexes is shown in the box plot 6.[28] We observe that the median sentiment for each party is slightly above

[25]Two of the 148 transcripts collected are in French. To make our results more consistent, we excluded these transcripts from the database.

[26]TextBlob is a popular Python library that provides a simple application programming interface (API) for common NLP tasks (see `https://textblob.readthedocs.io/en/dev/index.html` for details). An alternative to TextBlob is LegalBERT (`https://opensource.legal/projects/LegalBERT`). LegalBERT has the advantage of being specifically designed for legal texts. It is trained on a large corpus of legal documents (case law, statutes and regulations, contracts), which enables it to understand and process complex legal terminology and context. However, applying the LegalBERT algorithm instead of TextBlob does not produce any significant results in terms of the correlation between the sentiment indexes between the parties or with the rate of references to digital and classical evidence. We believe that this lack of results is due to LegalBERT's overspecialization in legal terminology. In fact, while LegalBERT excels at understanding formal legal documents, the court transcripts in our database are often conversational and argumentative. LegalBERT's specialization in formal legal language does not match the informal and interactive nature of the transcripts we have collected.

[27]See for instance Zaeem et al. (2020).

[28]The green line in the center of each box represents the median (50th percentile). The edges of the box correspond

zero, indicating an overall sentiment that is generally neutral to slightly positive. This may reflect the formality of the procedure and the resulting politeness of the parties involved. The spread of the sentiment indexes tends to be higher for victims and lower for the judge, and intermediate for the defense and the prosecution.[29]

to the first (Q1, 25th percentile) and the third (Q3, 75th percentile) quartiles, forming the interquartile range (IQR). The horizontal lines, or "whiskers," extend to the minimum and maximum values within 1.5 times the IQR from the quartiles. This indicates the range of most data points. Any points beyond the whiskers are considered outliers and represent extreme values that fall outside the typical range of the dataset.

[29]The defense has the highest coefficient of variation of the sentiment indexes (128.72%), followed by the victim (115.86%), the prosecutor (100.83%) and finally the judge (42.14%).

Figure 5: Evolution of the sentiment index for each party.

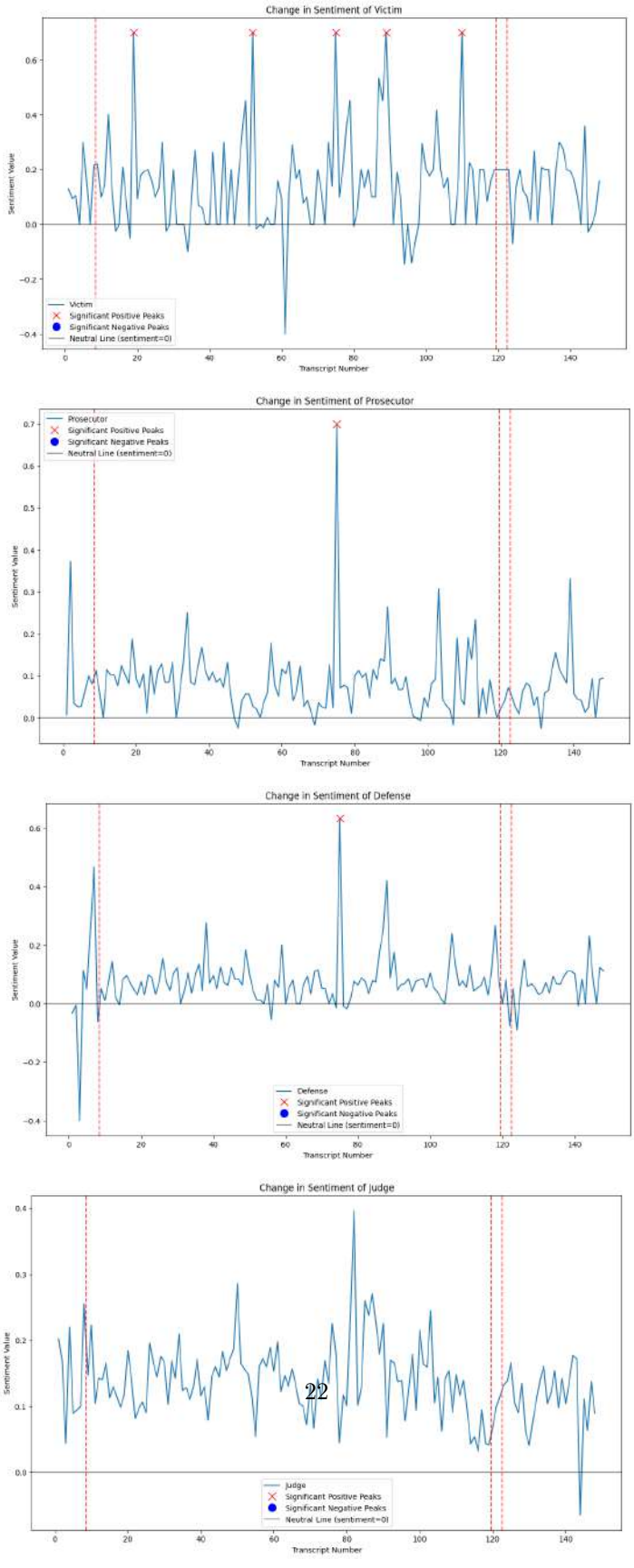
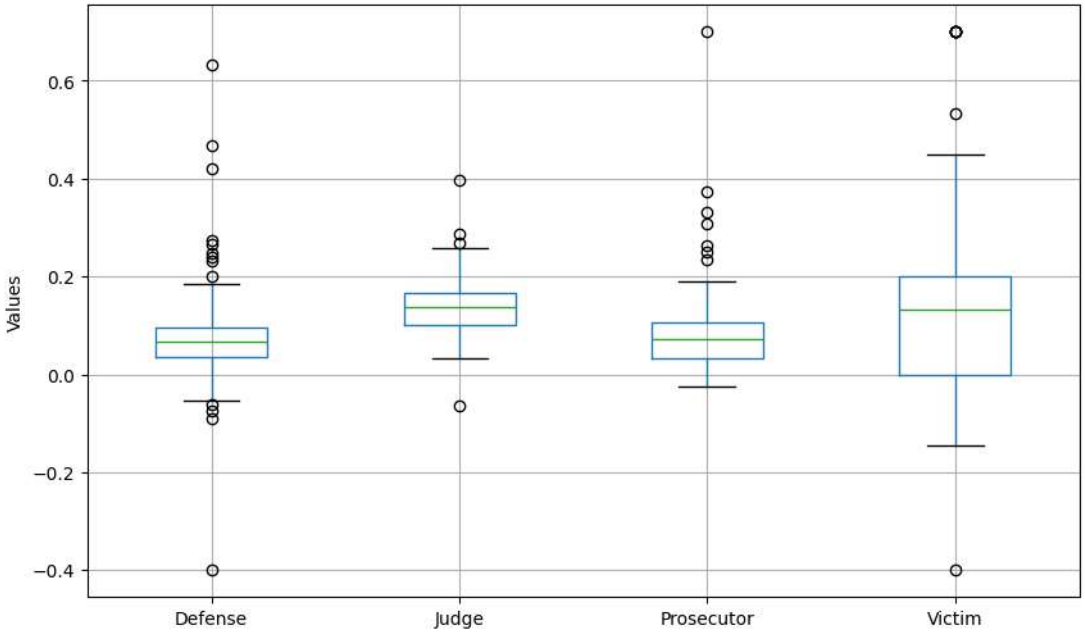


