American University Law Review

Volume 69 | Issue 6 Article 5

2020

A Break From Reality: Modernizing Authentication Standards for Digital Video Evidence in the Era of Deepfakes

John P. LaMonaga
American University Washington College of Law

A Break From Reality: Modernizing Authentication Standards for Digital Video Evidence in the Era of Deepfakes

Abstract

The legal standard for authenticating photographic and video evidence in court has remained largely static throughout the evolution of media technology in the twentieth century. The advent of "deepfakes," or fake videos created using artificial intelligence programming, renders outdated many of the assumptions that the Federal Rules of Evidence are built upon.

Rule 901(b)(1) provides a means to authenticate evidence through the testimony of a "witness with knowledge." Courts commonly admit photographic and video evidence by using the "fair and accurate portrayal" standard to meet this Rule's intent. This standard sets an extremely low bar—the witness need only testify that the depiction is a fair and accurate portrayal of her knowledge of the scene. In many cases, proponents' ability to easily clear this hurdle does not raise concerns because courts rely on expert witnesses to root out fraudulent evidence; thus, although the fraudulent evidence might pass the fair and accurate portrayal standard, it would later be debunked in court.

The proliferation of deepfakes severely complicates the assumption that technological experts will be able to reliably determine real from fake. Although various organizations are actively devising means to detect deepfakes, the continued proliferation and sophistication of deepfakes will make debunking fake video more challenging than ever. Witnesses who attest to the fair and accurate portrayal standard will likely not be able to identify subtle but important alterations in deepfakes. As a result, fraudulent evidence, authenticated through the Rule 901(b)(1) standard, will increasingly enter courtrooms with a decreasing ability for witnesses and courts to identify fakes. Because the technology to detect deepfakes lags behind the creation methods, deepfakes present a critical threat to courtroom integrity under the current standard.

The rising probability that juries see fake videos warrants a higher burden on the proponent of video evidence. Requiring additional circumstantial evidence to corroborate video evidence is a small but crucial step that will mitigate, but not solve, the coming deepfakes crisis. Further engagement around this topic is necessary to address the deepfakes crisis before it creates irreparable harm.

A BREAK FROM REALITY: MODERNIZING AUTHENTICATION STANDARDS FOR DIGITAL VIDEO EVIDENCE IN THE ERA OF DEEPFAKES

JOHN P. LAMONACA[*]

The legal standard for authenticating photographic and video evidence in court has remained largely static throughout the evolution of media technology in the twentieth century. The advent of "deepfakes," or fake videos created using artificial intelligence programming, renders outdated many of the assumptions that the Federal Rules of Evidence are built upon.

Rule 901(b)(1) provides a means to authenticate evidence through the testimony of a "witness with knowledge." Courts commonly admit photographic and video evidence by using the "fair and accurate portrayal" standard to meet this Rule's intent. This standard sets an extremely low bar—the witness need only testify that the depiction is a fair and accurate portrayal of her knowledge of the scene. In many cases, proponents' ability to easily clear this hurdle does not raise concerns because courts rely on expert witnesses to root out fraudulent evidence; thus, although the fraudulent evidence might pass the fair and accurate portrayal standard, it would later be debunked in court.

The proliferation of deepfakes severely complicates the assumption that technological experts will be able to reliably determine real from fake. Although various organizations are actively devising means to detect deepfakes, the continued proliferation and sophistication of deepfakes will make debunking fake video more challenging than ever. Witnesses who attest to the fair and accurate portrayal standard will likely not be able to identify subtle but important

[*] Senior Staff Member, *American University Law Review*, Volume 70; J.D. Candidate, May 2021, *American University Washington College of Law*; B.A. History, 2009, *Princeton University*. I would like to thank the *Law Review* staff for their tireless assistance with this piece and my family of law school friends who have made this law school experience irreplaceable. Finally, I am forever grateful to my parents, who have never failed to lead the way.

alterations in deepfakes. As a result, fraudulent evidence, authenticated through the Rule 901(b)(1) standard, will increasingly enter courtrooms with a decreasing ability for witnesses and courts to identify fakes. Because the technology to detect deepfakes lags behind the creation methods, deepfakes present a critical threat to courtroom integrity under the current standard.

