Unus Video, Nullus Video

Synthetic video generation and its potential impact on the assessment of the reliability of video evidence in the Dutch criminal system

Name:	Ivar van Dillen
Student number:	1256248, 312490
Date:	15-01-2019
Department:	Tilburg Institute for Law, Technology, and Society
First Supervisor:	S. Jameson
Second Supervisor:	T. Chokrevski

To my late mother, Mariëtte van Dillen - de Koning,
without whom I would not be where I am today.

With special thanks to my first and second supervisors, Shazade Jameson and Tomislav Chokrevski, whose comments and support have already proven invaluable.

Table of contents

Introduction

In the fall of 2017 a member of the internet forum Reddit published the outcome of their first experiments with a homemade deep-learning algorithm trained to replace faces and expressions of persons in videos.[1] The algorithm was freely available, and it did not take long for others to pick up on it. A total of two months was all it took for the algorithm to be transformed into an easy to use application. What was a rough set of code now only required a source video, a minute of footage of the transposed face, a decently powerful computer and several hours of processing time. Before long, users saw the possibilities provided by the technology. All kinds of content emerged: from altering movies by replacing the characters with the likeness of Nicholas Cage, to more questionable uses such as replacing actors in pornographic content with the likeness of celebrities.[2] With the birth of the deepfake, synthetic video generation entered public consciousness for the first time.

It too did not take long for the media to pick up on this, including in the Netherlands.[3] After the initial mostly negative outburst, several of the larger internet service providers reacted by banning deepfakes with pornographic content from their platforms.[4] With this move, most of the unrest regarding the technology disappeared as well. The technology itself however has not disappeared and is likely here to stay. In its short wake in the spotlight, it has already opened people's eyes to possible disruptions it can bring in the future. This includes several legal questions regarding the content produced.[5]

Among these legal questions lies the role of synthetic video generation in the courtroom. With the rise of the mobile phone, the access to cameras in our everyday world increased many times over, to en expected 5.07 billion worldwide users in 2019.[6] With these technological advances, digital evidence has increasingly started to enter the courtroom, with for example a tenfold increase of the use of email as evidence over the years 2000-2009 in the Netherlands.[7] On top of that, cases in which outings and imagery on social media such as FaceBook and messaging services such as WhatsApp played a major role have already taken place.[8] Judges and lawyers have to assess this evidence, deeming whether it is reliable enough to be used in court and what weight this evidence should carry.

[1] G. Oberoi, 'Exploring Deepfakes' (*Hackernoon*, March 5, 2018) <https://hackernoon.com/exploring-deepfakes-20c9947c22d9 > Accessed on April 22, 2018.

[2] D. Lee, 'Deepfakes Porn Has Serious Consequences' (*BBC*, February 3, 2018) <www.bbc.com/news/technology-42912529> Accessed on April 22, 2018.

[3] Redactie, 'Na fake nieuws ook fake porno? 'Dit gaat echt heel naar worden.'' (*NOS*, December 13, 2017) <https://nos.nl/op3/artikel/2207417-na-fake-nieuws-ook-fake-porno-dit-gaat-echt-heel-naar-worden.html> Accessed on May 5, 2018.

[4] M. Farokhmanesh, 'Deepfakes are disappearing from parts of the web, but they're not going away' (The Verge, February 9, 2018) <www.theverge.com/2018/2/9/16986602/deepfakes-banned-reddit-ai-faceswap-porn> Accessed on April 22, 2018.

[5] J. Christian, 'Experts Fear Face Swapping Tech Could Start an International Showdown' (*The Outline*, February 1, 2018) <https://theoutline.com/post/3179/deepfake-videos-are-freaking-experts-out?zd=2&zi=4n4jjajh> Accessed on April 22, 2018.

[6] GSMA, Measuring the Future of Mobile (*GSMA Intelligence data December 2017*, 2017) <www.gsma.com/mobileeconomy > Accessed on April 22, 2018.

[7] M. Stekelenburg, *De Betere Byte in de Strijd Om Het Gelijk* (Dissertation, Vrije Universiteit Amsterdam 2009), p. 241.

[8] J.J. Oerlemans, 'Veroordeling Voor "Uitreizen" Naar Syrië En de Rol Digitaal Bewijs' (2017) *Computerrecht* 242-243.

However, legal experts are generally not experts in the detection of modification or alteration of digital imagery. After all, they are legal experts trained in the theory and application of the law. Nuances such as pixel-patterns that might indicate a falsification to an expert might not be visible to the eye of the judge. It is possible for the judge to call in an expert to determine this, but for this to happen the judge must first deem this necessary. This leads to a paradox of knowledge: for the judge to call in an expert, they must first be knowledgeable enough to know that an expert is required.[9] Synthetic video generation, defined as artificial generation and alteration of falsified digital imagery at a highly realistic level, is already difficult detect even in these early days of the technology. It then begets the question whether the judge is knowledgeable enough to identify synthetically generated video.

Whereas it is commonly accepted that photographs can be easily edited, video editing has traditionally had a higher barrier to entry. Programs such as Adobe Photoshop are well-known, even part of the everyday language, and image editing is part of business as usual in many industries. Video editing has traditionally had a higher barrier of entrance. Generally every second of video requires the editing of 24 frames in a way that ends up looking smooth and realistic. Not only does this greatly increase the amount of work, it also increases the required know-how, computing power and time invested. Synthetic video generation could lower this barrier of entry significantly. By having the neural network do the heavy lifting, an average computer-literate person is able to do the work of what used to require a professional, at a fraction of the cost and time.[10] As such, the barrier to entry of alteration of video files might be greatly lowered.

All these factors combine into a possible problem. There is a new advancement in technology, leading to easy alteration of video data. The technology is commonly available and accessible. Meanwhile, evidence is graded by legal experts, generally not knowledgeable enough to such alterations. The technology behind deepfakes is still new, with this application of synthetic video generation being only in active development since December 2017. In this time, it has developed from a rough set of code that creates easy to spot fakes to an one-click app that produces realistic results. The technology is out in the open now and not only enthusiasts but also academics are working on commercializing and above all improving it. The Face2Face project is aimed at real-time replacement of a target person's expressions by the use of a second feed and a source actor.[11] The expressions made by the source actor are then rendered upon the target's face. A more commercial example is Lyrebird, a company spun out of the University of Montreal, aimed at creating realistic voice samples from as little as one minute of footage using the same kind of generative network.[12] With more than just enthusiasts interested in the technology, progress is bound to be made. It is not unthinkable that one day these alterations are no longer detectable to the human or even mechanical eye. Digital video evidence could become less reliable due to the possibility of it being synthetically generated.

[9] M. Evenblij, 'Gerechtelijk Deskundigen: Vechten Tegen De Kennisparadox' (2008) *Mr. Magazine* 6/7, p. 62-69.
<www.lrgd.nl/Portals/1/publicatieswebsite/Mrmagazine20080706.pdf> Accessed on April 23, 2018.

[10] A. Hern, 'My May-Thatcher Deepfake Won't Fool You But It's Tech May Change The World', (*The Guardian*, March 12, 2018)
<www.theguardian.com/technology/2018/mar/12/may-thatcher-deepfake-face-swap-tech-change-world> Accessed on Aprill 22, 2018.

[11] J. Thies and others, 'Face2face: Real-Time Face Capture And Reenactment Of RGB Videos' (2016) *2016 IEEE Conference on Computer Vision and Pattern Recognition*, p. 2387-2395.

[12] Lyrebird, 'about us' <https://lyrebird.ai> Accessed on May 4, 2018.

Evidence having a higher chance of being unreliable does however not have to result in it being inadmissible. The value that a digital video file can bring to a case might be too large for this altogether. On top of this, that it is possible for a video file to be synthetically generated does not mean that it actually is, and as such it should not be barred from being accepted as evidence in general. When it comes to fallible video evidence, one can see parallels with another type of evidence that has already gone through the development from heavy-weight evidence to a more careful approach: the eyewitness.

The eyewitness has always played a major role in the history of the courtroom. The eyewitness is a person who testifies their observations and experiences of what has transpired. The witness recalls the events from memory and relays them to the court in person or in writing. In the earlier half of the previous century, this was seen as heavy-weight proof and often carried with it a deciding factor in the judgement of cases.[13] Following formation of the misinformation effect, the finding that the memory of witnesses is highly malleable by misinformation, this view has drastically changed.[14] Psychological research has shown that the eyewitness can be influenced in many way and that as a result the memories surrounding the incident can be greatly altered. Some examples of research in this area are the findings that positive feedback can distort the eyewitness' memory,[15] the findings that mistaken eyewitness identification in lineups was the largest single factor contributing to incorrect judgements,[16] as well as the general negative influence of stress on the eyewitness' memory.[17] When the memories of an eyewitness as such are altered, his testimony is in his eyes not a faulty one and the witness is not lying. It is however not objectively accurate either. The witness testimony can be both false and true without an indication in the testimony alone whether it is false or true. The judge is deemed capable to discern whether testimonial evidence is accurate, but research indicates that this too might be doubtful.[18] Despite these flaws and doubts, eyewitness evidence is still used to this day.

Digital evidence, specifically video evidence, might have more in common with eyewitness evidence than it would appear to at first sight. The deepfake is only an early application of synthetic video generation. It is then not unthinkable that in the (near) future, synthetic video generation reaches such a level that it is indistinguishable to the human and perhaps even the mechanical eye. In such a circumstance, synthetic video evidence could be seen as the same as the eyewitness' testimony: the testimony could be both accurate and inaccurate in the same way that what is shown by video could be accurate or inaccurate. It would be impossible to tell without the use of supporting evidence, as the accuracy of the statement and the video on their own are unknown.

The deepfake can perhaps be seen as an early harbinger of the disruptions to be caused by synthetic video generation. Synthetic video evidence can show what really transpired, but could also show a false depiction without indication of which. Fallible evidence however

[13] G.L. Wells, A. Memon and S.D. Penrod, 'Eyewitness Evidence, Improving its Probative Value' (2006) 7(2) *Psychological Science in the Public Interest*, p. 45-75.

[14] E.F. Loftus, 'Planting misinformation in the human mind: A 30-year investigation of the malleability of memory' (2005) *Learning & Memory* 12, p. 361-366.

[15] G.L. Wells and A.L. Bradfield, '"Good, you identified the suspect:" Feedback to eyewitnesses distorts their reports of the witnessing experience' (1998) *Journal of Applied Psychology* 83(3), p. 360 - 376.

[16] G.L. Wells and E.A. Olson, 'Eyewitness Testimony' (2003) *Annual Review of Psychology* 45:279-95, p. 277-290.

[17] A.A. Ahanorian and B.H. Bornstein, 'Stress and Eyewitness Memory' in B.L. Cutler (ed), *Encyclopedia of Psychology and Law* (SAGE Publications, Newbury Park 2008).

[18] M. Sauerland, A.C. Krix, and M. Merckelbach, 'Identificaties Door Ooggetuigen: Waarom Een Rechtspsycholoog Handig Is' (2016) *Nederlands Juristenblad* 2016/1562.

does not necessarily mean it is inadmissible evidence, as is demonstrated by the use of the eyewitness in the court.

As such, I will try to discern how the Dutch criminal system currently regulates the use of digital evidence, specifically video evidence, and in what manner synthetic video generation can disrupt this system. I will then try and look deeper into evidence as a whole, looking at what the judge observes to determine whether evidence is reliable. To do so, I will try to discern indicators of reliability from the relevant literature. I will then apply these to digital video evidence, as well as look at how these indicators are disturbed by synthetic video generation. Finally, I will try and draw a comparison between the eyewitness and synthetic video evidence, to discern whether historical developments seeing to assessing the reliability of eyewitness evidence are perhaps applicable to synthetic video evidence.

To do so, I have formulated the following research question:

> *How does the development of synthetic video generation, as illustrated by the development of deepfakes, disrupt the use of digital video evidence in the Dutch criminal system and how does this compare to the use and regulation of eyewitness evidence in the courtroom?*

To come to a satisfactory answer for the research question, I have divided the research topic into three sub-questions which will each form their own chapter. These sub-questions are as follows:

1. *How does the Dutch criminal law system regulate digital video evidence and how does the development of deepfakes and the underlying development of synthetic video generation threaten this regulation?*
    1. *How does the Dutch criminal law system currently regulate evidence, and how does digital video evidence fit within this system?*
    2. *What are deepfakes and synthetic video generation?*
    3. *How does synthetic video generation disrupt the Dutch regulation of digital video evidence?*
2. *What makes evidence reliable, and how does this affect digital video evidence with regard to the disruption by synthetic video generation?*
    1. *How does the court assess the reliability of the evidence?*
    2. *What found indicators of reliability can be applied to digital video evidence and in what way are these indicators disrupted by synthetic video generation?*
3. *How can synthetic video evidence be compared to eyewitness evidence, and are there lessons to be learned from how the treatment eyewitness evidence has developed over the years that are applicable to synthetic video evidence?*
    1. *How can eyewitness evidence and synthetic video evidence be compared in regards to their respective reliability?*
    2. *Are there lessons to be learned from how eyewitness evidence has been developed over the years that are applicable to synthetic video evidence?*

The main methodology will consist of classical dogmatic research. This will be most apparent in the first sub-question, in which the current law and jurisprudence will be regarded and the current system of evidence will be explored. Here too I will take a deeper look at deepfakes as a herald of synthetic video generation, as well as its disruption within the system of evidence. This will mainly be done through the use of legal literature.