Figure 6: Boxplot of sentiment indexes for each party.



Correlation between the sentiments of each party. Table 7 shows the matrix of Pearson correlation coefficients of the sentiments expressed by each party. The table also includes the correlation coefficients with the rate of references to digital and classical evidence per 1000 words.

Table 7: Matrix of the Pearson correlation coefficients.

	Defense	Judge	Prosecutor	Victim	Rate Classical	Rate Digital
Defense	1					
Judge	0.0408	1				
	(0.6264)					
Prosecutor	0.3749***	0.2029**	1			
	(0.000)	(0.0143)				
Victim	0.2215***	0.0841	0.2736***	1		
	(0.0074)	(0.3146)	(0.0008)			
Rate Classical	-0.0637	0.2016**	-0.0120	-0.0202	1	
	(0.4463)	(0.0150)	(0.8863)	(0.8092)		
Rate Digital	0.0014	0.0645	0.0918	0.1847**	0.1404*	1
	(0.9865)	(0.4411)	(0.2720)	(0.0261)	(0.0921)	

Notes: In brackets, the p-values of the t-test.* $p < .05$, ** $p < .01$ and *** $p < .001$.

We find a significant, positive correlation between the prosecution and the defense sentiment indexes. Our results also show significant, positive correlations in the sentiment indexes between the prosecution and the victims, the prosecution and the judges, and the defense and the victims. Since these parties interact frequently in the context of the ICC proceedings, which tend to be adversarial, these correlations are not surprising. One possible interpretation of these positive correlation coefficients is that each party reacts to the opposing party's tone by using the same tone (*i.e.*, the harsher an intervention is, the harsher the opposing party's reaction will be).[30]

Examining the correlation between the sentiment indexes of the parties and their references to the two types of evidence, we find a statistically significant and positive correlation between references to classical evidence and judges' sentiment indexes. Additionally, we find a statistically significant and positive correlation between references to digital evidence and the sentiment indexes of the victims' representatives. One possible explanation for these correlation results is that judges and victims have different preferences regarding the means of evidence: judges are more positive toward classical evidence, while victims respond more positively to digital evidence. Despite the fact that

[30]Because of the small sample size, we should exercise caution when interpreting these results, especially regarding the victims, since they intervene to a lesser extent.

the prosecution relies more on digital evidence and the defense relies more on classical evidence (as highlighted in subsection 5.1), we find no correlation between the sentiment indexes of these parties and the rate of references to both types of evidence. A possible explanation for this lack of correlation is that the sentiments of the defense and the prosecution are more related to the topic and the tone of the discussion (and perhaps the content of the evidence presented) than to its form (digital or classical).