The rising probability that juries see fake videos warrants a higher burden on the proponent of video evidence. Requiring additional circumstantial evidence to corroborate video evidence is a small but crucial step that will mitigate, but not solve, the coming deepfakes crisis. Further engagement around this topic is necessary to address the deepfakes crisis before it creates irreparable harm.

TABLE OF CONTENTS

"*[R]eality is not external. Reality exists in the human mind, and nowhere else.*"
—George Orwell[1]

INTRODUCTION

Artificial intelligence and machine learning have enabled unprecedented leaps in mankind's capability to solve the most pressing issues of the twenty-first century.[2] Programmers and doctors have worked together to create artificially intelligent programs that synthesize data from millions of patients to diagnose illness with greater precision and speed than ever before.[3] Soon, self-driving cars will relieve humans of the deadliest

[1] GEORGE ORWELL, 1984 249 (New American Library ed. 1961) (1949).

[2] Machine learning is a subset of the broader application of artificial intelligence. While machine learning takes many forms, the "core notion is that the machine would be able to take data and learn . . . without human intervention." Vijay Singh, *What Is the Difference Between Machine Learning and Artificial Intelligence?*, DATA SCI. CENT. BLOG (Sept. 22, 2018, 9:00 PM), https://www.datasciencecentral.com/profiles/blogs/what-is-the-difference-between-machine-learning-and-artificial [https://perma.cc/XAG7-XBX2].

[3] Donna Marbury, *How Health Systems Are Using AI and Future Predictions*, MANAGED HEALTHCARE EXECUTIVE (Aug. 8, 2018), https://www.managedhealthcareexecutive.com/article/how-health-systems-are-using-ai-and-future-predictions [https://perma.cc/QJ6P-RC53]; *New AI Model Tries to Synthesize Patient Data like Doctors Do*, PAC. NORTHWEST NAT'L LABORATORY (Nov. 12, 2019), https://www.pnnl.gov/news-media/new-ai-model-tries-synthesize-patient-data-doctors-do [https://perma.cc/J7FW-PPX6]; *see* Emily Mullin, *FDA Approves AI-Powered Diagnostic that Doesn't Need a Doctor's Help*, MIT TECH. REV. (Apr. 11, 2018), https://www.technologyreview.com/f/610853/fda-approves-first-ai-powered-diagnostic-that-doesnt-

threat on our highways (ourselves).[4] However, notwithstanding the tremendous promise of improvement that artificial intelligence brings to our world, future generations may someday remember December 2017 as a seminal moment of the digital age that exposed the danger of advanced technological capabilities. As an internet technology website, Motherboard, first reported with great despair, in December 2017, a Reddit user with the online handle "deepfakes" created a series of videos utilizing new techniques that grafted the faces of several well-known actresses into pornographic videos.[5] Reddit, along with several pornographic websites, quickly featured explicit videos in which Daisy Ridley, Gal Gadot, and other actresses had never actually appeared.[6]

The level of sophistication of this technology was still blossoming; Motherboard reported that "[i]t's not going to fool anyone who looks closely. Sometimes the face doesn't track correctly and there's an uncanny valley effect at play, but at a glance it seems believable."[7] However, over the past several years, "deepfakes"—colloquially named after the otherwise unidentified Reddit user who circulated fake pornographic videos—have evolved from videos whose alterations are reasonably discernible by the naked eye to fakes that are challenging for both the human eye and machine detection software to distinguish from real videos.[8] This progression is predominantly due to the

need-a-doctors-help [https://perma.cc/J4XT-DWDP] (providing an example of diagnostic software that detects illness using patient data).

[4] Suhasini Gadam, *Artificial Intelligence and Autonomous Vehicles*, MEDIUM (Apr. 19, 2018), https://medium.com/datadriveninvestor/artificial-intelligence-and-autonomous-vehicles-ae877feb6cd2 [https://perma.cc/L4JN-X34J].