The second and third subquestion too shall consist mainly of classical dogmatic research, but will include other methodologies as well. Both shall at least partially depend on interdisciplinary research, with mainly legal-psychological studies regarding the reliability of witnesses and reliability in general as the interdisciplinary focus. These are not to be used on their own, but as reinforcement of legal theories introduced and explored in these chapters. It would in my eyes be irresponsible to not explore these theories when it comes to the reliability of the eyewitness: a subject that is deftly interwoven with the human psyche, and of which most criticism comes from academia in legal psychology.

This thesis shall not include the more technological aspects, nor the solutions that could be provided by technological means. An example of this would be the evaluation of the effects of gradual degradation of hardware inside the video camera, which could then be used as a digital fingerprint of the footage in the judicial process. The more technological aspect surely plays an important role in this dilemma, but observing this would fall far outside of the scope of this thesis. Uses of the law shall mostly be restricted to the Dutch system where it is possible and appropriate, and systems of law alike to it where it is not.

The first chapter consists of an introduction to the topic, the research-questions and the methodology. The second chapter looks at the current situation and the technological innovation causing the disruption. The third chapter consists of an analysis of legal theories pertaining to the reliability of evidence, as well as standards of law, in an attempt to isolate elements of reliability, that will then be applied to fallible digital video evidence. The fourth chapter attempts to forge a link between the treatment of eyewitness evidence and synthetic video evidence by making use of the found indicators of reliability, and observes whether any lessons learned from the treatment of eyewitness evidence could be applied to synthetic video evidence.The fifth chapter consists of the final conclusion, in which the sub-questions will be summarised and applied to provide an answer to the main research question.

Chapter 2 A foundation to the disruption

How does the Dutch criminal law system regulate digital video evidence and how does the development of deepfakes and the underlying development of synthetic video generation threaten this regulation?

Digital evidence is a fairly recent development, coming into existence entirely after the initial creation of the Dutch Criminal Procedural Code[19] ["CPC"] in 1926. The code has since been amended and updated, but still uses the original framework. This framework is seemingly starting to become unfit for today's modernizing world.[20] To understand the disruptions synthetic video generation might cause, it is of importance to first see how the Dutch system of evidence works, including the role of digital video evidence within this system. Synthetic video generation too requires some exposition to understand the possibilities. Finally, I will look at how synthetic video generation disturbs the system of evidence.

2.1 *The law as is*

How does the Dutch criminal law system currently regulate evidence, and how does digital video evidence fit within this system?

A distinction is made between four types of systems of evidence. These are the negative legal system of evidence, the positive legal system of evidence, the free system of bare conviction, and the free system of reasoned conviction. The free systems of conviction have no limitations to what is considered lawful evidence, it is only of importance whether the evidence convinces the court. A free system of reasoned conviction requires that the court motivate its conviction. Does this obligation not exist, then it is a free system of bare conviction.[21]

The Dutch law uses a negative legal system of evidence within its criminal law system.[22] The legal systems are bound to the types of evidence as written in the law. The law will offer a limited list of acceptable types of evidence, and the court may not base its conviction on evidence that does not fall within these accepted types of evidence. The distinction between the positive and negative legal system of evidence lies with the power the court's appreciation has within the proces. In the positive legal system, the court is lawfully required to come to a guilty judgement when a certain amount of evidence is presented. To illustrate: a murder has occurred, and the prosecutor has presented three pieces of acceptable evidence. The law dictates that the court is required to come to a guilty verdict when three or more of such pieces of evidence are presented. The court might not be convinced of the suspect's guilt, but has to come to a guilty judgement based upon this evidence. Such a requirement does not exist within the Netherlands, making it a negative system of evidence. This negative legal system of evidence forms the first defining character of the Dutch system of evidence.

The second defining characteristic is that the Dutch criminal system of evidence contains several rules regarding the so-called 'minimum of proof'.[23] These rules somewhat bind the

[19] Wetboek van Strafvordering 1921.

[20] Commissie modernisering opsporingsonderzoek in het digitale tijdperk, *Regulering Van Opsporingsbevoegdheden In Een Digitale Omgeving* (report, 2018), ch. 1, 8.

[21] G.J.M. Corstens and M.J. Borgers, *Het Nederlands Strafprocesrecht* (8th edition, Kluwer 2014), p. 757.

[22] Wetboek van Strafvordering 1921, Art. 338.

[23] Commissie modernisering (n20).

court in their appreciation of the evidence. They generally see to the requirement that the evidence presented originates out of more than one source. This is mostly evident in testimonial evidence, as well as material evidence related to it. A witness statement on its own can as such not be used to come to a guilty verdict, no matter how convincing the statement is.[24] Such a system is rather unique, being employed in this manner solely in the Netherlands.[25]

The third defining characteristic consists of the freedom of appreciation that is due to the court.[26] The court is granted a large amount of freedom of appreciation to interpret the legal evidence presented. There are no obligations to come to a guilty judgement once certain conditions have been met, instead leaving this to the appreciation of the court. There is however an obligation to judge innocent when the minimum of proof is not met.[27] If evidence is contested, an obligation lies upon the court to substantiate the motivation. There are some exceptions to this duty to state reasons, such as facts of common knowledge. Such facts can for example include natural phenomenon or the public character of roads,[28] and generally consist of facts that are ought to be known or are easily learned through publicly accessible sources.[29] Within the acceptable evidence as illustrated by the law, it is the court's appreciation that determines the final verdict. If if the evidence does not convince the court of the guilt of the suspect beyond a reasonable doubt, the suspect is to be spoken free.

The fourth and final defining characteristic consists of the accusatory nature of the trial.[30] The burden of proof lies with the prosecutor, not with the suspect. The burden of proof lying with the suspect is found to be incompatible with the assumption of innocence. This characteristic mainly sees to this burden of proof, and thus falls outside of the scope of this thesis.

These four characteristics from the basis of the Dutch criminal system of evidence, forming the foundation for the rules as described within the CPC. The articles 338 through 344a CPC specifically see to the use of evidence. Art. 338 CPC states that the judge can only come to a guilty judgement if he is convinced by the lawfully acceptable evidence.

Evidence has to be both lawful and convincing. For evidence to be lawful, it has to be both reliable and acceptable. The reliability sees to its authenticity and its integrity. The evidence has to connect to the suspect, and the evidence has to be accurate. The acceptability sees to the negative legal system, limiting the acceptable evidence to several categories. Art. 339 CPC states all acceptable types of evidence, with art. 340 through 344a CPC further specifying these acceptable types. A total of five acceptable types are listed, as will be discussed below.

The personal observation of the judge forms the first accepted type.[31] The personal observation of the judge consists of three elements.[32] First of all, the observation consists of whatever the judges themself observes. The observation is broad in nature, and can include

[24] Wetboek van Strafvordering 1921, Art. 342(2).

[25] M.J. Dubelaar, Betrouwbaar getuigenbewijs (Kluwer 2014), p. 207-209.

[26] Corstens (n21), p. 758

[27] Dubelaar (n25).

[28] Memorie van Toelichting, Ontwerp tot vaststelling van een wetboek van strafvordering, der Koningin aangeboden door de Staatscommissie voor de herziening van het Wetboek van Strafvordering, ingesteld bij Koninklijk Besluit van 8 April 1910, no. 17, deel 1 Ontworpen Wetboek (Landsdrukkerij 1913), p. 82.

[29] HR 11 januari 2011, ECLI:NL:HR:2011:BP0291, NJ 2011/116.

[30] Corstens (n21), p. 758.

[31] Wetboek van Strafvordering 1921, Art. 339(1)(1) jo 340.

[32] Wetboek van Strafvordering 1921, Art. 340.

all senses. For example, an observation can be made of a nervous reaction by the suspect, but can also consist of the act of looking at a video presented during the trial. Such an observation generally does not have to be made known.[33] Second, only the observation by the judge themself is valid evidence. Finally, the observation must be made during the trial itself. If an observation falls outside of the trial, it cannot be used as lawful evidence. This is a strict requirement, as for example an observation made just after adjourning the court was deemed as unacceptable.[34]

The suspect's statement forms the second accepted type.[35] It consists of a statement made by the suspect during the trial, about the circumstances of the case, which they themselves have observed or experienced. Observing sees to the sensory external observations made by the suspect, such as seeing, hearing, or tasting. Experiences sees to feelings experienced by the suspect themself, such as experiencing pain, sadness, or anger. The statement can also consist of something the suspect has said outside of the trial about his experiencing of the circumstances of the case.[36]

A suspect's statement is subject to the minimum of proof, and is as such by itself not enough to come to a guilty judgement. The minimum of proof requires at least one additional piece of evidence, originating from another source, to support the witness statement. With the suspect's statement not being enough to come to a guilty judgement, it will be less appealing to put undue pressure upon the suspect to obtain a confession.[37] Though this supporting evidence should generally come from another source than the suspect, it is not defined what this should consist of otherwise. In practice it is seen that supporting evidence as such can be rather superficial, with for example the confession of a theft supported by the statement that the good belongs to the victim being found as satisfactory.[38]

The witness' statement forms the third accepted type of evidence.[39] It consists of a statement made by the witness during the trial, about the circumstances of the case, which he himself has observed or experienced. The observation and experience is the same as for the suspect's statement, referring to the observation and the experience. The statement is limited to these. The witness is not allowed to guess as to the motivations of the suspect, nor as to their actions, as this falls within the domain of the judge.[40]

Unlike the suspect's statement, the witness' statement does not contain a paragraph regarding the use of the witness' statement made outside of the trial. The law as is would thus not allow the use of the witness' statement in such a way. In practice however an important workaround has been implemented in the 'De Auditu'-case.[41] A testimonium de auditu consists of a witness stating that he has heard someone else state something, essentially being alike to hearsay. As such, hearsay evidence is lawful evidence. Judges however generally value such hearsay evidence as of low value.

[33] The exception to this is when the other party can't reasonably be expected to have been prepared for the observation. This is rather casuistic and up to the judge. See: HR 15 december 2009, ECLI:NL:HR:2009:BJ2831, NJ 2011/78.

[34] HR 29 augustus 2006, ECLI:NL:HR:2006:AX6414, NJ 2007/134.

[35] Wetboek van Strafvordering 1921, art. 339(1)(2) j° 341.

[36] Wetboek van Strafvordering 1921, art. 341(2).

[37] L.C. Besier and A.J. Blok, *Het Nederlandsche Strafproces Deel II* (Tjeenk Willink 1925), p.141.

[38] HR 30 juni 2009, ECLI:NL:HR:2009:BH3704, NJ 2009/495.

[39] Wetboek van Strafvordering 1921 art. 339(1)(3) j° 342.

[40] Corstens (n21), p. 778.

[41] HR 20 december 1926, ECLI:NL:HR:1926:BG9435, LJN BG9435.

The expert's statement forms the fourth accepted type of evidence.[42] It consists of a statement made by an expert in the relevant field, during the trial, about what their knowledge and field teaches about the facts and circumstances of the case. An expert is a person with specific knowledge in a certain field, allowing them to provide relevant information or do relevant research.[43] The statement made by the expert must relate to their field, and the expert is under oath when they make the statement.[44] The opposite party is allowed to oppose the expert's findings with their own.[45]

Written documents form the fifth and final accepted type of evidence.[46] The article detailing the written documentation gives a list of five subtypes of accepted documentation. Of these five types, four see to specific documentation, whereas the last category sees to any documentation not covered by the previous four. The first four types in order are: judgements and decisions, official police reports and other such documents made by investigating officers and authorized persons containing their observations and experiences regarding the facts and circumstances, official documentation from public authorities regarding their relevant domain, and expert witness' written reports about their findings. The fifth subtype includes all other written documents, but does require additional supporting evidence.

These four characteristics and five types of valid evidence form in essence the core of the Dutch criminal evidence system. Of course, beyond this basis things can get rather complex rather quickly, due to the almost a hundred years of continual development of the law. The Dutch criminal procedural code was set up in 1926, after a twelve year development.[47] This code is still in use today, having seen many amendments, adaptations and additions due to the changing world.[48] Digital evidence is a newer development coming forth from this changing world, logically not having been available at the inception of the criminal procedural code. In a world that is more and more digitized, evidence too follows this trend. [49] The widespread increase in the use of mobile phones, computers, and the internet has lead to a multitude of sources that can generate digital evidence. For example, a mobile phone can tell you a lot about someone's life, revealing location, contacts, interests, and many more personal information.[50] This evidence can provide vital information to the case.

Digital evidence however does not have its own traditional location within the system of evidence, and as such cannot be directly submitted as evidence. Criminal procedural codes are often complex structures, safeguarding the rights of all parties involved in the criminal process. As such, they are difficult to completely renew, and are often behind on modern developments. Digital evidence has been incorporated into many jurisdictions, often by applying the old rules to the new subject. This too is largely the case when it comes to evidence in the Dutch criminal procedure.

Digital evidence, in the same manner as photographic evidence or a murder weapon itself, does not fall within one specific category of acceptable evidence, as described above. It

[42] Wetboek van Strafvordering 1921 art. 339(1)(4) j° 343.

[43] Wetboek van Strafvordering 1921 art. 51i(1).

[44] Wetboek van Strafvordering 1921 art. 51m(2).

[45] European Convention of Human Rights 1950, art. 6.

[46] Wetboek van Strafvordering 1921 art. 339(1)(5) j° 344.

[47] Commissie Modernising Opsporingsonderzoek in het Digitale Tijdperk (n19), paragraph 1.1.

[48] For example: Wet Computercriminaliteit I 1993; Wet Computercriminaliteit II 2006.

[49] M. Dubelaar and G. Vanderveen, 'Beeld en geluid in het strafproces: Implicaties van de opkomst van (audio)visuele technieken en materialen voor communicatie en besluitvorming in de strafrechtspraktijk' (2009) *Nederlands Juristenblad* 1530.