In summary, the findings of section 5 show that the prosecution and defense attorneys use digital and classic evidence differently in terms of the number of references. The prosecution relies more on digital evidence than the defense. However, we do not observe any difference in the emotional tone associated with the use of digital or classical evidence. Furthermore, the emotional tones of the prosecution and defense attorneys are significantly and positively correlated.

6 Conclusion

In this paper, we examine the use of digital *versus* classical evidence at the International Criminal Court.[31] Despite its usefulness, particularly when access to the territory is impossible, the probative value of digital evidence is often questioned or limited due to the difficulty of formally identifying and authenticating its source. Additionally, there is a high risk that the information is simply incorrect, *e.g.*, due to disinformation or misinformation. Our goal is to determine how much digital evidence is used in court, and how the prosecution uses it alongside classical evidence to meet the standard of proof during the pretrial and trial stages. We also aim to determine whether the prosecution and defense adopt different strategies regarding the use of different types of evidence.

First, we develop a theoretical model to understand how the standard of proof may affect how the prosecution collects both types of evidence given their different probative values. We assume that digital evidence is only corroborative and has lower probative value due to authentication difficulties specific to this type of evidence. Furthermore, we assume that the cost of collecting each type of evidence may vary (*e.g.*, the relative cost of collecting classical evidence may be high if the crime scene is not easily and safely accessible by investigators). Our theoretical results show that as the standard of proof becomes more stringent, the amount of digital and classical evidence produced by

[31] As a reminder, we consider digital evidence to be information transmitted over the internet, while classical evidence is primarily witness testimony.

the prosecutor increases. The relative increase in the collection of each type of evidence depends on the marginal collection costs, and the extent to which the two types of evidence are complementary.

Second, we apply textual analysis on trial transcripts to count the references to each type of evidence at different stages of the trial. Our empirical investigation shows that digital evidence is referred to less frequently than classical evidence. The intensity of references to both types of evidence increases between the pretrial and trial phases as the standard of proof rises. However, there are important differences in how each party uses the two types of evidence. Focusing specifically on the prosecution and defense in the Al Hassan case, we find that the prosecution's rate of references to digital evidence is higher than the defense's, especially during the pretrial phase.

Third, we use natural language processing, and specifically sentiment analysis, to measure the emotional tone of the parties involved in the Al Hassan trial. For each party in each transcript, we calculate a sentiment index and analyze how it changes throughout the case. We find that the spread of these indexes is greater for the victims' representatives than for the other parties. Furthermore, we observe a significant and positive correlation between the sentiment indexes of the prosecution and the defense, which may be due to the adversarial context and cross-examination aligning their emotional tones. Lastly, we observe no correlation between the sentiments of the defense or prosecution and their use of digital or classical evidence. Consequently, although the prosecution relies more on digital evidence than the defense, their emotional tones do not diverge significantly.

One limitation of our work is the timing of the crimes in relation to the use of social networks and digital communications. The hearing transcripts relate to crimes that began over a decade ago. Specifically, the Al Hassan case involves crimes that began in 2012. Over time, social networks have become more widespread. For instance, the recent Russian invasion of Ukraine has resulted in numerous digital recordings of Russian abuses. However, eventual trials related to the Ukrainian situation may not occur in the near future.

Appendix

Appendix A: Tables for Section 2

Table 8: Transcripts by year of crime start

	n	%
2010	225	43.2
2011	3	0.6
2012	152	29.2
2013	141	27.1
Total	521	100

Table 9: Keywords related to classical evidence

Statistic	Mean	Min	Max	Per 1000 words
Testimonial	0.1	0	7	0.003
Testimony	6.6	0	53	0.7
Witness	29.3	0	173	2.63

Note: All keyword variants are included. For **Testimonial**, the plural form *Testimonials* is also included. For **Testimony**, the plural form *Testimonies* is also included. For **Witness**, the plural form *Witnesses* is counted, but occurrences of *Mr Witness* and *Ms Witness* are excluded, as well as instances found in the expressions *"The witness gives evidence via video link"* and *"The witness speaks"*.

Table 10: Keywords related to digital evidence

Statistic	Mean	Min	Max	Per 1000 words
Facebook	0.3	0	18	0.02
Instagram	0.0	0	0	0
Linkedin	0.0	0	0	0
Snapchat	0.0	0	0	0
Telegram	0.002	0	1	0.00008
Tiktok	0.0	0	0	0
Twitter	0.05	0	13	0.013
Whatsapp	0.01	0	1	0.0011
Flickr	0.0	0	0	0
Youtube	0.2	0	18	0.009
Viadeo	0.0	0	0	0
Digital	0.1	0	11	0.013
Video	16.3	0	209	1.57
Footage	2.0	0	62	0.13
Internet	0.2	0	12	0.023
Hach	0.02	0	3	0.0031
Broadcast	1.4	0	23	0.12
Social Media	0.1	0	5	0.0062
Social Network	0.02	0	3	0.0026
Online Broadcast	0.0	0	0	0
Video Footage	0.6	0	11	0.042
Digital Evidence	0.004	0	1	0.00016
Open Data	0.0	0	0	0

Note: All keyword variants are included. For **Video** and **Footage**, the plural forms *Videos* and *Footages* are also counted. Occurrences of *Video* appearing in the expression *"The witness gives evidence via video link"* are excluded. All other keywords are counted exactly as they appear. For **Broadcast**, the following forms are included: *Broadcasts*, *Broadcasted*, and *Broadcasting*. For **Social Network**, the plural form *Social Networks* is included. For **Video Footage**, the plural form *Video Footages* is also included. For **Digital Evidence**, the plural form *Digital Evidences* is also counted.