[5] Samantha Cole, *AI-Assisted Fake Porn Is Here and We're All Fucked*, VICE (Dec. 11, 2017, 2:18 PM), https://www.vice.com/en_us/article/gydydm/gal-gadot-fake-ai-porn [https://perma.cc/AUR4-W36D].

[6] Samantha Cole, *We Are Truly Fucked: Everyone Is Making AI-Generated Fake Porn Now*, VICE (Jan. 24, 2018, 1:13 PM), https://www.vice.com/en_us/article/bjye8a/reddit-fake-porn-app-daisy-ridley [https://perma.cc/P2D4-GASP].

[7] Cole, *supra* note 5. Japanese roboticist Masahiro Mori coined the concept "uncanny valley," used to describe a psychological phenomenon that occurs as a robot or android's visual resemblance to the human likeness improves; our subconscious enjoyment of the visual experience increases until the robot's likeness reaches a certain level of sophistication, at which point many feel "repulsive affects" that some describe as "creepy" or "eerie." Shensheng Wang, Scott O. Lilienfeld & Philippe Rochat, *The Uncanny Valley: Existence and Explanations*, 19 REV. GEN. PSYCHOL. 393, 393, 396 (2015).

[8] Editorial Board, *A Reason to Despair About the Digital Future: Deepfakes*, WASH. POST (Jan. 6, 2019, 7:10 PM), https://www.washingtonpost.com/opinions/a-reason-to-despair-about-the-digital-future-deepfakes/2019/01/06/7c5e82ea-0ed2-11e9-831f-

advancement of processes for creating deepfakes that use machine learning programs to continuously improve the fidelity of the videos and render increasingly lifelike representations.[9]

The coming proliferation of deepfakes has created no shortage of alarms in the legal, political, and social spheres, in which scholars predict countless challenges to organized society, ranging from celebrity harassment to political and governmental manipulation.[10] Some scholars have already rushed to address regulatory challenges that deepfakes pose and identify civil remedies for victims of deepfake videos.[11] For example, many state privacy torts do not account for artificial rather than actual depictions of the victim,[12] and First Amendment precedent is ill-equipped to deal with the expression of non-obscene but nonetheless manipulative fake videos.[13] However, despite some recognition that fake video is an imminent threat to courtroom integrity, lawmakers have done little to address the manner in which our evidentiary standards

3aa2c2be4cbd_story.html?utm_term=.f4f9e1e7b293 ("Deepfakes are also inherently hard to detect. The technology used to create them is trained in part with the same algorithms that distinguish fake content from real—so any strides in ferreting out false content will soon be weaponized to make that content more convincing.").

[9] *See infra* Part I.A.

[10] *See, e.g.*, Hallie Jackson, *Fake Obama Warning About 'Deep Fakes' Goes Viral*, MSNBC (Apr. 19, 2018), https://www.msnbc.com/hallie-jackson/watch/fake-obama-warning-about-deep-fakes-goes-viral-1214598723984 (highlighting director Jordan Peele's effort to educate the public about deepfakes by creating a realistic fake video of Barack Obama).

[11] *See, e.g.*, Elizabeth Caldera, Comment, *"Reject the Evidence of Your Eyes and Ears": Deepfakes and the Law of Virtual Replicants*, 50 SETON HALL L. REV. 177, 178 (2019) (arguing that the Federal Trade Commission is the best choice among administrative agencies to regulate deepfake technology); Douglas Harris, *Deepfakes: False Pornography Is Here and the Law Cannot Protect You*, 17 DUKE L. & TECH. REV. 99, 102–03 (2019) (arguing that a federal criminal law prohibiting fake pornographic videos is necessary to address deepfakes because state tort and non-consensual pornography laws are insufficient); Russell Spivak, *"Deepfakes": The Newest Way to Commit One of the Oldest Crimes*, 3 GEO. L. TECH. REV. 339, 340–41 (2019) (examining whether various state defamation or privacy tort causes of action are viable remedies or if they conflict with First Amendment protections).

[12] *See* Danielle Keats Citron, *Sexual Privacy*, 128 YALE L.J. 1870, 1921–24, 1939 (2019) (highlighting the disconnect between privacy torts and pornographic deepfake videos because the fake video represents no physical intrusion or truthful, private facts).