[50] B.J. Koops, 'Privacy Spaces' (2018) *West Virginia Law Review* 121.

can't be submitted directly to the court as it is not listed as an accepted type of evidence, but this can be circumvented in multiple ways. First of all, digital evidence can fall under the personal observation of the judge. For example, if one of the parties has digital video evidence, they can simply display the video to the judge during the trial. The judge makes an observation of the display, which can then be used as valid evidence to come to a judgement. Second, an investigating officer can detail the video in an official report. The evidence then consists formally of a written document, yet materially contain the digital evidence due to the transcription. Third, the digital video file as is can be paired with an expert's statement or written report validating the authenticity of the evidence. This is not a limited list, and other possibilities are also open. For example, it would theoretically be acceptable to have a witness testify about observing the contents of a video.

There are no additional requirements made of digital evidence by the law.[51] When a police or an expert's report has been submitted detailing a digital video file, the judge is under no obligation to look at the video itself. There are no additional rules regarding authenticity, nor trustworthiness of the material.[52] This is in contrast with how evidence is normally handled within criminal law.

In practice too, there seems to be a lack of regulation when it comes to digital evidence. Digital evidence is mostly collected and safeguarded by forensic experts. This is handled by the Dutch Forensic Institute. The Forensic Institute has published so-called 'Forensisch-Technische normen'; forensic standards describing how to handle evidence after a crime.[53] Unless they are directly derived from the law, these standards are not legally binding. Following the standards however does strengthen the integrity of the evidence collected. There is a standard when it comes to recordings of calls, as well as a concept proposal for mobile phones, but there are no general standards yet seeing to digital evidence, nor for digital video evidence specifically. As such, digital evidence does not have any special legal or non-legal requirements.

2.2 <u>The heralding deepfake</u>
What are deepfakes and synthetic video generation?

The first deepfake was published on November 2nd 2017, on a social media platform called Reddit. Reddit is an American social news aggregation, web content rating, and discussion website, relying largely on user-submitted content and discussion.[54] It is currently the sixth most visited site in the world by unique views, as well as being the third most visited site in the United States, passing giants as Facebook and Twitter.[55] Reddit allows the user to create so-called 'subreddits', sub-forums with a specific theme, goal or subject. It is in one of these subreddits that the first publications of deepfakes occurred.

On November 2nd 2017, the user 'Deepfakes' created one such subreddit (bearing the same name) dedicated to sharing his content with others.[56] The deepfakes he created were mostly

[51] Dubelaar (n49).

[52] Dubelaar (n49).

[53] NFI, 'Factsheet' (2007) <http://www.fomat.nl/nfi_fsnormen.pdf> Accessed on 26 May 2018.

[54] Reddit Originals, 'How reddit works' (30 July 2014), <https://redditblog.com/2014/07/30/how-reddit-works-2/> Accessed on 27 May 2018.

[55] Alexa top 500 global sites and US sites. <www.alexa.com/topsites> Accessed on 27 May 2018, <www.alexa.com/topsites/countries/US> Accessed on 27 May 2018.

[56] <www.reddit.com/r/deepfakes> Accessed on 14 June 2018.

of explicit content, and were according to him made through the use of open-source libraries, namely Keras and TensorFlow. On 8 January 2018 the user "DeepFakeApp" set up his own subreddit, and used this to launch FakeApp, an application aimed at providing a desktop tool for creating photorealistic face swap videos using AI.[57] In a month, a smaller community of about 2500 gathered here. It was not until the 25th of January that the user 'Derpfakes' caused an explosion in popularity and exposure by posting several movie clips in which the actors were replaced with the actor Nicholas Cage. Several days later, the first blanket ban for deepfakes occurred on the hosting provider Gfycat.com due to the questionable ethics surrounding the explicit content creation.[58] On February 7th 2018 and onward, reddit came out with a statement relating to an update of its policies, and with this update applied a ban to deepfakes aimed at explicit content.[59] Whilst this move is fully understandable, it in practice however also lead to the termination of non-explicit subreddits dedicated to the technology. With this move, the mainstream exposure to deepfakes mostly ended. The communities however seem to have merely migrated from Reddit to less moderated and visible places, such as imageboards and private forums.

Deepfakes are an early form of synthetic video generation. Synthetic video generation is the artificial generation and alteration of falsified digital imagery at a highly realistic level.[60] It is not restrained to only video, but as discussed before can also include images and audio.[61] Synthetic video generation makes use of machine learning, a specific subset of artificial intelligence. The idea of machine learning is about as old as computers themselves are, with the first mention originating in 1959.[62] However, due to the trial-and-error nature and the subsequent high power requirement of the process, it has not been feasible to employ machine learning at a large scale until recent years.

Currently, most synthetic video generation is done through the use of an autoencoder in a generative adversarial network ['GAN']. An autoencoder is a unsupervised artificial neural network, aimed at taking an input and turning it into a unique and reversible code.[63] A generative adversarial network uses two neural networks in contest with each other in a zero-sum game.[64] One network creates and the other network controls whether it was done right, giving feedback to the generating network.

[57] <www.reddit.com/r/fakeapp> Accessed on 14 June 2018, <www.fakeapp.org> Accessed on 14 June 2018.

[58] S. Cole, 'AI-Generated Fake Porn Makes Have Been Kicked Off Their Favorite Host: Reddit is still silent' (Motherboard, 31 January 2018). <https://motherboard.vice.com/en_us/article/vby5jx/deepfakes-ai-porn-removed-from-gfycat> Accessed on 17 June 2018.

[59] Reddit, 'Update on site-wide rules regarding involuntary pornography and the sexualization of minors' (Reddit, 7 February 2018). <www.reddit.com/r/announcements/comments/7vxzrb/update_on_sitewide_rules_regarding_involuntary/> Accessed on 17 June 2018.

[60] R. Chesney and D.K. Citron, 'Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security' (2019 forthcoming) *California Law Review* 107, p. 2-4. <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3213954> Accessed on 1 November 2018.

[61] Chesney (n60), p. 4.

[62] A.L. Samuel, 'Some Studies in Machine Learning Using the Game of Checkers' (1959) *IBM Journal of Research and Development* Vol. 3, Issue 3.

[63] Oberoi (n1).

[64] I. Goodfellow and others, 'Generative Adversarial Nets" (2014) *Advances in Neural Information Processing Systems* 27 (NIPS 2014).

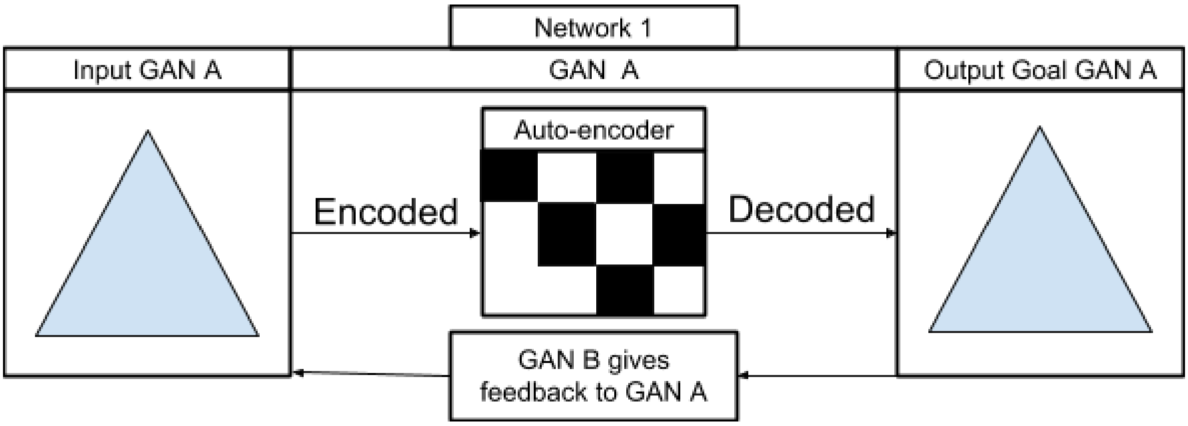

Figure 1: a single network

To illustrate it a bit less technically: an image of a triangle is given as input, and the output goal is the same triangle. GAN A first encodes the image into a unique code, then decodes that code again to the same triangle. GAN B judges on how well the output compares to the input. In doing so, GAN A over time learns what elements of the triangle should be encoded to still obtain an acceptable triangle when decoded. A different image input results in a different code, which then decodes back into the different image to obtain the output goal.

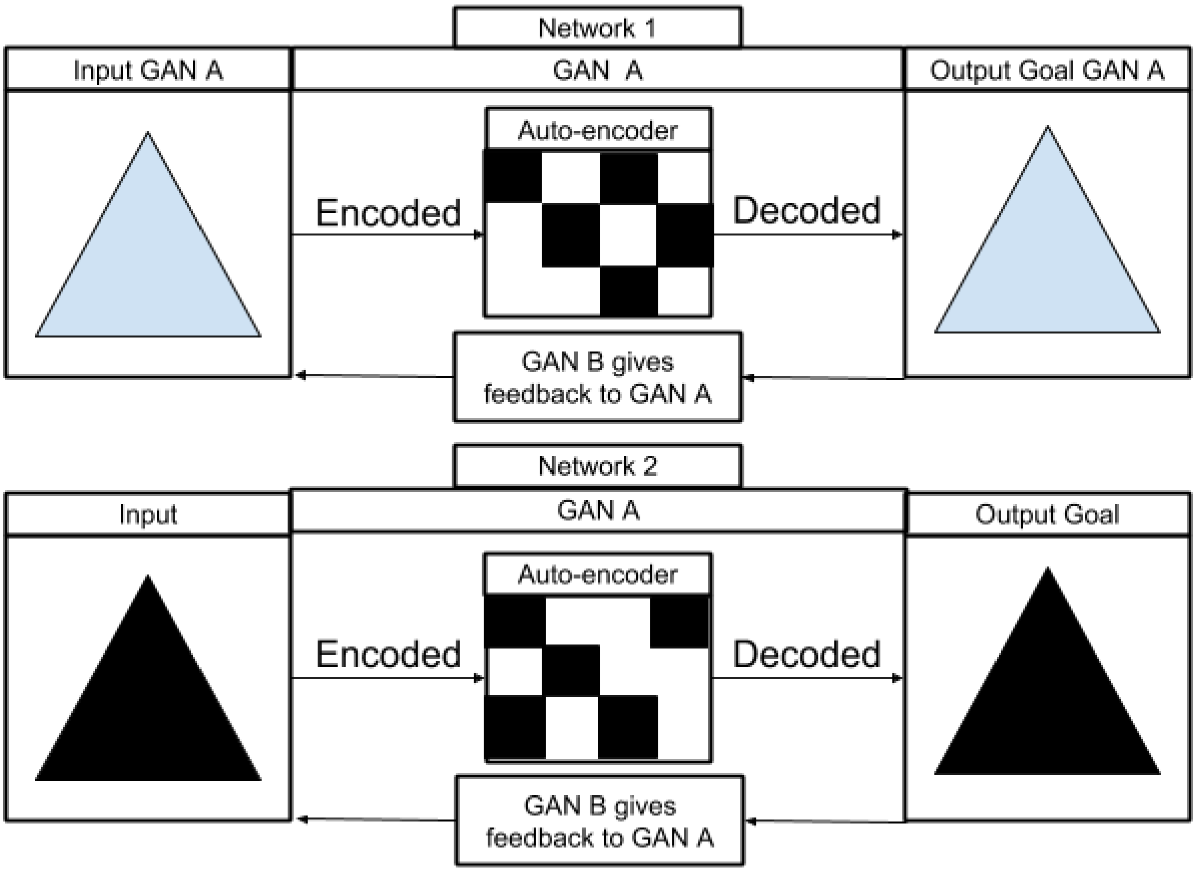

Figure 2: two seperate networks

Deepfake-technology at the moment uses a combination of these networks. After training the network, a black triangle is requested as the output goal instead of a white triangle. GAN A is trained in making white triangles, and will try to make the black triangle with this knowledge. GAN B then gives feedback to how well GAN A remade the output goal. This will eventually result in GAN A combining both the white triangle and the black triangle into an acceptable output, such as a grey triangle.

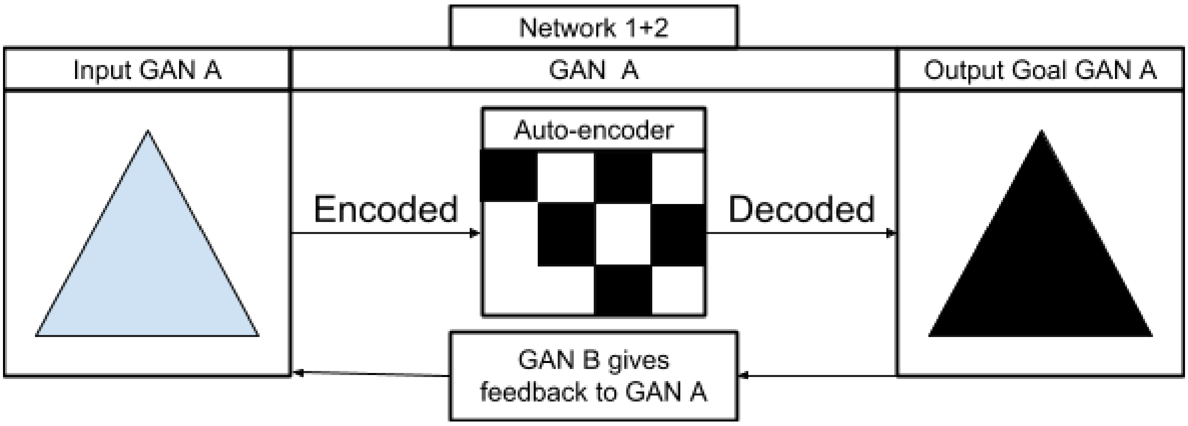

Figure 3: combining the networks

This application of GANs allows for self-improving systems with highly realistic outcomes, and will likely form the basis of synthetic video generation in the future as well. It enables realistic generation of large amounts of footage at a reduced time and price.