Table 11: References to classical evidence in hearing transcripts

	n	%
No reference to classical evidence	17	3.3
Reference to classical evidence	504	96.7
Total	521	100

Table 12: References to digital evidence in hearing transcripts

	n	%
No reference to digital evidence	100	19.2
Reference to digital evidence	421	80.8
Total	521	100

Appendix B: The Optimal Levels of Classical Evidence.

The Lagrange function resulting from (1) is:

$$L(C,D) = \alpha C + \beta D + \lambda(f(C,D) - \overline{x}) \tag{4}$$

From the first-order conditions, we have:

$$\begin{aligned} \frac{\partial L}{\partial C}(C,D) &= \alpha + \lambda \frac{\partial f}{\partial C}(C,D) = 0 \\ \frac{\partial L}{\partial D}(C,D) &= \beta + \lambda \frac{\partial f}{\partial D}(C,D) = 0 \end{aligned} \tag{5}$$

The constraint in (1) is binding ($\lambda > 0$), because otherwise the prosecutor can always decrease $\alpha C + \beta D$ by decreasing either C or D without violating that constraint. Thus, choosing C and D such that $f(C,D) > \overline{x}$ cannot be optimal.

From (5) and the binding constraint, the optimal levels of classical evidence (C^*) and digital evidence (D^*) are characterized by:

$$\begin{cases} \frac{\frac{\partial f}{\partial C}(C^*,D^*)}{\frac{\partial f}{\partial D}(C^*,D^*)} &= \frac{\alpha}{\beta} \\ f(C^*,D^*) &= \overline{x} \end{cases} \tag{6}$$

Effect of an increase in the standard of proof on C^* and D^*.

By differentiating the first equation of (2) with respect to $\overline{x}$, we get:

$$\frac{\partial C^*}{\partial \overline{x}} \underbrace{\left[\beta \frac{\partial^2 f}{\partial C^2}(C^*,D^*) - \alpha \frac{\partial^2 f}{\partial C \partial D}(C^*,D^*)\right]}_{\equiv X} + \frac{\partial D^*}{\partial \overline{x}} \underbrace{\left[\beta \frac{\partial^2 f}{\partial C \partial D}(C^*,D^*) - \alpha \frac{\partial^2 f}{\partial D^2}(C^*,D^*)\right]}_{\equiv Y} = 0 \tag{7}$$

Similarly, by differentiating the second equation of (2), we get:

$$\frac{\partial C^*}{\partial \overline{x}} \frac{\partial f}{\partial C}(C^*,D^*) + \frac{\partial D^*}{\partial \overline{x}} \frac{\partial f}{\partial D}(C^*,D^*) = 1 \tag{8}$$

We can write the system of equations formed by (7) and (8) in matrix form:

$$\begin{pmatrix} X & Y \\ \frac{\partial f}{\partial C}(C^*,D^*) & \frac{\partial f}{\partial D}(C^*,D^*) \end{pmatrix} \begin{pmatrix} \frac{\partial C^*}{\partial \overline{x}} \\ \frac{\partial D^*}{\partial \overline{x}} \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \tag{9}$$

We assume that the determinant of the first matrix, denoted H in the following, is negative (it is a necessary condition of the minimization program (1)).

Applying Cramer's rule to (9) yields:

$$\frac{\partial C^*}{\partial \overline{x}} = \frac{\begin{vmatrix} 0 & Y \\ 1 & \frac{\partial f}{\partial D}(C^*, D^*) \end{vmatrix}}{H} = -\frac{Y}{H} \tag{10}$$

From which it follows that:

$$\text{sign } \left(\frac{\partial C^*}{\partial \overline{x}} \right) = \text{sign } (Y) > 0 \tag{11}$$

And:

$$\frac{\partial D^*}{\partial \overline{x}} = \frac{\begin{vmatrix} X & 0 \\ \frac{\partial f}{\partial C}(C^*, D^*) & 1 \end{vmatrix}}{H} = \frac{X}{H} \tag{12}$$

From which it follows that:

$$\text{sign } \left(\frac{\partial D^*}{\partial \overline{x}} \right) = \text{sign } (-X) > 0 \tag{13}$$

Condition for which the quantity of classical evidence increases relatively more with the standard.

We compare (10) and (12). For a small increase in the standard, the quantity of classical evidence increases more than the quantity of digital evidence if:

$$\begin{aligned} \frac{\partial C^*}{\partial \overline{x}} > \frac{\partial D^*}{\partial \overline{x}} &\Leftrightarrow X + Y > 0 \\ &\Leftrightarrow \beta \frac{\partial^2 f}{\partial C^2}(C^*, D^*) - \alpha \frac{\partial^2 f}{\partial D^2}(C^*, D^*) + (\beta - \alpha) \frac{\partial^2 f}{\partial C \partial D}(C^*, D^*) > 0 \end{aligned} \tag{14}$$

Appendix C: References by Each Participant Depending on the Trial Phase.