[13] *See* Spivak, *supra* note 11, at 358–64 (addressing the obscenity and child pornography exceptions to the First Amendment restraint on prohibiting a communication based on its content).

for authenticating photographic and video evidence must adapt to counter this threat.[14]

This Comment addresses the need for heightened evidentiary standards to counter the dangerous consequences of deepfakes, a need that is likely to become a central focus to our judicial process as prosecutors, plaintiffs, and defendants all turn to the courts to redress the threat and harms that deepfakes cause. Courts currently rely on an evidentiary standard that assumes authenticating witnesses have sufficient personal knowledge to attest to a photograph's or video's authenticity;[15] this standard is now inadequate to meet the intent of the Federal Rules of Evidence. Recent amendments to the Federal Rules of Evidence in 2017 aimed to address the growing influx of electronic media, such as social media posts or websites, into courtrooms.[16] However, the 2017 amendments did not replace or circumvent existing authentication requirements; instead, they allow the proponent of the evidence to offer authentication by certification rather than demanding witness testimony, which can be both costly and time-consuming.[17] Since the 2017 amendments, deepfakes have burst into the national consciousness, and their potentially devastating consequences demand further examination into the authentication standard for photographic and video evidence. Ultimately, current authentication standards for photographs and video fail to account for the inability of witnesses, even those present at the scene depicted, to determine reality from forgery.

Part I of this Comment explores deepfake video creation and the unique difficulty in authenticating or debunking them. The novel creation process that utilizes machine learning networks not only enables extraordinarily high-fidelity forgeries but also severely complicates detection capabilities. Part I also introduces the psychological effect

[14] *See, e.g.*, Jeff Ward, *10 Things Judges Should Know About AI*, 103 JUDICATURE 12, 17 (2019) (positing that the risk to "fundamental civic institutions and processes" may be undermined if the "current rules of evidence do not keep pace with these advances").

[15] *See* FED. R. EVID. 901(b)(1) (describing "[t]estimony that an item is what it is claimed to be" as sufficient to satisfy the authenticity requirement).

[16] *See* FED. R. EVID. 902(13)–(14) (declaring certified records and data as self-authenticating evidence, requiring "no extrinsic evidence of authenticity"). The amendments are "largely a reflection of the digital world in which we live." Ramona L. Lampley, *Something Old and Something New: Exploring the Recent Amendments to the Federal Rules of Evidence*, 57 WASHBURN L.J. 519, 519–20 (2018) (providing a practical explanation and analysis of the impact of the 2017 amendments on Rules 803 and 902).

[17] Lampley, *supra* note 16, at 525.

known as suggestibility, which makes deepfakes especially dangerous because of the human memory's susceptibility to recall events that never happened, compounding the deepfakes problem. Part II outlines the current legal standard that courts use to lay a foundation for the authenticity of video evidence to satisfy the requirement of Rule 901(a) of the Federal Rules of Evidence, primarily through Rule 901(b)(1) or Rule 901(b)(9).

Part III argues that, because of the high fidelity of deepfakes, witnesses no longer meet the recollection element of the personal knowledge standard established by Rule 602 to act as a witness with knowledge to testify that a video is a fair and accurate portrayal of a scene. Witnesses can only attest to the fair and accurate portrayal standard by augmenting their recollection with speculation, and because of the psychological effects of suggestibility, are likely to believe the gaps that their memories have filled. Combined with the conflation of illustrative and substantive evidence that photography and video creates, courts are likely to admit substantive evidence for a jury to consider under far lower standards than Rule 901(a) intended. This Comment recommends a new addition to Rule 901 to establish a foundation of authenticity outside of the presence of the jury to mitigate the risk of unfair prejudice. This recommendation aims to alleviate the problem of deepfakes in the courtroom but admittedly does not solve the problem entirely.

Lastly, this Comment concludes that the current legal standard for establishing a foundation to authenticate videos fails to meet the original intent behind the evidence rules of authentication in light of new and continuously developing photographic and video technology. Transitioning to a heightened evidentiary standard is necessary to anticipate the upcoming deepfakes crisis in our courtrooms, rather than reacting to it as the technology permeates our society.