Within the current stage of the development of the technology, this has not caused any problems when it comes to evidence yet. Synthetically generated videos are still generally easy to spot, having certain telltale signs. These signs range from image artifacts to specific faults in the algorithm leading to persons depicted in generated videos not blinking as often as they should.[65] However, developments in computer science are generally fast-paced. An example of this are the leaps made between Face2Face[66], published in 2016, and Deep Video Portraits[67], published in 2018. Compared to Face2Face, Deep Video Portraits allows not only for the transposition of expressions, but also the transfer of head pose, eye direction and blinking.

The limitations of the technology lie mostly within the available computing power and knowledge.[68] The process requires a large amount of computing power to synthetically generate realistic imagery. This is however seen from our current standing. Looking at the progress made within the development of computer hardware, the limitation of power is only a matter of time. The other limiting factor is the available knowledge, mainly within the field of generative networks.[69] Generative networks are however a hot topic, and this technology in general developing further and becoming more commonly accessible will too further development of synthetic video generation.[70] Synthetic video generation also seems to have private interest. This is already visible in the entertainment industry, with NVidia's recent exploration in applying synthetic generation to generate realistic faces, which can be used in commercial products.[71] These private interests can further support growth of the technology. An increase in power, as well as an increased sophistication of generative networks will decrease the limitations of synthetic video generation. It is as such not unthinkable that synthetic video generation matures in the near future , leading to falsified videos indistinguishable from real ones to both the human and mechanical eye.

[65] Y. Li, C. Lyu and S. Lyu, 'In Ictu Oculi: Exposing AI Generated Fake Face Videos by Detecting Eye Blinking' (2018) University of Albany <https://arxiv.org/abs/1806.02877> Accessed on 31 October 2018.

[66] Thies (n11), p. 2387-2395.

[67] H. Kim and others, *Deep Video Portraits* (Paper, Stanford University, 2018) <https://web.stanford.edu/~zollhoef/papers/SG2018_DeepVideo/page.html> Accessed on 31 October 2018.

[68] Chesney (n60), p. 8-9.

[69] Chesney (n60), p. 8-9.

[70] Chesney (n60), p. 8-9.

[71] T. Karras, S. Laine, T. Aila, 'A Style-Based Generator Architecture for Generative Adversarial Networks' (12 December 2018) Cornell University <https://arxiv.org/abs/1812.04948> Accessed on 28 December 2018.

2.3 Identifying the disruption

How does synthetic video generation disrupt the Dutch regulation of digital video evidence?

When looking at the requirement of lawfulness, synthetic video generation causes a disruption within the reliability of the evidence. The requirement of reliability is primarily disturbed by the technology of synthetic video generation. The reliability of evidence consists of the authenticity and the integrity of the evidence. Respectively, in what manner the evidence relates to the suspect, and whether the evidence is accurate and displays what truly happened. The reliability of evidence is an important matter in the criminal procedure. Unreliable evidence can lead to wrongful convictions. This is generally seen as unacceptable, due to the ramifications on the life and rights of the wrongfully convicted.

Synthetically generated videos have inherent problems when it comes to integrity, as they do not display the truth as it truly happened. As such, they should logically not be allowed to be presented as depictions of said truth. The difficulty here however lies within the detection of such evidence, specifically synthetically generated videos. Forgeries are getting more difficult to detect with advancing technology, and synthetic video generation has the potential to slip by due to the pre-eminent reliability video has.[72]

Digital video evidence is currently deemed pre-eminently reliable.[73] Everyone is aware of the possibility of image manipulation, with brands such as PhotoShop being common household names.[74] Realistic video editing on the other hand is much more rare, mostly reserved to Hollywood studios and others with deep pockets. Realistic video editing is both difficult and expensive, and even then still is easy to detect. Even with highly realistic traditional video generation, the uncanny valley-effect often reveals something isn't quite right with the imagery.[75] As such, people are quick to assume that what they see in a video is real. The costs of creating or traditionally altering such a video are often too high for it to be an option. If the video is traditionally altered or created, then it is often easy to tell whether it is real or not. This difficulty to create realistically altered or generated video is however quickly fading with the development of synthetic video generation.[76]

The requirement of acceptability strengthens the disruption observed within the requirement of reliability. The legal system as is has no separate category for digital evidence, despite its growing importance. Instead, digital evidence is submitted through the currently available types of lawful evidence. These accepted types of evidence are however not made for digital evidence, and do not foresee possible complications that digital evidence might bring.

The personal observation of the judge is one of the more vulnerable categories when it comes to synthetically generated videos. The judge is not under an obligation to be knowledgeable about what is observed. The law brings little guarantees when it comes to digital evidence, as there are no special requirements. Meanwhile, visual footage has a large effect on human beings, eliciting empathy with the victim and apathy with the suspect. With video evidence, the line between merely observing and experiencing is a blurred one.

[72] Chesney (n60), p. 5.

[73] W.D.H. Asser, *Asser Procesrecht 3 Bewijs* (2nd revision, Kluwer, 2017), p. 2017/249.

[74] Chesney (n60), p. 5.

[75] M. Mori, 'Uncanny Valley' (1970) *Energy* 7(4), p. 33-35.

[76] Chesney (n60), p. 7.

[77] The judge is human too, and cannot escape the effects that video evidence is shown to have on humans. Despite this, a guilty verdict can be based majorly on the personal observation of the judge.[78] If the judge as such is convinced by synthetic video evidence, the law does not limit the judge to come to such an unacceptable outcome.

The expert witness is allowed to give information about the technology, but is not allowed to go outside their field. Such a judgement would fall within the domain of the judge, as it assesses the weight of the evidence.[79] The expert witness however can be called in to verify the integrity of the evidence in person or in a written report. With the advancement of synthetic video generation, it is however not unthinkable that this will become increasingly difficult. In recent years, forensic experts have been struggling with altered photography.[80] It would be only logical that as the technology improves, synthetic video generation too will become more difficult to detect overtime. With the quick development and inherent power of generative networks, perhaps a day might come in which the expert cannot distinguish between a real and synthetic video. The expert witness' assessment of the reliability might as such be only a temporary bandaid.

The police report is an important piece of evidence within the Dutch criminal system, even carrying with it a special exemption to the minimum of proof.[81] The police report is an official document and contains all steps the investigating officers have taken in the investigation, containing all the facts observed and experienced by the investigating officer. A vulnerability here lies once again in the person observing the synthetically generated video. The investigating officer describes the video within their report, and this description is then acceptable evidence at face value. The investigating officer however lacks both the knowledge of the expert, as well as the expected capability of detecting inauthentic evidence of the judge. Through the police report, these already weakened safeguards are circumvented.

The requirement that the evidence is convincing does seemingly not hinder the disruption caused by synthetic video generation. Video evidence is a generally influential type of evidence. In Douglas e.a. 1997 it was found that showing pictures, video files and animations to judges and jury members leads to an increase of guilty convictions.[82] In Whalen & Abrams 2007 it was shown that the showing of such visual evidence also leads to higher compensation of damages.[83] In Kassin & Garfield 1991, it was shown that after showing a video about a murder case, evidence was more often accepted as being beyond a reasonable doubt.[84] Visual stimuli, such as video, are quick to generate an emotional response, both of empathy towards the victim as well as apathy towards the suspect.[85] There is no discernable variation in this effect, and legal experts and judges seem to also be susceptible. As such, visual evidence seems to have an inherent weight to it.

[77] A.M. van Woensel and F. van Laanen, 'Commentaar op Artikel 340' in M.S. Groenhuijsen and others (eds.), *Losbladige commentaar op het Wetboek van Strafvordering / Melai* suppl. 163.

[78] HR 3 juli 2007, ECLI:NL:HR:2007:BA4994, NJ 2007/412.

[79] Y. Buruma, 'Betrouwbaar Bewijs' (2009) *Delikt en Delinkwent* 23.

[80] Chesney (n60), p. 5.

[81] Wetboek van Strafvordering 1921, art. 344(2).

[82] K.S. Douglas, D.R. Lyon and J.R.P. Ogloff, 'The impact of graphic photographic evidence on mock jurors' decisions in a murder trial: Probative or prejudicial?' (1997) *Law and Human Behavior* 21(5), p. 485-501.

[83] Dubelaar (n49).

[84] S.M. Kassin, and D.A. Garfield, 'Blood and guts - General and trial-specific effects of videotaped crime scenes on mock jurors'(1991) *Journal of Applied Social Psychology,* 21(18), p. 1459-1472.

[85] Kassin (n84).

2.4 Conclusion

The law as is places digital evidence into a system that is increasingly unfit for it. Digital evidence is not allowed to be submitted as is, but is instead inserted into the system using the currently available means. This system is threatened by the development of synthetic video generation, heralded by the public surge of deepfakes. The developmental cycle of synthetic video generation has only recently started, and is bound to increase over time due to diffusal of its limiting factors.

Synthetically generated videos are inherently unreliable, as they do not display the truth accurately. Evidence is required to be both reliable and convincing. Despite being inherently unreliable, the difficulty of detection as well as the current inherent reliability of video lead to the evidence being possibly accepted. The system of evidence as written in the law worsens this problem by having digital evidence be acceptable through different means without additional safeguards or requirements. The requirement of the evidence being convincing does not lessen this problem, as videos seem to have an inherent convincing effect to them.

This raises the question whether the current system is indeed capable of handling synthetical video evidence. To determine this, we must first find out how the judge determines whether evidence is reliable or not. In other words, how does the judge determine the reliability of the evidence, and is this applicable to digital video evidence?

Chapter 3 Assessing reliability

What makes evidence reliable, and how does this affect digital video evidence with regard to the disruption by synthetic video generation?

The rise of synthetic video generation is problematic due to the way video is regarded as inherently reliable. At the rate of development currently behind synthetic video generation, it is not unthinkable that such content might find its way into the criminal system of evidence before this inherent reliability wears off. It is also questionable whether the parties involved in validating this reliability are capable of doing so with synthetic video evidence.

To properly observe the problem at hand, we must first look at what it is that makes evidence reliable. There are many meanings of 'reliable', ranging from the everyday use of 'being able to be relied on as honest or truthful'[86], to the scientific 'giving the same result on successive trials'[87]. Translation too proves difficult, as the Dutch word for reliability can also mean trustworthiness or dependability. Whereas there is no codified meaning of the word, within Dutch law the reliability of evidence consists of its authenticity and integrity. The authenticity relates to whether or not it truly links the evidence to the suspect. The integrity relates to the value of the evidence itself. The authenticity and the integrity relate the most to the accuracy of the evidence, and whether the evidence displays the truth. In other words, how well the evidence displays what truly happened. Unreliable evidence goes against these values, and as such should not be allowed.[88]

The judge has the freedom to assess the value of evidence within the borders laid out by the law. When it comes to the reliability of evidence, the Dutch criminal procedural code however does not specify what this reliability is, nor when it applies. This is left to the judgement of the court. The judge is deemed capable of fully determining whether evidence submitted is reliable or not. In practice, there is however surprisingly little known about exactly how the judge values the reliability of evidence.[89]

When it comes to the acceptation of evidence, the judge has an obligation to motivate why they have accepted evidence when this is disputed by the opposing party. Using these motivations, it is possible to get an indication what convinced the judge of the reliability of the evidence. It is however not a perfect method, as the indications seem to largely skew towards testimonial evidence. This is a logical disbalance, as it are mostly testimonial statements that are disputed due to their inherent proneness to unreliability. Due to this prevalence, existing research too largely focuses on testimonial authenticity. As such, this is to be kept in mind when reading these findings.

First, a framework will be set up as to analyze the assessment of the reliability of evidence. The framework consists of indicators of reliability: aspects of the evidence that can give the judge an indication of whether it is reliable or not. This framework will then be applied to synthetically generated video evidence to see whether the judge is capable of assessing the reliability of this type of evidence.

3.1 *The judge's tools*

How does the court assess the reliability of evidence?

[86] Thesaurus.com, <https://www.thesaurus.com/browse/reliable> Accessed on 16 July 2018.

[87] Merriam Webster, <https://www.merriam-webster.com/dictionary/reliable> Accessed on 16 July 2018.

[88] HR 30 maart 2004, ECLI:NL:PHR:2004:AM2533, NJ 2004/376, r.o. 3.6.4; HR 29 januari 2013, ECLI:NL:HR:2013:BY0816, NJ 2013/414.

[89] L. Stevens, 'Bewijs Waarderen' (2014) *Nederlands Juristenblad* 40, p.2844-2845.

By law, the judge is the final arbiter when it comes to the authenticity of the evidence.[90] The court must be convinced of the indicted by the presented evidence. For the evidence to be acceptable, it must be reliable. Over the years, the judge has evolved a toolkit of sorts, containing different aspects of the evidence that are repeatedly observed in the court's motivations on the reliability of evidence.

In my own observations I have found that three of these indicators are almost always observed, whereas the others play a more optional role depending on the case and the nature of the evidence. There are three primary indicators of reliability by which the court can determine the authenticity of the evidence. These are the accuracy of the evidence, the consistency of the evidence, and the completeness of the evidence. These three indicators often influence each other, and as such looking at one requires a look at the others. These indicators will be discussed first.