Number of words. Table 13 provides the number of words of each participant by phase in the Al Hassan case, as well as the proportion of words for each participant by phase.

Table 13: Total number of words and percentages for participants in pretrial and Trial

Participant	**pretrial**		**Phase 1 - P**		**Phase 2 - V**		**Phase 3 - D**		**TOTAL**	
	Nbr	**%**	**Nbr**	**%**	**Nbr**	**%**	**Nbr**	**%**	**Nbr**	**%**
P	54488	62.1	162300	36.6	1192	5.4	55811	28.5	273791	36.6
D	19577	22.3	97807	22.1	3129	14.1	76550	39.1	197063	26.3
V	7836	8.9	14188	3.2	10641	47.9	3906	2.0	36571	4.9
J	5841	6.7	168650	38.1	7253	32.6	59636	30.4	241380	32.2
TOTAL	**87742**	100.0	**442945**	100.0	**22215**	100.0	**195903**	100.0	**748805**	100.0

Reference to digital and classical evidence by the participants. Table 14 and 15 indicates respectively the number of references to both digital and classical evidence by the participants during the pretrial and trial phases.

Table 14: Total number of words referring to classical evidence for participants in pretrial and trial

Participant	**pretrial**	**Trial**			**TOTAL**
		Phase 1 - P	**Phase 2 - V**	**Phase 3 - D**	
Office of the Prosecutor (P)	70	1270	1	283	**1624**
Defense (D)	82	688	24	755	**1549**
Victims' Repr. (V)	4	71	28	14	**117**
Judges (J)	26	1004	40	330	**1400**
TOTAL	**182**	**3033**	**93**	**1382**	**4690**

Table 15: Total number of words referring to digital evidence for participants in pretrial and trial

Participant	**pretrial**	**Trial**			**TOTAL**
		Phase 1 - P	**Phase 2 - V**	**Phase 3 - D**	
P	161	1370	0	95	**1626**
D	25	339	5	209	**578**
V	9	0	0	4	**13**
J	2	50	0	15	**67**
TOTAL	**197**	**1759**	**5**	**323**	**2284**

Appendix D: Methodology and Examples of the Hearing Transcript Cleaning Process

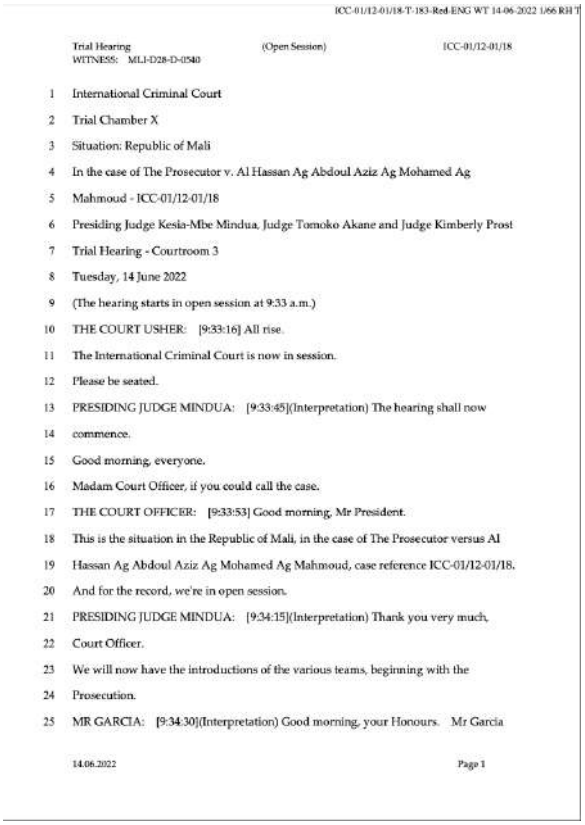
- **First:** a cleaning process had been applied on each transcript to remove headers, footers, and page line numbers.
- **Second:** we performed some textual preprocessing on the transcripts to capture each speaker's words during the hearing, with a particular focus on Q&A segments (see examples of Figure 7 and Figure 8).
- **Third:** we applied frequency analysis methods to tabulate the occurrences of our keywords. Note that to ensure comprehensive coverage of words related to both digital and classical evidence, we carefully selected specific terms and their synonyms for each category. To enhance this process, we leveraged the *Thesaurus API*[32], which automates the search for synonyms. To ensure semantic comprehensiveness, we employed the Thesaurus API not merely as a synonym generator but as a systematic extension tool. For each core keyword (e.g., "video", "witness"), the API provided a list of related terms, which we then manually curated to retain only those relevant within the legal and evidentiary context. This hybrid approach of automated expansion and manual filtering enhanced the reliability of our word occurrence analysis by accounting for lexical variations, while minimizing noise from irrelevant synonyms.