I. DEEPFAKES BACKGROUND

Anyone remotely familiar with graphic design can attest to the relative ease with which various programs, such as Adobe Photoshop, can modify digital images. In fact, a post on Adobe's blog "Adobe Life" invites Photoshop users to "reimagine[] reality."[18] The technology behind deepfakes, however, elevates this ability to a level previously

[18] *How Our Photoshop Floor Reimagines Reality*, ADOBE LIFE BLOG (Apr. 4, 2018), https://blogs.adobe.com/adobelife/2018/04/04/adobe-photoshop-floor [https://perma.cc/GCZ7-3GXY].

unreachable for mainstream graphic design programs. Understanding how deepfakes technology ushers in a new era of manipulation requires grasping two concepts: first, how the creators use machine learning algorithms to generate videos with human likenesses at unprecedented levels of fidelity, and second, how this creation process frustrates current methods of determining real from fake.

A. *Deepfakes Creation Through Generative Adversarial Net Machine Learning Cycles*

The use of advanced machine learning techniques to create fake videos burst onto the scene in December 2017.[19] The near-apocalyptic journalism that followed Motherboard's exposure of the exploits of the "deepfakes" user on Reddit quickly caught the attention of technology commentators,[20] mainstream news outlets,[21] and the government.[22] Although the concept of doctoring digital photography (or other evidence, for that matter) is not new,[23] the budding creation process behind deepfakes enables creators to mimic reality in a devastatingly realistic fashion.

At the core of this new technology is a process called a generative adversarial net (GAN). University of Montreal Ph.D. student Ian Goodfellow led a 2014 scientific paper that first introduced GAN models.[24] In the paper, the authors articulated a process in which two

[19] *See* Cole, *supra* note 5 (describing the fake videos that first introduced the world to the concept of "deepfakes").

[20] *See, e.g.*, Karen Hao, *Deepfakes Have Got Congress Panicking. This Is What It Needs to Do*, MIT TECH. REV. (June 12, 2019), https://www.technologyreview.com/s/613676/deepfakes-ai-congress-politics-election-facebook-social [https://perma.cc/AN78-4BFC] (explaining Congress's early efforts to draft a deepfakes regulation bill to "spark a more nuanced conversation" rather than to actually pass the bill into law).

[21] *See, e.g.*, Kevin Roose, *Here Come the Fake Videos, Too*, N.Y. TIMES (Mar. 4, 2018), https://www.nytimes.com/2018/03/04/technology/fake-videos-deepfakes.html [https://perma.cc/JBT8-F2LH] (highlighting deepfakes' potential to be wielded as an ideological tool).

[22] *See, e.g.*, Deepfakes Report Act of 2019, H.R. 3600, 116th Cong. (2019) (requiring "the Secretary of Homeland Security to publish an annual report" on deepfakes and other digital forgery technology).

[23] *See* David Levi Strauss, *Doctored Photos–The Art of the Altered Image*, TIME (June 13, 2011), https://time.com/3778075/doctored-photos-the-art-of-the-altered-image [https://perma.cc/4LQ8-CHFG] (demonstrating the existence of doctored photography since at least 2011).

[24] Ian J. Goodfellow et al., *Generative Adversarial Nets*, ARXIV (June 10, 2014), https://arxiv.org/pdf/1406.2661.pdf [https://perma.cc/ME86-SN74]. Although

machine learning algorithms are simultaneously pitted against one another.[25] One of these programs is a generative model that creates new data samples; the other, known as a discriminator model, evaluates this data against a training dataset for authenticity.[26] The discriminator model estimates the probability that the sample came from the generative model (a machine creation) or sample data (a real-world reference).[27] These models are known as neural networks because they mimic organic brain function, with interconnected nodes layered to process information far more vast and complex than traditional computer algorithms.[28] These two neural networks operate in a cyclical fashion and learn from each other—the generative model program is learning to create false data, and the discriminator model is learning to identify whether the data is artificial.[29] The result is a process by which each element of the GAN model learns the other's methods in a "constant escalation";[30] the generative model constantly improves its ability to create data sets that have a lower probability of failing the detection algorithm as the discriminator model learns to keep up, a process that continuously improves the fidelity of the creation.[31] This