The accuracy of the evidence forms the first primary indicator of reliability. It sees to how much the evidence relates to the truth.[91] It forms the prime interest of the judge. A distinction is made between the partial accuracy and the overall accuracy of evidence. The partial accuracy refers to the distinct parts that a piece of evidence is made up of. The overall accuracy refers to the accuracy of the evidence as a whole. A piece of evidence can be partially inaccurate without being wholly inaccurate.

For example, person A. stands accused of murder. An expert witness however has written an expert's report about the authenticity of his video evidence, stating that it is wrong. Unbeknown to the expert, their expert's report is however partially based on outdated information that has since been proven wrong. The report is partially inaccurate, but not wholly. The court is not allowed to use the untruthful part to base any of their argumentation or judgements on. It is up to the court to determine the weight of the report as a whole, and whether the inaccurate part devalues this weight.

The Dutch High Court has confirmed the existence of this indication and its role within the assessment of the reliability of the evidence.[92] For evidence to be acceptable, it has to be believable and corresponding to the truth.[93] The High Court however does not indicate what the reliability it wants to see is, nor how the court should come to such a conclusion.[94] Even such an elementary view on the use of evidence however brings complications within a system as complex as procedural criminal law, as it essentially forbade the use of statements made by the suspect or witness that were deemed untruthful by the court. This in turn clashes with art. 341 CPC, of which the explanatory memorandum indicates that any such statement made should be available for the court to use as they deem necessary.

The consistency of the evidence forms the second primary indicator of reliability.[95] The facts and circumstances are generally presented in a certain way by the evidence, and the consistency sees to both the internal and external consistency. The internal consistency refers to the consistency of the evidence in isolation. An example of this would be that of the witness contradicting themselves within their statement. The external consistency sees to the consistency of the evidence in relation to other evidence. If for example security camera footage shows the suspect walking on the street at a certain time, but all other

[90] Wetboek van Strafvordering 1921, art. 338.

[91] Dubelaar (n25), p. 166.

[92] HR 14 september 1992, ECLI:NL:PHR:1992:AC3716, NJ 1993/54; HR 23 september 2008, ECLI:NL:HR:2008:BD3902, NJ 2008/525.

[93] HR 14 september 1992, ECLI:NL:PHR:1992:AC3716, NJ1993/54.

[94] Stevens (n89), paragraph 3.1.

[95] Dubelaar (n25), p. 168-170.

evidence points towards the suspect being somewhere else at that time, then there is a clear external inconsistency between the evidence.

Whereas material evidence is generally internally consistent due to its nature, testimonial evidence can differ greatly in consistency. Testimonies coming from a single source are likely to at least contain a few inconsistencies.[96] This is due to various reasons, including the mental state of the witness, the manner in which the questions are asked, as well as the general working of the memory.[97] These testimonial inconsistencies are split within two categories: incidental and intentional.

The incidental inconsistency sees to the inconsistency within testimonial evidence created without the testifier willingly doing so. It generally does not greatly influence the reliability of the testimonial evidence, but can still give an indication of reliability. Some manner of incidental inconsistency will also almost always be apparent in testimonial evidence. Witnesses can forget or misremember parts, as well as be guided by leading questions on what to remember.[98]

The intentional inconsistency sees to the inconsistency within testimonial evidence created willingly by the testifier. An example of this would be the suspect who changes his statement after being presented with contradicting evidence. Intentional inconsistency can influence the reliability of the evidence in a greater manner than the incidental inconsistency, as it shows a willingness to provide inaccurate information.

The completeness of the evidence forms the third major indicator of reliability.[99] It refers to in what manner the evidence incorporates all facts, circumstances, and details. The more the evidence shows about the occurrence, the more complete it is considered. This completeness can be of influence on the assessment of the consistency and accuracy as well. Details can for example play an important role in determining the external consistency of evidence. If specific details are confirmed by multiple sources, there is a higher likelihood that the evidence is reliable.

The details shown in the evidence generally play a central role in assessing the completeness of the evidence. A distinction is made between central and peripheral details. Central details see to details entailing critical elements of the case, whereas peripheral details see to details that are not deemed critical.[100] For the court, evidence missing central details is generally a larger indicator of unreliability than evidence missing peripheral details.[101] This is not to say that the missing of central details is a good indicator of reliability, as what is deemed critical will differ for each person. A trained investigating officer will for example be much more likely to pick up details that the court deems important, whereas a witness might focus on something completely different in the heat of the moment.

Inconsistencies in details are generally not seen as as a sign of unreliability, but are regarded as mostly irrelevant.[102] This mainly goes for peripheral details. The inconsistencies in these details are often excused as naturally occurring, being the result of what the party deemed critical, as well as the working of the memory. The observing party is not expected

[96] Dubelaar (n25), p. 168.

[97] Dubelaar (n25), p. 168-169.

[98] G. Odinot, G. Wolters, and A. van Giezen, ‘Accuracy, confidence and consistency in repeated events’ (2013) *Psychology, Crime & Law* 19/7, p. 3.

[99] Dubelaar (n25), p. 170.

[100] Dubelaar (n25), p.171.

[101] Dubelaar (n25), p. 171.

[102] Stevens (n89), p. 2842.

to be perfectly aware of all details, nor to have a perfect memory. Minor discrepancies within the details of the story are as such not a large indicator of unreliability.

These three major indicators of reliability tend to be looked at in combination. These primary indicators are more general in nature. They are either discussed directly, or are present in an indirect manner within the motivation. As the truth is the main goal, the accuracy of the evidence always plays a role. The consistency and the completeness of the evidence both play a large role in determining this accuracy.

There are however another three factors that play a role when determining the authenticity of evidence.[103] These secondary indicators are the source of the evidence, the presentation of the evidence, and the creation of the evidence. Due to the mostly testimonial direction of previous research as well as motivations, these have a significant testimonial bias. As these still form part of the tools the court uses to determine the authenticity of evidence, they are however not without value when observing digital evidence.

The first category of secondary indicators of reliability sees to the source of the evidence. This refers to the party that delivered the evidence as is. When it comes to testimonial evidence such as the witness' statement, it is the witness themself that is considered the source of the evidence. When it comes to material evidence, it can be the party that submitted this evidence or the party that created this evidence. Within the observed research, the source of the evidence can be divided into three subcategories. These are the sincerity of the source of the evidence, the competence of the source of the evidence, and other elements relating to the source of the evidence.

The sincerity of the source of the evidence refers to the belief held by the submitter that the evidence is accurate, and as such a truthful representation of reality.[104] Logically the accuracy of the evidence is an important aspect, and reasons to believe that the source of the evidence is insincere can lead to a decrease of the perceived reliability of the evidence. To determine the sincerity of the evidence, it is of importance to determine the motivation of the source of the evidence.

This motivation can be determined in multiple ways.[105] The position of the source of the evidence itself can form an indicator. Evidence submitted by a co-defendant implicating the other suspect can for example be tainted by the wish of the source to unburden himself. A witness in a personal relationship with the suspect too has personal gain when it comes to unburdening the suspect. The attitude of the source of the evidence can also form an indicator.[106] If the suspect during the trial strategically decides to keep quiet and talk at opportune moments, as well as leave out certain information, then this can be seen as an indication of unreliability of the evidence. This is not limited to outright untruths, but can also be applied to conversational implicatures.[107] In such implicatures, the source speaks truthfully, but intends to create an inaccurate image for the receiver.

The competence of the source of the evidence refers to the ability of the source to have made the observance displayed within the evidence, understand this observance, and their capability of relaying this observance.[108] It sees mostly to testimonial evidence, relating to the capability of witnesses to relay their observed information. In this regard, two distinct groups are seen as inherently having a higher chance of delivering unreliable evidence. These are on one hand children, and on the other hand mentally-limited individuals. These

[103] Dubelaar (n25), p. 181-193.

[104] Dubelaar (n25), p. 182.

[105] Dubelaar (n25), p. 183.

[106] Stevens (n89).

[107] Dubelaar (n25), p. 182.

[108] Dubelaar (n25), p. 183.

are distinguished from the average adult due to their limitations in memory, speech, ability to think abstractly, as well as increased suggestibility.[109]

Other source-bound elements refers to any remaining elements of the source that do not fall within the above two types.[110] The literature speaks mostly of the impartiality of the source. In testimonial evidence, the manner in which the source made an observation also matters. An example given refers to the education of the source and the possibility of such expertise to influence the assessment of the reliability of the evidence.[111]

The second category of secondary indicators of reliability sees to the presentation of the evidence. The manner in which the evidence is presented has influence on the assessment of the reliability of the evidence. In literature, the presentation of evidence too can be divided in three categories. These are the emotion represented within the presentation of evidence, the assuredness of the presentation of the evidence, as well as the medium of the presentation of the evidence.

The emotion represented within the presentation refers to the manner in which emotion and behavior displayed within the delivery method influence the receiver. Logically, this generally refers to testimonial evidence. It can however find a way into other evidence as well. An example of this would be a heart-wrenching letter found at the scene, indicting the suspect. The emotions displayed in such a letter can have an influence on the court in its assessment of the reliability of the evidence. In testimonial evidence, both vocal and non-vocal elements play a role. Such elements include body language, facial expressions, or intonation. The court's task to try and discern whether someone is lying too falls within this category. Whereas research has indicated that there is no specific pattern of behavior indicating lying, it is still a natural reaction.[112]

The assuredness of the presentation refers to the manner of certainty with which the evidence is presented. In testimonial evidence, this mostly refers to the witness and the assuredness of their witness' statement. Strongly told statements tend to be more believable, as it suggests a certain knowledge of the truth. This has however been found incorrect, as the certainty of the statement is much more closely connected to the personality of the witness than to the statement itself.[113] Research has shown shown that there is only a negligible connection between the assuredness with which a statement is made and the accuracy of the statement itself.[114] Despite this, the assuredness of the presentation can still influence the assessment of the authenticity of the evidence. This is a seemingly human bias, and the court too seems to be not fully immune to it.

The medium through which the evidence is presented can also be of influence on the assessment of the reliability of the evidence.[115] It refers to the manner in which the receiver obtains the evidence. There is still relatively little known about how audiovisual technological developments influence the courtroom.[116] There does however seem to be a difference in the inherent reliability of the available mediums. As discussed before, photographs and videos evoke a stronger reaction than words alone due to their visual

[109] Dubelaar (n25), p. 184.

[110] Dubelaar (n25), p. 184.

[111] Dubelaar (n25), p. 184.

[112] Dubelaar (n25), p. 186.

[113] H.L.G.J. Merckelbach, I.E.L. Candel and H.F.M. Crombag, 'De Goede Getuige' (2003) *Trema* nr. 6, p. 217.

[114] G. Wolters and G. Odinot, 'Zijn zekere getuigen betrouwbare getuigen?' in P.J. van Koppen and others (eds.) *Reizen met Mijn Rechter: Psychologie van het Recht* (Kluwer 2010, p. 538.

[115] Dubelaar (n25), p. 189-193.

[116] N. Feigenson, 'Visual Evidence' (2010) *Psychonomic Bulletin & Review* 02, p. 149-154.

nature.[117] Pictures and videos are also seen as generally reliable by both the common man and the court.[118] The assessment of the reliability of written evidence too is subject to the choice of medium.[119] The manner in which a statement is written down is of influence on the assessment of the reliability.

The third category of secondary indicators of reliability sees to the creation of the evidence. The manner in which the evidence is initially created has influence on the assessment of the authenticity of the evidence. The literature describes two such types. These are the influences on the observation, as well as the influences on the source.

The influences on the observation see to any influences on the observation or experience.[120] These include for example the amount of light at the time of making the observance, the duration of the observance, as well as the surroundings in which the observance was done. [121] Witnessing a murder in an empty field leads to a different observation than witnessing a murder in a bustling city. There is a limit to the quality of the observation when it comes to humans, as well as a limited capacity of preserving and recalling these observations.The influences on the source see to any outside influences on the source itself. These can include influences that change the frame of reference, causing the person to interpret the evidence in a new light, but also things that are capable of altering the memory itself, such as leading questions. Humans have a limited capability of preserving and recalling memories, and these are often malleable.[122]

I personally distinguish a fourth type of the creation of evidence. This fourth category relates to the handling of the evidence. Traditionally this has seen to the handling of material and documentary evidence, validating the integrity of the evidence since it has been obtained by the investigating officer. With the introduction of digital evidence, this has become more important. With the development of synthetic audio generation, it will be important in regards to testimonial evidence as well.

The handling of the evidence is mainly determined by its lack of influences on the material itself. Whereas the creation of the evidence sees to how outside influences alter the source and the observation, the handling of the evidence sees to whether or not there have been any influences on the evidence itself during the investigation. This is generally done in two ways, by implementing a chain of custody, as well as by implementing a chain of evidence.

The chain of custody refers to the chain of people who have handled the evidence from the moment of collection. This includes the officer that collected the evidence, and the forensic researcher performing analytic research, but also the postman handling the evidence during transport. It is not relevant whether the person actively took part in the research, merely that the person had access to the evidence for an amount of time.

The chain of evidence sees to the evidence itself. It details where the evidence has been, as well as what has happened to it. A forensic researcher running two different tests on the evidence will for example have two actions in the chain of evidence, whilst having one in the chain of custody.

[117] N. Feigenson, C. Spiesel and R.K. Sherwin, *Law in the Digital Age: How visual communication technologies are transforming the practice, theory and teaching of law* (NYLS Legal Studies Research Paper No. 05/06-6, Barbados Group Working Paper no. 05-06, 2005) <https://ssrn.com/abstract=804424> Accessed on 2 August 2018.

[118] Asser (n73), p. 2017/249.

[119] Dubelaar (n25), p. 192.

[120] Dubelaar (n25), p. 136.

[121] Dubelaar (n25), p. 136-138.

[122] Dubelaar (n25), p. 138.