  Furthermore, in order to improve the accuracy of our word counts, we also established specific conditions for counting the selected terms. Certain phrases were excluded to avoid misclassification of words, such as "The witness testifies via video link" for the term "video" and "Mr. Witness" or "Ms. Witness" for the word "witness".
- **Fourth:** we matched the word count for each speaker in both the pretrial and trial transcripts to their respective roles in the case.

It should be noted that among the publicly available transcripts on the ICC website, there are some transcripts with sections that have been redacted and thus omitted from public access (see example of Figure 9). This affects the completeness of the data available for analysis and the results.

[32]From API Ninjas: https://api-ninjas.com/api/thesaurus

Figure 7: Transcript sample from the *Al Hassan* case in the trial phase. The first image (left) shows the publicly available transcript from the ICC website, and the second image (right) shows the transcript after cleaning and reorganization.



ICC-01/12-01/18-T-183-Red-ENG WT 14-06-2022 1/66 RH T

Trial Hearing (Open Session) ICC-01/12-01/18
WITNESS: MLI-D28-D-0540

International Criminal Court
Trial Chamber X
Situation: Republic of Mali
In the case of The Prosecutor v. Al Hassan Ag Abdoul Aziz Ag Mohamed Ag Mahmoud - ICC-01/12-01/18
Presiding Judge Kesia-Mbe Mindua, Judge Tomoko Akane and Judge Kimberly Prost
Trial Hearing - Courtroom 3
Tuesday, 14 June 2022
(The hearing starts in open session at 9:33 a.m.)
THE COURT USHER: [9:33:16] All rise.
The International Criminal Court is now in session.
Please be seated.
PRESIDING JUDGE MINDUA: [9:33:45](Interpretation) The hearing shall now commence.
Good morning, everyone.
Madam Court Officer, if you could call the case.
THE COURT OFFICER: [9:33:53] Good morning, Mr President.
This is the situation in the Republic of Mali, in the case of The Prosecutor versus Al Hassan Ag Abdoul Aziz Ag Mohamed Ag Mahmoud, case reference ICC-01/12-01/18.
And for the record, we're in open session.
PRESIDING JUDGE MINDUA: [9:34:15](Interpretation) Thank you very much, Court Officer.
We will now have the introductions of the various teams, beginning with the Prosecution.
MR GARCIA: [9:34:30](Interpretation) Good morning, your Honours. Mr Garcia

14.06.2022 Page 1

Nom: 10 THE COURT USHER
Heure: 9:33:45
Commentaire:
All rise.
The International Criminal Court is now in session.
Please be seated.

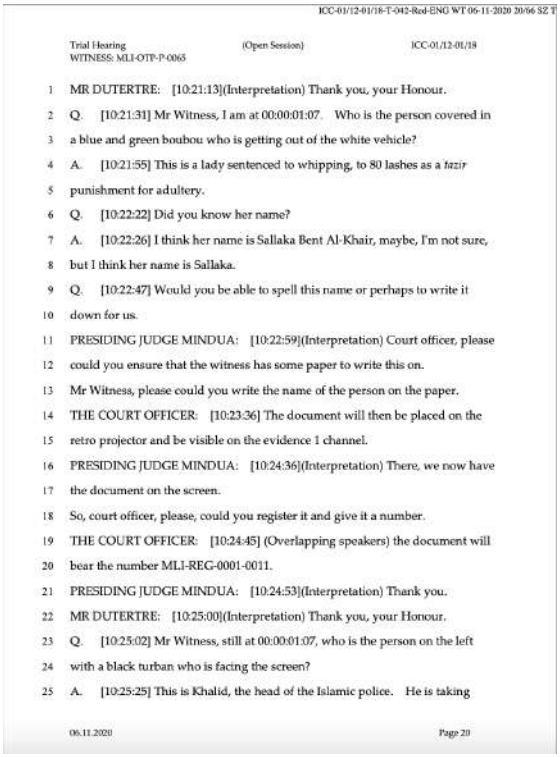
Nom: 13 PRESIDING JUDGE MINDUA
Heure: 9:33:53
Commentaire:
(Interpretation) The hearing shall now commence.
Good morning, everyone.
Madam Court Officer, if you could call the case.

Nom: 17 THE COURT OFFICER
Heure: 9:34:15
Commentaire:
Good morning, Mr President.
This is the situation in the Republic of Mali, in the case of The Prosecutor versus Al Hassan Ag Abdoul Aziz Ag Mohamed Ag Mahmoud, case reference ICC-01/12-01/18.
And for the record, we're in open session.

Nom: 21 PRESIDING JUDGE MINDUA
Heure: 9:34:30
Commentaire:
(Interpretation) Thank you very much, Court Officer.
We will now have the introductions of the various teams, beginning with the Prosecution.

Nom: 25 MR GARCIA

Figure 8: Transcript sample from the *Al Hassan* case in the trial phase - Q&A segment. The first image (left) shows the publicly available transcript from the ICC website, and the second image (right) shows the transcript after cleaning and reorganization.