Goodfellow, now a research scientist at Google's "Brain Team," coined the modern term "GAN" and is credited with materializing GAN coding into reality, the idea of pitting machines against each other to learn has roots in the early years of computer programming. Martin Giles, *The GANfather: The Man Who's Given Machines the Gift of Imagination*, MIT TECH. REV. (Feb. 21, 2018), https://www.technologyreview.com/s/610253/the-ganfather-the-man-whos-given-machines-the-gift-of-imagination [https://perma.cc/4P5N-7AX6]; *see also* A.L. Samuel, *Some Studies in Machine Learning Using the Game of Checkers*, 3 IBM J. RES. & DEV. 210, 211 (1959) (exploring attempts to teach programs to play checkers strategically against one another in the early years of computer science growth).

[25] Chris Nicholson, *A Beginner's Guide to Generative Adversarial Networks (GANs)*, PATHMIND, https://pathmind.com/wiki/generative-adversarial-network-gan [https://perma.cc/JEY9-K283].

[26] *Id.* "The generative model can be thought of as analogous to a team of counterfeiters, trying to produce fake currency and use it without detection, while the discriminative model is analogous to the police, trying to detect the counterfeit currency." Goodfellow et al., *supra* note 24, at 1.

[27] Nicholson, *supra* note 25.

[28] Chris Nicholson, *A Beginner's Guide to Neural Networks and Deep Learning*, PATHMIND, https://pathmind.com/wiki/neural-network [https://perma.cc/WXD6-Y5NS].

[29] Nicholson, *supra* note 25.

[30] *Id.*

[31] *Id.*

continuous process enables the generative model to build a dataset that avoids the pitfalls that would normally give away a fraud.[32]

There are countless commercial and consumer applications of GAN technology. Chris Nicholson[33] has aptly described the breadth of GAN's incredible scientific potential, stating that "[GAN models] can learn to mimic any distribution of data. That is, GANs can be taught to create worlds eerily similar to our own in any domain: images, music, speech, prose. They are robot artists in a sense, and their output is impressive—poignant even."[34] The artistic applications are endless. Some fields, such as the film industry, have already employed ultra-lifelike human likenesses using a variety of methods.[35] Researchers are also developing GAN technology for commercial purposes such as enabling shoppers to picture what an article of clothing looks like on

[32] Spivak, *supra* note 11, at 343–44. Spivak provides a useful illustration of GAN models by applying the cyclical learning process to signature styles of famous authors. A GAN programmer could train a discriminator model to learn the styles of, for example, James Joyce to the point where it can identify the author's prosaic style embedded within other textual samples. *Id.* The generative model then creates new data sets (new pages of prose) for the discriminator to attempt to determine whether the new data was written by the generative model or came from the actual library of James Joyce. *Id.* After the generative model reveals the discriminator model to be right or wrong, the two models repeat the process continuously, with the generator fixing its mistakes until the discriminator can no longer reliably predict the probability of creation versus original. *See id.* Programmers can apply the same process to depictions of human movements and human voice. *Id.* at 351.

[33] Chris Nicholson is the CEO of Pathmind Inc., a Silicon Valley artificial intelligence services provider. PATHMIND, https://pathmind.com/about [https://perma.cc/9FY9-64SP] (last visited August 6, 2020).

[34] Nicholson, *supra* note 25.