Through the use of these indicators the court is deemed able to assess the reliability of the evidence presented. There is however a logical bias towards testimonial evidence within these indicators. It is as such the question whether these indicators of reliability will function when it comes to digital evidence, specifically when tainted by synthetic video generation.

3.2 Applying the tools to the problem at hand

What found indicators of reliability can be applied to digital video evidence and in what way are these indicators disrupted by synthetic video generation?

Despite their mostly testimonial nature, the found indicators can also largely be applied to digital video evidence. As stated before, reliability of evidence is deemed to consist of its authenticity and its integrity. The authenticity of the evidence sees to whether or not the evidence actually connects to the evidence. For example, if a crime is committed on a public library pc using the suspect's credentials, but the suspect is not visible on the security footage, then is is likely that the authenticity of the evidence falters. The evidence does not connect to the suspect. The integrity sees to whether or not the evidence accurately displays the truth. Digital evidence is easily edited, and traces of editing can be covered up too. This leads to an ongoing cat-and-mouse game between those trying to falsify evidence, and those trying to detect falsified evidence. Technology however increasingly makes the detection of evidence of lesser integrity difficult.[123] When it comes to synthetic video generation, it is this integrity that is mainly of interest to us.

The primary indicators of accuracy, consistency and completeness are all still applicable, and even the main indicators of the integrity of the evidence when it comes to undetectable synthetic video evidence. The secondary indicators too can give the judge an inclination towards whether the evidence is of suitable integrity.

The accuracy is the main disturbed area. Synthetically generated video evidence is inherently inaccurate, as it does not show the truth as is. Synthetic video evidence however can be presented as real. It is here that the consistency and completeness form the primary indicators to the court for assessing the integrity of the evidence. The completeness and the consistency of the evidence will form the court's major tools to detect and assess the integrity of synthetically generated video evidence. As it is fabricated, it is bound to be inconsistent with reality at some points.

The secondary indicators too can give the court an inclination towards the integrity, and as such reliability of the evidence. The source of the evidence can play a large role, as it still requires some technical know-how to make use of the current software. The attitude, as well as motivation of the source can play a large role in this. If the source is known to be interested in synthetic video generation, and this source submits video evidence proving he was not guilty, then it is logical to question this evidence.

The presentation of the evidence however makes things more difficult due to its medium. As discussed before, video has a degree of inherent reliability. This inherent reliability is not yet weakened by technologies such as synthetic video generation, as these are only recently developing and have not yet reached mainstream knowledge. Whereas people are largely aware that photo manipulation is an everyday occurrence, synthetic video manipulation is still seen as largely isolated to Hollywood blockbusters and outside of the reach of the average person.

[123] Chesney (n60), p. 5.

The third category seeing to the creation of the evidence mostly doesn't apply to digital evidence, due to its testimonial nature. The handling of the evidence is of more importance, but fails to secure the integrity of the evidence when it comes to synthetic video evidence. It safeguards the integrity of the evidence within the process itself, but the evidence here is already of questionable integrity once it enters the system. The chains of custody and evidence prevent the other party from claiming that the evidence was changed during the investigation, but offer little else when it comes to the protection of the integrity of the evidence.

The court might very well still be capable of detecting synthetic video evidence, but this will be increasingly more difficult as the technology develops. The court is to be assumed trained and skilled in assessing the primary and secondary indicators, and these indicators still play a role within the detection of synthetic video evidence. A discrepancy within the consistency and the completeness of the evidence should arouse suspicion in the same manner that such a discrepancy would in testimonial evidence. The source of the evidence too can be an important indicator of a possibly questionable integrity of the evidence. Many of the indicators are however not as suitable to assessing digital evidence as they are to assessing testimonial evidence. This testimonial bias is a logical conclusion of years of assessing the reliability of mainly testimonial evidence, but now too highlights the lack of such a development of the assessment of digital evidence. As such, it is like using a fretsaw to try and cut down a tree. It might be capable and will eventually get the job done, but it is not quite as effective.

An additional problem exists that originates within the criminal procedural code as is. As discussed before, digital evidence does not fall within the acceptable evidence as laid out in art. 339 CPC. It is however an increasingly important type of evidence, and as such valuable to the court. The system as is however facilitates the increased difficulty in assessing the reliability caused by synthetic video evidence. As discussed before, digital evidence is submitted mainly through the existing acceptable categories of evidence. The main means of submission consist of the personal observation of the judge, and the expert witness' testimony or report.

When it comes to the personal observation of the judge, the problem lies within the detection of the synthetic video generation. Humans are inherently likely to deem video evidence as reliable.[124] Judges and juries also share this bias. Synthetic video generation results in realistic falsification, striving to be as indistinguishable as possible from real footage. The detection in first instance as such is a problem that should not be overlooked. Whether the judge is capable of this is questionable. The judge is first and foremost a legal expert, and to expect every judge to be an expert in every field he judges on is at best unreasonable. The court as such might not have the knowledge required to detect synthetic video generation in the first place.

This problem is strengthened by the knowledge paradox.[125] Experts might be able to explain the notions of the integrity of the evidence, but as long as the court is not aware of the possibility of synthetic video generation and its possible misuse within the court, the chances are low that an expert is called upon to do so in the first place. To determine that an expert's knowledge is necessary, a certain amount of knowledge must be held by the court in the first place. In the current stage and spread of the technology, it is questionable whether this knowledge currently exists within the court.

[124] Asser (n73), p. 2017/249.

[125] Evenblij (n9), p. 62-69.

Finally, the current approach circumvents possible safeguards when it comes to digital video evidence. Each acceptable category of evidence has some safeguard build into them, to validate that the decision-making process is based upon a sound foundation. A clear example of such a safeguard is the 'unus testis, nullus testis'-approach as seen in testimonial evidence, requiring a conviction to be based on more than one testimonial source. These safeguards however date to the inception of the procedural criminal code, and as such were not created with the intention of them being used with digital evidence. Meanwhile, digital evidence is prone to falsification, and falsification is increasingly difficult to detect. In a system as heavily procedural as the criminal one, one would think that digital evidence would not be as loosely defined and handled as it is.

3.3 Conclusion

The court has several tools at hand to assess the reliability of the evidence. This reliability consists of the authenticity and the integrity. These tools consist of three primary categories, as well as four secondary categories. This distinction is mine, as I have found the three primary indicators of accuracy, completeness and consistency to play a role in nearly all assessments of reliability. The secondary categories of the person of the witness, the presentation of the evidence, and the creation of the evidence, are more situational, but can still give the court an indication towards the reliability of the evidence.

These tools are however largely based on testimonial evidence and do not see as much to digital evidence. Whereas the court is likely still capable to assess the reliability of synthetic video evidence using these existing indicators, it is most likely not quite as effective as when used to assess testimonial evidence.

Additionally, the law as is strengthens some problems encountered. For the court to assess the reliability of the evidence, it must first be aware that there is a possible unreliability. With synthetic video generation the court might however have no inclination that it is looking at a falsification. A knowledge paradox occurs: to further examine a problem in evidence it is first be necessary to identify the problem. Finally, the system as is was not made to handle digital evidence, and as such digital evidence is merely subject to the safeguards dictated by whatever means it is submitted, not by safeguards that would be more fitting towards the nature of digital evidence.

Chapter 4 The witness as example

How can synthetic video evidence be compared to eyewitness evidence, and are there lessons to be learned from how the treatment of eyewitness evidence has developed over the years that are applicable to synthetic video evidence?

Where there are humans, there is some form of law to guide their co-existence. There are thousands of years of legal development that cumulate into the system that we have today. The technological progress and accompanying digitalization of the society cause new disruptions that the law must face. With the extensive history of the law, the disruption observed might however in essence be much alike to one observed before. Whereas these disruptions might cause traditional systems to no longer be as suitable as they once were, the same traditional systems might give us a view of a possible solution in the way a previous disruption was dealt with.

For synthetic video generation, this solution might lie within the treatment of eyewitness' evidence. The eyewitness has been a central and important, yet also inherently unreliable type of evidence that has been discussed for thousands of years. The eyewitness is an inherently flawed form of evidence, yet ever still an important category. The eyewitness is a natural form of evidence.[126] When a crime is observed by a witness, it is only logical that this witness gets to tell what they saw. The witness' statement can fill in the voids left by the other evidence and shed light on things that other evidence cannot as easily observe. The witness however is also prone to outside influence, is inherently coloured in their observation, and is limited in their memory as well as in their ability to express those memories in the right words.[127]

The problems faced by the judge with the witness is in many ways alike to the disruptions caused by synthetic video evidence. There is evidence. This evidence might be unreliable, but whether it is or not is difficult to detect. The evidence, when reliable, might however be of importance to the case, and dismissing it completely could have unwanted consequences for both suspect, victim, and prosecutor. To perhaps find a possible answer in our own past, I will explore the relation between eyewitness evidence and whether it contains a possible manner to handle synthetic video evidence.

4.1 *<u>The silent witness in comparison</u>*

How can eyewitness evidence and synthetic video evidence be compared in regards to their respective reliability?

When it comes to the reliability of the evidence, the same problems can be observed in both the eyewitness evidence and the synthetic video evidence. This is especially the case when it comes to the integrity of the evidence. To demonstrate this, the indicators of reliability as discussed in the previous chapter will be revisited and applied to the eyewitness' testimony, as well as compared to the problems arising with synthetic video evidence. Eyewitness evidence interferes with several of the found elements of reliability in the same way that synthetic video evidence does. Most notably, the core problem of the inaccuracy of the evidence is similar in nature. When it comes to eyewitness evidence however, the law and

[126] R.H. de Bock, *Tussen Waarheid en Onzekerheid* (Dissertation, 2011), p. 237-240.

[127] Asser (n73), p. 253.

legal practice are more capable of detecting such an inaccuracy. This is most apparent within the secondary indicators found.

As discussed in chapter 3, the indicators of reliability can give the court an indication towards the reliability, consisting of the integrity and authenticity, of the evidence. I distinguish three primary indicators of reliability, namely the accuracy of the evidence, the consistency of the evidence, and the completeness of the evidence. Apart from these generally present indicators, four other more optional secondary types of indicator were identified. These are the source of the evidence, the presentation of the evidence, the creation of the evidence, and the handling of the evidence. With the aid of these indicators, the court is deemed capable of determining whether the evidence is reliable.

Similar to synthetic video evidence, the problem of the accuracy of eyewitness evidence is twofold. The first problem is the inaccuracy itself, which is unwanted in evidence. The second problem is that the inaccuracy is also difficult to detect. Research has shown that there is no specific tell that people are lying, nor that humans in general are capable of detecting lies.[128] This difficulty of detection will be further discussed first, before returning to the primary and secondary indicators of reliability. For the time being, we can conclude that the accuracy of both the witness statement and synthetic video evidence is generally at a certain risk, and that this potential inaccuracy is difficult to detect in both.

Testimonial evidence is prone to both intentional and unintentional errors. The judge is deemed capable of discerning whether the testimonial evidence is reliable or not.[129] The question however remains whether that is truly the case. Literature has shown that judges differ in their assessment of the reliability of evidence; an unwanted effect at best due to its effect on the legal certainty and inequality this can create.[130] Despite these problems, the judge might however be more capable of judging humans than judging data.

First, there is an inherent bias towards the detection of unreliability within human evidence. Judges are human. One of the things humans are inherently skilled at is judging other humans.[131] Human behaviour as such still plays a role within the process. Whereas is has been shown that there are no telltale signs of lying, as well that humans are not very good at detecting whether or not people lie, this does not keep us from trying.[132] This can be seen in the secondary indicators of reliability. These secondary indicators often see to more humane elements that are inherent to for example the assessment of a testimony. The secondary indicator of the presentation of the evidence sees to this especially, as both the emotion and the assuredness of the presentation are deemed as indicators of the reliability of the evidence. These are however social values that say little of the actual reliability of the evidence, yet seemingly do play an important part in the process of assessing the reliability of evidence.

These social-oriented indicators are however much less applicable to synthetic video evidence, leaving the judge in total with less indicators on whether the evidence should be deemed reliable. Synthetic video evidence itself can display emotion if so desired by the creator, but this emotion is hardly a true expression as it would be during a testimony. It's predetermined and calculated into the creation of the evidence, losing much of its indicatory power. These social-oriented indicators fall away, leaving less indicators for the judge to base his assessment on.

[128] Dubelaar (n25).

[129] Sauerland (n18).

[130] Sauerland (n18).

[131] H.H. Kelley, 'Attribution Theory in Social Psychology' (1967) *Nebraska Symposium on Motivation* 15, p. 192-238.

[132] Dubelaar (n25), p. 186.

Second, there is a historical development when it comes to the protection of the reliability of unreliable evidence. This development has largely not yet occured when it comes to unreliable digital evidence. The assessment of testimonial evidence is an age-old practice. Records dating back to the Roman times show disputes surrounding testimonial evidence, discussing the questionable value of having only a single witness.[133] This 'unus testis, nullus testis'-approach too can be found in both the Old and the New Testament.[134] The idea that a single witness might not tell the whole truth is an old one. Through the use of such ideas as the unus testis-approach these risks of reliability have been reduced as early as two thousand years ago. Additionally, over the years a shift has occurred within the perceived reliability of testimonial evidence.