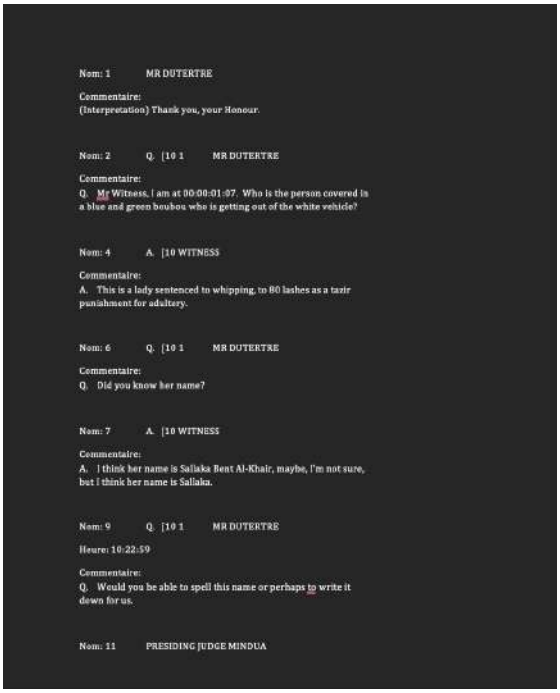


ICC-01/12-01/18-T-042-Red-ENG WT 06-11-2020 20/66 SZ T

Trial Hearing (Open Session) ICC-01/12-01/18
WITNESS: MLI-OTP-P-0065

MR DUTERTRE: [10:21:13](Interpretation) Thank you, your Honour.
Q. [10:21:31] Mr Witness, I am at 00:00:01:07. Who is the person covered in a blue and green boubou who is getting out of the white vehicle?
A. [10:21:55] This is a lady sentenced to whipping, to 80 lashes as a *tazir* punishment for adultery.
Q. [10:22:22] Did you know her name?
A. [10:22:26] I think her name is Sallaka Bent Al-Khair, maybe, I'm not sure, but I think her name is Sallaka.
Q. [10:22:47] Would you be able to spell this name or perhaps to write it down for us.
PRESIDING JUDGE MINDUA: [10:22:59](Interpretation) Court officer, please could you ensure that the witness has some paper to write this on. Mr Witness, please could you write the name of the person on the paper.
THE COURT OFFICER: [10:23:36] The document will then be placed on the retro projector and be visible on the evidence 1 channel.
PRESIDING JUDGE MINDUA: [10:24:36](Interpretation) There, we now have the document on the screen. So, court officer, please, could you register it and give it a number.
THE COURT OFFICER: [10:24:45] (Overlapping speakers) the document will bear the number MLI-REG-0001-0011.
PRESIDING JUDGE MINDUA: [10:24:53](Interpretation) Thank you.
MR DUTERTRE: [10:25:00](Interpretation) Thank you, your Honour.
Q. [10:25:02] Mr Witness, still at 00:00:01:07, who is the person on the left with a black turban who is facing the screen?
A. [10:25:25] This is Khalid, the head of the Islamic police. He is taking

06.11.2020 Page 20



Nom: 1 MR DUTERTRE
Commentaire:
(Interpretation) Thank you, your Honour.

Nom: 2 Q. [10 1 MR DUTERTRE
Commentaire:
Q. Mr Witness, I am at 00:00:01:07. Who is the person covered in a blue and green boubou who is getting out of the white vehicle?

Nom: 4 A. [10 WITNESS
Commentaire:
A. This is a lady sentenced to whipping, to 80 lashes as a tazir punishment for adultery.

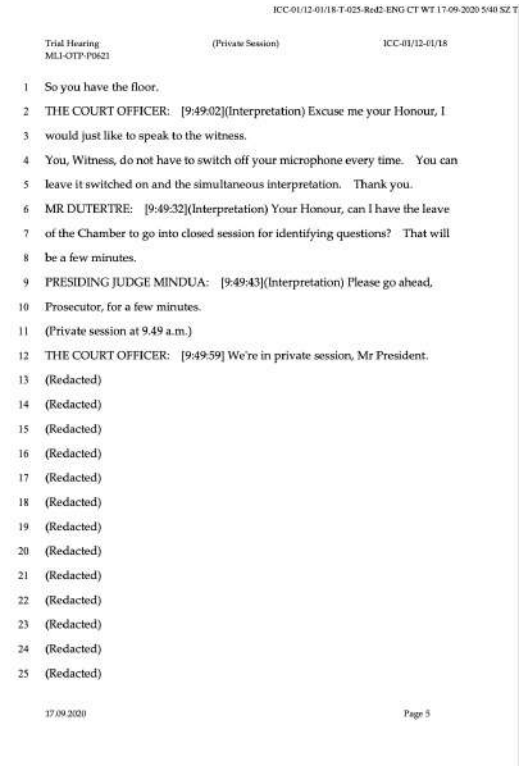
Nom: 6 Q. [10 1 MR DUTERTRE
Commentaire:
Q. Did you know her name?

Nom: 7 A. [10 WITNESS
Commentaire:
A. I think her name is Sallaka Bent Al-Khair, maybe, I'm not sure, but I think her name is Sallaka.

Nom: 9 Q. [10 1 MR DUTERTRE
Heure: 10:22:59
Commentaire:
Q. Would you be able to spell this name or perhaps to write it down for us.

Nom: 11 PRESIDING JUDGE MINDUA

Figure 9: Transcript sample from the *Al Hassan* case in the trial phase - private session.
The first image (left) shows the publicly available transcript from the ICC website, and the second image (right) shows the transcript after cleaning and reorganization.



ICC-01/12-01/18-T-025-Red2-ENG CT WT 17-09-2020 5/40 SZ T

Trial Hearing (Private Session) ICC-01/12-01/18
MLI-OTP-P0621

So you have the floor.
THE COURT OFFICER: [9:49:02](Interpretation) Excuse me your Honour, I would just like to speak to the witness.
You, Witness, do not have to switch off your microphone every time. You can leave it switched on and the simultaneous interpretation. Thank you.
MR DUTERTRE: [9:49:32](Interpretation) Your Honour, can I have the leave of the Chamber to go into closed session for identifying questions? That will be a few minutes.
PRESIDING JUDGE MINDUA: [9:49:43](Interpretation) Please go ahead, Prosecutor, for a few minutes.
(Private session at 9.49 a.m.)
THE COURT OFFICER: [9:49:59] We're in private session, Mr President.
(Redacted)
(Redacted)
(Redacted)
(Redacted)
(Redacted)
(Redacted)
(Redacted)
(Redacted)
(Redacted)
(Redacted)
(Redacted)
(Redacted)
(Redacted)

17.09.2020 Page 5

of the Chamber to go into closed session for identifying questions? That will be a few minutes.

Nom: 9 PRESIDING JUDGE MINDUA

Heure: 9:49:59

Commentaire:
(Interpretation) Please go ahead, Prosecutor, for a few minutes.
(Private session at 9.49 a.m.)

Nom: 12 THE COURT OFFICER

Heure: 10:00:15

Commentaire:
We're in private session, Mr President.
(Redacted)
(Redacted)
(Redacted)
(Redacted)
(Redacted)
(Redacted)
(Redacted)
(Redacted)
(Redacted)
(Redacted)
(Redacted)
(Redacted)
(Redacted)
2
3
4
5
6
7
8
9
Page redacted – Private session
2
3

References

Aronson, J. D. (2018). The utility of user-generated content in human rights investigations. New Technologies for Human Rights Law and Practice, pages 129–48.

Ash, E., Chen, D. L., and Galletta, S. (2022). Measuring judicial sentiment: Methods and application to us circuit courts. Economica, 89(354):362–376.

Demougin, D. and Fluet, C. (2008). Rules of proof, courts, and incentives. The RAND Journal of Economics, 39(1):20–40.

Emons, W. and Fluet, C. (2009). Accuracy versus falsification costs: The optimal amount of evidence under different procedures. The Journal of Law, Economics, & Organization, 25(1):134–156.

Frankenreiter, J. and Livermore, M. A. (2020). Computational methods in legal analysis. Annual Review of Law and Social Science, 16(1):39–57.

Froeb, L. M. and Kobayashi, B. H. (2001). Evidence production in adversarial vs. inquisitorial regimes. Economics Letters, 70(2):267–272.

Galanter, M. (1974). Why the haves come out ahead: Speculations on the limits of legal change. Law & Soc'y Rev., 9:95.

Hellwig, K. (2021). The potential and the challenges of digital evidence in international criminal proceedings. International Criminal Law Review, 22(5-6):965–988.

Jouannet, E. (2007). Remarques théoriques: La preuve comme reflet des évolutions majeures de la société internationale. In et Jean-Marc Sorel, H. R. F., editor, La preuve devant les juridictions internationales. Paris: Pedonne.

Kaplow, L. (2012). Burden of proof. The Yale Law Journal, pages 738–859.

Laux, J. (2018). A new type of evidence? cyberinvestigations, social media, and online open source video evidence at the icc. Archiv des völkerrechts, 56(3):324–360.

Mungan, M. C. (2011). A utilitarian justification for heightened standards of proof in criminal trials. Journal of Institutional and Theoretical Economics (JITE)/Zeitschrift für die gesamte Staatswissenschaft, pages 352–370.

Obidzinski, M. and Oytana, Y. (2019). Identity errors and the standard of proof. International Review of Law and Economics, 57:73–80.

Posner, R. A. (1999). An economic approach to the law of evidence. Stan. L. Rev., 51:1477.

Rizzolli, M. (2016). Adjudication type i and type ii errors. Forthcoming in the Encyclopedia of Law and Economics.

Rizzolli, M. and Saraceno, M. (2013). Better that ten guilty persons escape: punishment costs explain the standard of evidence. Public Choice, 155(3):395–411.

Talley, E. L. (2013). Law, economics, and the burden(s) of proof. In Research Handbook on the Economics of Torts, pages 305–329. MA: Edward Elgar.

Wippman, D. (2006). The Costs of International Justice. The American Journal of International Law, 100:21.

Zaeem, R. N., Li, C., and Barber, K. S. (2020). On sentiment of online fake news. In 2020 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), pages 760–767. IEEE.