Whereas once someone's honor and word were a highly valuable good, nowadays these carry with them less power. This is combined with a shift in understanding that a reputable person does not necessarily mean that his testimony too is reputable. Due to its historically important role within the court and its natural status as evidence, additional safeguards have been developed within the law. Within Dutch law, this can be seen in the different requirements and limits the law states on for example the witness' testimony. The witness is limited within his testimony to his own observations and experiences, and the verdict cannot be based upon a single testimonial source but must be supported by additional evidence.[135] The requirements of these supporting evidence then have seen years of jurisprudence to determine what exactly the requirements are and what it is that has to be supported.[136]

Digital evidence as a whole, and especially synthetic video evidence, are however a rather new phenomenon. Digital evidence as a whole has only come largely into play in the last few decades, only truly surging in use the past decade.[137] Technology continually develops, and society adapts new technologies quickly. The court and the law at times have trouble keeping up with these developments. An example of this is the use of mobile phones as evidence.[138] Whereas these not fifteen years ago were mostly dumb devices capable of calling, texting, and maybe run tetris, nowadays they are essentially small carry-on computers that contain a lot of details on your private life.[139] These personal and intimate zones are historically well protected within criminal law and its possible infringements. However, this protection does not seem apparent when looking at mobile phones.[140] Social norms, and in turn the law, take time to adapt to these technological advancements. When it comes to digital evidence as a whole this adaptation has only recently begun.

Finally, judges are aware that testimonial evidence is inherently prone to errors, be they intentional or unintentional. Testimonial evidence is treated with due suspicion, and rightly so. As discussed before, this might not be the case with digital video evidence.

The accuracy of the evidence is the most important indicator, as it sees to how much the statement relates to the truth. All evidence has the potential to give a false indication of what occured, but this potential is larger when it comes to eyewitness evidence. The witness might have his own reasons to tell his story with a certain twist. The witness is not

[133] Seneca the Elder, *Controversiae*, 7.5.4.

[134] Deuteronomium, 19:15; and 2 Korinthe, 13:1.

[135] Wetboek van Strafvordering 1921, art. 339(1)(3) jo 342.

[136] HR 30 juni 2009, ECLI:NL:HR:2009:BH3704, NJ 2009/495.

[137] Stekelenburg (n7), p. 241.

[138] Koops (n49).

[139] K. Shilton, 'Four Billion Little Brothers? Privacy, mobile phones, and ubiquitous data collection' (2009) *Communications of the ACM* 52/11, p. 48-53.

[140] Koops (n50), p. 13-31.

a perfect observer, and as such will selectively see and remember things based on what he deems important.[141] The witness' memories might be influenced by outside parties, such as with leading questions in a police interrogation. These are only a few examples of reasons, both intentional and unintentional, that can lead to an inadequate accuracy of the witness' statement. As discussed in chapter 3, synthetic video generation is inherently inaccurate. It is a realistic falsification, and as such is always the product of the source that created it and selected the input. As such, both video evidence and testimonial evidence have a heightened chance to be inaccurate, but this inaccuracy is more difficult to detect in synthetic video evidence.

The consistency and the completeness of the evidence often play a large role in the witness statement, and both form a major indicator of the reliability. Internal and external consistency play an important role in this. The internal consistency sees to how often the statement contradicts itself. If a statement includes two different date for the same occurrence, it is logical that at least one of these must be wrong. The external consistency sees to how often the testimonial evidence and other evidence contradict each other.

The consistency and completeness of the evidence play a similar role when it comes to synthetic video evidence. The internal and external consistency of the evidence play an important role in both, as inconsistencies are often the most apparent indicator that something is amiss. A judge cannot magically sense whether someone is lying, but must come to this conclusion based on the facts and circumstances presented to him. This goes in a similar manner for the synthetic video evidence. If it is presented as legit, it is up to the court to determine why it is not acceptable.

The secondary indicators of reliability are of more interest when it comes to the relation between synthetic video evidence and testimonial evidence. It is here that we can see a bias towards testimonial evidence. This is logical, as it is mostly testimonial evidence that requires a motivation on its reliability by the court. As such, tools to assess the reliability of mostly testimonial nature are formed over time. These indicators are also easier to identify, as they are used more often.

The source of the evidence traditionally sees to the person making the witness' statement. This is apparent within its subcategories. The sincerity of the source of the evidence sees to in what manner the source itself believes that their evidence is accurate. This is one of the subcategories that is still applicable to synthetic video evidence, as the motivation of the source can grant insight into the accuracy of the evidence. If the husband of a suspect suddenly delivers synthetic video evidence that could fully prove that suspect had nothing to do with it, then this personal relationship can be seen as a reason to doubt the reliability of the evidence in the same manner it would were the husband to give a witness' statement.

The competence of the source sees to the ability of the source to have made the observance displayed within the evidence, understand this observance, as well as their capability to relay this observation. Just the wording makes it apparent that this is largely aimed at testimonial evidence. The witness ought to be capable of making an adequate observation, understand this observation, as well as relay this observation. If the source is not deemed competent, then this is an indicator that the reliability of the evidence might not be adequate.

Whereas it cannot be directly applicable as is to synthetic video evidence, the inverse could be applied. Synthetic video generation is still a rather technical process that not everyone would be capable of. As such, the known capability of being able to synthetically generate evidence might be reason in itself to question the reliability of submitted evidence. It must

[141] Asser (n73), p. 2017/253.

however be pointed out that as time passes and the limitations as explored in chapter 2 decrease, this inverse application of the competence of the source also decreases in potential as more people will gain access to the technology and the knowledge.

The secondary indicator of the presentation of the evidence too has a bias towards detecting unreliability within testimonial evidence, once again apparent within its subcategories. The emotion and the assuredness of the presentation see mostly to testimonial evidence, as they assess the means of delivery of the testimony and the effects of it upon the court and the perceived reliability. The medium however does make a difference between the two. As the name suggests, synthetic video evidence is rather bound to one medium. This medium however has a certain degree of inherent reliability to it.[142] The indicators of the emotion and the assuredness of the presentation are as such not as applicable to synthetic video evidence, but the choice of medium of the presentation can result in synthetic video evidence being deemed more reliable.

Finally, the personally identified secondary indicator of the handling of the evidence sees to the manner in which the evidence has been dealt with after becoming part of the investigation. It sees much more to material evidence than the other secondary indicators, and is mainly of importance when it comes to the integrity of the evidence. It does not apply to eyewitness evidence, but also fails to prevent the disruption as caused by synthetic video evidence. The chain of custody and evidence see to it that the evidence is not tampered with during the investigation and trial, but do not and cannot guarantee the integrity of the evidence before the investigation and trial. This secondary indicator does nothing to assess whether the evidence is unreliable outside of the handling of the investigating parties.

As such, the essence of the problem of both eyewitness evidence and synthetic video evidence is largely the same. Both have an inherent inaccuracy to them. This inaccuracy is difficult to detect, but for different reasons. This will be discussed further below. There is a logical bias found within the secondary indicators of reliability to assess the reliability of testimonial evidence. These secondary indicators are somewhat translatable to synthetic video evidence, but when it comes to criminal law it begets the question whether this is to a satisfactory degree.

4.2 Unus Video, Nullus Video

Are there lessons to be learned from how eyewitness evidence has been developed over the years that are applicable to synthetic video evidence?

If the problem in essence is the same, then perhaps something could be learned of the years of development to secure the reliability of eyewitness evidence. Whereas the assessment of synthetic video evidence itself will have to develop over time based on the technology and its effects, the safeguards implemented by the law are something that can be looked at now.

The witness statement falls under art. 339(1)(3) jo. 342 CPC, where it is described as a statement made during the trial, about the facts and circumstances experienced by the witness. The statement must see to the witness' own experience or observation, and cannot include guesswork or conclusions. The requirement of the statement being made in the courtroom is in practice of much lesser effect, as the court allows hearsay statements. An important element of the judgement of the witness statement lies within the Unus Testis

[142] Asser (n73).

approach.[143] The evidence that the suspect has committed the alleged crime, cannot be based solely on the testimony of one witness. The statement must at least be supported by one additional piece of evidence. The high court has elaborated on this in multiple cases.

In Unus Testis I[144], the High Court dealt with a case in which the victim was allegedly raped by her at the time husband. There had been eleven years between the alleged act and the police report. The only available evidence was the statement of the victim, and a de-auditu statement. A statement from a witness however cannot be used as sole basis for a conviction. As such, the High Court stated that a conviction cannot be made on one witness statement alone, but that this statement must be supported by other evidence to guarantee the soundness of the judgement.

In Unus Testis II[145], about an intimidation in the city of Gouda, the High Court found that merely the statement of the victim in combination with the suspect carrying a knife was too meagre evidence to come to a guilty conviction. The mere carrying of a knife by the suspect did not support the statement of the victim enough that he was threatened by the suspect. Once again, the High Court reiterates that it cannot give a general guideline as to when supporting evidence supports the statement enough, as this greatly depends on the facts and circumstances. It however refers to other cases in which it has judged on such evidence.

The requirements for what was acceptable have been developed in mostly case law. In subsequent cases[146], the High Court expanded upon what counted as evidence being supporting enough to carry the testimony. Such supporting evidence does not have to confirm everything in the statement, but has to touch upon it in specific areas. If the witness states that during an abduction she saw glimpses of industrial chimney, and other evidence shows that the route the abductors took passed an industrial zone, then this can serve as supporting evidence to the statement. Without this supporting evidence, however, the statement would not be open to the judge to use as basis for a conviction. The high court however also stated that it could not give general rules, as it must be judged on a case-by-case basis whether the supporting evidence is strong enough.

Whereas the silent and the non-silent witness do differ from each other, this is not to say that nothing can be learned from the approach taken with the witness statement. Especially the unus-testis-approach for witness testimonies might be an elegant inspiration to the problem introduced by synthetic video generation. It recognizes that the witness is a flawed source, but still values its important role in the legal system as natural evidence. The added safeguard serve as an additional line of defense to guarantee the soundness of the judgement.

An unus-testis-approach for digital video evidence could in theory be easily translated and function in essence in much the same way as it does with the witness statement. A single instance of digital video evidence on its own would no longer be enough to come to the conviction of the alleged crime, as is possible in the current system. Instead, it would need to be supported by additional evidence. This evidence does not need to confirm the entirety of the digital video evidence, but only touch upon it in certain areas, verifying the contents of the video evidence.

This approach would not eliminate the problem with altered digital video evidence at its core, but it would however possibly decrease the effect such evidence may have in the

[143] Wetboek van Strafvordering 1921, art. 342(2).

[144] HR 30 juni 2009, ECLI:NL:HR:2009:BG7746, NJ 2009/495.

[145] HR 30 juni 2009, ECLI:NL:HR:2009:BH3704, NJ 2009/495.

[146] HR 13 juli 2010, ECLI:NL:HR:2010:BM2452, NJ 2010/515; HR 20 december 2016, ECLI:NL:HR:2016:2911, NJ 2017/91.

courtroom in the same manner the unus-testis approach has on testimonial evidence. In a similar manner as the unus-testis approach, it would not be part of the reliability check of the evidence done by the judge, but take place afterwards.[147] The judge is still expected to determine whether the evidence is reliable with their own toolset and judgement, and the reliability check remains an important part of the judgement as a whole.

Added to this is however the additional safeguard. Even in the hopefully exceptional circumstances in which inaccurate evidence is faultily deemed reliable, the additional requirement of supporting evidence would put up a significant threshold to the successful use of altered digital video evidence in the courtroom. Such evidence would need to be supported by additional evidence, confirming at least partially the shown imagery.

4.3 Conclusion

Synthetic video evidence disturbs the use of video evidence within the court greatly with its difficult to detect inherent inaccuracy. Any piece of video evidence could be synthetic video evidence, and it is up to the court to determine the reliability of such evidence. There is however another type of evidence that is inherently prone to falsehoods, be they intentional or unintentional. The witness forms an important and even natural type of evidence that still sees great use in today's court.

The problem in essence between the use of eyewitness evidence and synthetic video evidence is the same. The evidence has a higher chance of being inaccurate, but this inaccuracy is difficult to detect. Synthetic video evidence is however more difficult to detect than false eyewitness evidence. Humans are inherently skilled at judging other humans, including their testimonies. Eyewitness evidence has existed for a long time, and has been regarded with due knowledge of possible unreliability for a time. To counter this unreliability, additional safeguards have been developed over time in law and jurisprudence. These safeguards do not yet exist for synthetic video evidence. Finally, the court might simply not be aware of the possibility of synthetic video evidence.

This testimonial bias is visible within the other indicators of reliability as well. Whereas still applicable, these are likely not nearly as effective as they are when assessing testimonial evidence. This is not to say that a possible answer to the problem can not be found within the handling of testimonial evidence. One such safeguard is the Unus Testis-approach, requiring testimonial evidence to be supported by additional evidence before it can be accepted. This approach could also be applied to a create an Unus Video-approach. Whereas this would not eliminate the core problem of synthetic video evidence, it would however greatly increase the difficulty of using synthetic video evidence.

[147] P.J. van Koppen, *Overtuigend Bewijs:Indammen van Rechterlijke Dwalingen* (Nieuw Amsterdam 2011), p. 248-249.

Conclusion

Synthetic video generation has the potential of disrupting the Dutch criminal system of evidence. The Netherlands uses a negative legal system, limiting the acceptable types of evidence. Digital evidence does currently not fall within these acceptable types as is, but can be presented in several ways. Synthetic video generation sees to the artificial realistic falsification or alteration of digital imagery. These are inherently inaccurate, yet this inaccuracy can be difficult to detect. Through these different forms of submission, this inaccurate evidence has an increased chance of being accepted.

Several indicators of reliability can be found in the literature. I have made a distinction between primary and secondary indicators. The primary indicators were generally taken into consideration, whereas the secondary indicators are largely dependent on the type of evidence. The primary indicators of reliability consist of the accuracy, the consistency and the completeness of the evidence. The secondary indicators consist of the source of the evidence, the presentation of the evidence, and the creation of the evidence. The current indicators of reliability are developed to handle testimonial unreliability. As such, whereas the court might still be capable of assessing the reliability of digital video evidence, the effectiveness might be lessened greatly.

The mere heightened chance of evidence being inaccurate in combination with the difficulty of detection however should not fully exclude digital video evidence. There are parallels between digital video evidence and eyewitness evidence. Both share the same base problem of a heightened and difficult to detect potential inaccuracy. The secondary indicators of reliability strengthen the assessment of the reliability of eyewitness evidence. This does not occur in the same manner for digital evidence. The approach taken when it comes to the treatment of eyewitness evidence can form a basis on how to treat synthetic video evidence. As an example, the implementation of an Unus Testis-approach for digital video evidence would make the acceptation of synthetic video evidence much more difficult.

I am of the opinion that the system as is fails to properly safeguard the stakes at play of the participants of the criminal system when it comes to synthetic video evidence. Evidence forms a core part of the system, and it simply requires the evidence to be reliable. The current system in place sufficiently warrants the assessment of the reliability of eyewitness evidence, but this does not translate completely to the assessment of the reliability of synthetic video evidence. The primary indicators can still function in the court's assessment, but the secondary indicators largely see to evidence of a testimonial nature. Whereas this problem will likely solve itself over time due to the creation of new secondary indicators, there is still a period of time in which the assessment of synthetic video evidence is weakened. The current situation regarding the types of evidence, specifically with digital evidence not falling within a specific type of evidence, work to weaken the assessment further. A separate type of acceptable evidence for digital evidence could strengthen the integrity, authenticity, and reliability of digital evidence as a whole.

As such, synthetic video generation has the potential to disrupt the use of digital video evidence within the Dutch criminal system by undermining the reliability of the evidence. The stakes at play for those participating in the criminal system are generally high, and an improper assessment of the reliability of evidence could lead to unacceptable results.

This conclusion would be insincere without stating this thesis' limitations, as it are these limitations that greatly affect the research. First, it observes a newly emerging technology that is largely unexplored, as is also evident in available research. Digital evidence as a whole is only in the past decades entering the courtroom en masse, and synthetic video generation might not play a role for years to come.

Both digital evidence, as well as the assessment of the reliability of evidence are surprisingly understudied fields within the Netherlands. There are calls for more research into these fields, but these seem to have largely gone unanswered.[148] The available sources have a largely testimonial basis. Nonetheless, I do believe that these sources are authoritative, and can be sufficiently applied to the question at hand. As most logically see to testimonial evidence primarily, this has to be done carefully when applying it to the assessment of digital evidence. Additional research into the actual practical assessment of the reliability of digital evidence in general is however long overdue, and it could be a great benefit to shed additional light on this mostly still obscure topic.

The developed framework of the indicators of reliability could form a useful tool for future research. Whereas originally mostly grounded in research regarding testimonial evidence, I am of the conviction that it can be applied to other types of evidence when taking into account the split between primary and secondary indicators. The primary indicators form the core of the assessment, whereas the secondary indicators can give additional information depending on the type of evidence and the facts and circumstances of the case. This allows for a review of the assessment of the reliability of evidence other than testimonial evidence, and as such opens up possibilities in fields that have so far seen little research. Specifically the assessment of digital evidence falling outside of the scope of the original accepted types of evidence, such as digital photographs or videos could be examined through the use of this framework. Additionally, the framework could be expanded with further primary or secondary indicators through additional research within case law.

[148] Stevens (n89), p. 2842.

Literature

Law

Wetboek van Strafvordering 1921

Memorie van Toelichting, Ontwerp tot vaststelling van een wetboek van strafvordering, der Koningin aangeboden door de Staatscommissie voor de herziening van het Wetboek van Strafvordering, ingesteld bij Koninklijk Besluit van 8 April 1910, no. 17, deel 1 Ontworpen Wetboek (Landsdrukkerij 1913)

Computercriminaliteit I 1993

Computercriminaliteit II 2006

Cases

HR 20 december 1926, ECLI:NL:HR:1926:BG9435, LJN BG9435.

HR 14 september 1992, ECLI:NL:PHR:1992:AC3716, NJ 1993/54.

HR 30 maart 2004, ECLI:NL:PHR:2004:AM2533, NJ 2004/376.

HR 29 augustus 2006, ECLI:NL:HR:2006:AX6414, NJ 2007/134.

HR 3 juli 2007, ECLI:NL:HR:2007:BA4994, NJ 2007/412.

HR 23 september 2008, ECLI:NL:HR:2008:BD3902, NJ 2008/525.

HR 30 juni 2009, ECLI:NL:HR:2009:BH3704, NJ 2009/495.

HR 15 december 2009, ECLI:NL:HR:2009:BJ2831, NJ 2011/78.

HR 13 juli 2010, ECLI:NL:HR:2010:BM2452, NJ 2010/515.

HR 11 januari 2011, ECLI:NL:HR:2011:BP0291, NJ 2011/116.

HR 29 januari 2013, ECLI:NL:HR:2013:BY0816, NJ 2013/414.

HR 20 december 2016, ECLI:NL:HR:2016:2911, NJ 2017/91.

Books

Ahanorian, A.A., and Bornstein, B.H., ‘Stress and Eyewitness Memory’ in B.L. Cutler (ed), *Encyclopedia of Psychology and Law* (SAGE Publications, Newbury Park 2008)

Asser, W.D.H., *Asser Procesrecht 3 Bewijs* (2nd revision, Kluwer, Deventer 2017)

Besier, L.C., and Blok, A.J., Het Nederlandsche Strafproces Deel II, (Tjeenk Willink, Deventer 1925)

Corstens, G.J.M. and Borgers, M.J., *Het Nederlands Strafprocesrecht* (8th edition, Kluwer, Deventer 2014).

Dubelaar, M.J., *Betrouwbaar getuigenbewijs* (Kluwer, Deventer 2014)

Stekelenburg, M., *De Betere Byte in de Strijd Om Het Gelijk* (Vrije Universiteit Amsterdam, Amsterdam 2009)

Van Koppen, P.J., *Overtuigend Bewijs:Indammen van Rechterlijke Dwalingen* (Nieuw Amsterdam, Amsterdam 2011)

Van Woensel, A.M., and Van Laanen, F., ‘Commentaar op Artikel 340’ in M.S. Groenhuijsen and others (eds.), *Losbladige commentaar op het Wetboek van Strafvordering / Melai* (Kluwer, Deventer 2007)

Wolters, G., Odinot, G., 'Zijn zekere getuigen betrouwbare getuigen?' in P.J. van Koppen and others (eds.) *Reizen met Mijn Rechter: Psychologie van het Recht* (Kluwer, Deventer 2010).

Papers

De Bock, R. H., *Tussen Waarheid en Onzekerheid* (Dissertation Tilburg University, 2011)

Li, Y., and Lyu, C., and Lyu, S., 'In Ictu Oculi: Exposing AI Generated Fake Face Videos by Detecting Eye Blinking' (2018) University of Albany <https://arxiv.org/abs/1806.02877> Accessed on 31 October 2018.

Feigenson, N., Spiesel, C. and Sherwin, R.K., *Law in the Digital Age: How visual communication technologies are transforming the practice, theory and teaching of law* (NYLS Legal Studies Research Paper No. 05/06-6, Barbados Group Working Paper no. 05-06, 2005)

Karras, T., Laine, S., and Aila, T., 'A Style-Based Generator Architecture for Generative Adversarial Networks' (12 December 2018) Cornell University <https://arxiv.org/abs/1812.04948> Accessed on 28 December 2018.

Articles

Buruma, Y., 'Betrouwbaar Bewijs' (2009) *Delikt en Delinkwent* 23.

Chesney, R. and Citron, D.K., 'Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security' (2019 forthcoming) *California Law Review* 107

Douglas, K.S., and Lyon, D.R., and Ogloff, J.R.P., 'The impact of graphic photographic evidence on mock jurors' decisions in a murder trial: Probative or prejudicial?' (1997) *Law and Human Behavior* 21(5).

Dubelaar, M., and Vanderveen, G., 'Beeld en geluid in het strafproces: Implicaties van de opkomst van (audio)visuele technieken en materialen voor communicatie en besluitvorming in de strafrechtspraktijk' (2009) *Nederlands Juristenblad* 1530.

Evenblij, M., Gerechtelijk Deskundigen: vechten tegen de kennisparadox, *Mr. Magazine* (6/7) 2008.

Feigenson, N., 'Visual Evidence' (2010) *Psychonomic Bulletin & Review* 02.

Goodfellow, I. and others, 'Generative Adversarial Nets" (2014) *Advances in Neural Information Processing Systems* 27 (NIPS 2014).

Kassin, S.M., and Garfield, D.A., 'Blood and guts - General and trial-specific effects of videotaped crime scenes on mock jurors'(1991) *Journal of Applied Social Psychology,* 21(18).

Kelley, K.K., 'Attribution Theory in Social Psychology' (1967) *Nebraska Symposium on Motivation* 15

Koops, B.J., 'Privacy Spaces' (2018) *West Virginia Law Review* 121

Loftus, E.F., 'Planting misinformation in the human mind: A 30-year investigation of the malleability of memory' (2005) *Learning & Memory* 12.

Merckelbach, H.L.G.J., Candel, I.E.L., Crombag, H.F.M., 'De Goede Getuige' (2003) *Trema* nr. 6.

Mori, M., 'Uncanny Valley' (1970) *Energy* 7(4)

Odinot, G., Wolters, G., Van Giezen, A., 'Accuracy, confidence and consistency in repeated events' (2013) *Psychology, Crime & Law* 19/7

Oerlemans, J.J., 'Veroordeling Voor "Uitreizen" Naar Syrië En de Rol Digitaal Bewijs' (2017) *Computerrecht* 242-243.

Samuel, A.L., 'Some Studies in Machine Learning Using the Game of Checkers' (1959) *IBM Journal of Research and Development* Vol. 3, Issue 3.Sauerland, M., Krix, A.C. and Merckelbach, M., 'Identificaties Door Ooggetuigen: Waarom Een Rechtspsycholoog Handig Is' (2016) *Nederlands Juristenblad* 2016/1562

Shilton, K., ‘Four Billion Little Brothers? Privacy, mobile phones, and ubiquitous data collection’ (2009) *Communications of the ACM* 52/11

Stevens, L, ‘Bewijs Waarderen’ (2014) *Nederlands Juristenblad* 40.

Thies, J., and others, ‘Face2face: Real-Time Face Capture And Reenactment Of RGB Videos’ (2016) *2016 IEEE Conference on Computer Vision and Pattern Recognition*

Wells, G.L., Memon A. and Penrod S.D., ‘Eyewitness Evidence, Improving its Probative Value’ (2006) 7(2) *Psychological Science in the Public Interest*

Wells, G.L. and Bradfield, A.L, ‘“Good, you identified the suspect:” Feedback to eyewitnesses distorts their reports of the witnessing experience’ (1998) *Journal of Applied Psychology* 83(3)

Wells, G.L., and Olson, E.A., ‘Eyewitness Testimony’ (2003) *Annual Review of Psychology* 45:279-95

Reports

Commissie modernisering opsporingsonderzoek in het digitale tijdperk, *Regulering Van Opsporingsbevoegdheden In Een Digitale Omgeving* (report, 2018)

GSMA, Measuring the Future of Mobile (*GSMA Intelligence data December 2017*, 2017) <www.gsma.com/mobileeconomy > Accessed on April 22, 2018.

Websites and Blogs

Christian, J., ‘Experts Fear Face Swapping Tech Could Start an International Showdown’ (*The Outline*, February 1, 2018) <https://theoutline.com/post/3179/deepfake-videos-are-freaking-experts-out?zd=2&zi=4n4jjajh> Accessed on April 22, 2018.

Cole, S., ‘AI-Generated Fake Porn Makes Have Been Kicked Off Their Favorite Host: Reddit is still silent’ (motherboard, 31 January 2018), <https://motherboard.vice.com/en_us/article/vby5jx/deepfakes-ai-porn-removed-from-gfycat> Accessed on 17 June 2018.

Farokhmanesh, M., ‘Deepfakes are disappearing from parts of the web, but they’re not going away’ (The Verge, February 9, 2018) <www.theverge.com/2018/2/9/16986602/deepfakes-banned-reddit-ai-faceswap-porn> Accessed on April 22, 2018.

Hern A., ‘My May-Thatcher Deepfake Won’t Fool You But It’s Tech May Change The World’, (*The Guardian*, March 12, 2018) <www.theguardian.com/technology/2018/mar/12/may-thatcher-deepfake-face-swap-tech-change-world> Accessed on Aprill 22, 2018

Lee, D., ‘Deepfakes Porn Has Serious Consequences’ (*BBC*, February 3, 2018) <www.bbc.com/news/technology-42912529> Accessed on April 22, 2018

Oberoi, G., ‘Exploring Deepfakes’ (*Hackernoon*, March 5, 2018) <https://hackernoon.com/exploring-deepfakes-20c9947c22d9 > Accessed on April 22, 2018

Nederlands Forensisch Instituut, Factsheet (*NFI*, March 2007) <http://www.fomat.nl/nfi_fsnormen.pdf> Accessed on 31 October 2018.

Redactie, ‘Na fake nieuws ook fake porno? ‘Dit gaat echt heel naar worden.’’ (*NOS*, December 13, 2017) <https://nos.nl/op3/artikel/2207417-na-fake-nieuws-ook-fake-porno-dit-gaat-echt-heel-naar-worden.html> Accessed on May 5, 2018.

Reddit Originals, ‘How reddit works’ (30 July 2014), <https://redditblog.com/2014/07/30/how-reddit-works-2/> Accessed on 27 May 2018.