[35] *See, e.g.*, ROGUE ONE: A STAR WARS STORY (Lucasfilm 2016). The film prominently features actor Peter Cushing in his 1977 role as Grand Moff Tarkin, twenty-two years after Cushing's death in 1994. Jason Guerrasio, *The Actor Behind the CGI Tarkin in 'Rogue One' Tells Us How He Created the Character*, BUS. INSIDER (Jan. 9, 2017, 12:35 PM), https://www.businessinsider.com/cgi-moff-tarkin-rogue-one-guy-henry-2017-1 [https://perma.cc/WEA4-M6SB]. Industrial Light & Magic used the related but distinct technology of computer graphic imaging with motion capture dots techniques to recreate Cushing's likeness. *Id.* Another example of the film industry's use of GAN technology is *Finding Jack*. *See* FINDING JACK (Magic City Films, forthcoming 2020). The film stars James Dean in a leading role sixty-four years after his death in a 1955 car crash. Jesse Damiani, *James Dean to Be Digitally Resurrected to Appear in His Fourth Film, 'Finding Jack'*, FORBES (Nov. 7, 2019, 8:32 AM), https://www.forbes.com/sites/jessedamiani/2019/11/07/james-dean-to-be-digitally-resurrected-to-appear-in-his-fourth-film-finding-jack/#1d5fff933102 [https://perma.cc/7RCV-PUV4].

a particular person (without the burden of actually trying it on)[36] or devising stronger encryption techniques to protect confidential information and communications online.[37]

Naturally, as benign use of the technology spreads, the dark side of video manipulation is accelerating with equal speed as GAN modeling becomes more widely accessible to those with less noble intentions.[38] Actor and director Jordan Peele created deepfake videos of Barack Obama making speeches that never happened to highlight their danger to civil society.[39] Politicians are a natural target for deepfake creators because of the volume of publicly available photographs and videos of politicians for the creators to utilize. Malign creators, whether domestic or foreign, can use deepfakes to further drive America's political polarization and create the sort of "dystopia" that Jordan Peele warned of in his message.[40]

Further, despite Reddit's and several pornographic websites' efforts to ban deepfake pornography,[41] malicious actors can still create and distribute deepfake celebrity or otherwise nonconsensual pornographic material in other less regulated corners of the internet. As the software to create lifelike deepfakes proliferates, the degree of difficulty and the skill required to create such videos is dropping, leaving convincing and powerful weapons in the hands of a larger number and greater variety of malevolent actors.[42]

[36] Donggeun Yoo et al., *Pixel-Level Domain Transfer*, ARXIV (Nov. 28, 2016), https://arxiv.org/pdf/1603.07442.pdf.

[37] Martín Abadi & David G. Anderson, *Learning to Protect Communications with Adversarial Neural Cryptography*, ARXIV (Oct. 21, 2016), https://arxiv.org/pdf/1610.06918.pdf [https://perma.cc/H4GA-J23Q].

[38] *See* Rory Cellan-Jones, *Deepfakes Videos 'Double in Nine Months'*, BBC (Oct. 7, 2019), https://www.bbc.com/news/technology-49961089 [https://perma.cc/5WBG-UX93] (discussing a September 2019 study from cybersecurity company Deeptrace that found 14,698 deepfake videos online compared to only 7,964 in December 2018).

[39] Kaylee Fagan, *A Viral Video that Appeared to Show Obama Calling Trump a 'Dips–' Shows a Disturbing New Trend Called 'Deepfakes'*, BUS. INSIDER (Apr. 17, 2018, 4:48 PM), https://www.businessinsider.com/obama-deepfake-video-insulting-trump-2018-4 [https://perma.cc/BX63-RVNK].

[40] Roose, *supra* note 21 (predicting that "[p]eople will share them when they're ideologically convenient and dismiss them when they're not").

[41] Janko Roettgers, *Reddit, Twitter Ban Deepfake Celebrity Porn Videos*, NASDAQ (Feb. 7, 2018, 2:11 AM), https://www.nasdaq.com/articles/reddit-twitter-ban-deepfake-celebrity-porn-videos-2018-02-07.

[42] *See* Larry N. Zimmerman, *Cheap and Easily Manipulated Video*, 87 J. KAN. B. ASS'N 20, 20 (2018) (comparing the dismissive attitudes following Hollywood's video

B. The Challenge of Finding Reliable and Lasting Detection Methods

As GAN programming continues to develop and expand, the ability to detect deepfakes becomes increasingly important in a variety of disciplines. The challenge of reliably and consistently detecting deepfakes further evinces the new era of digital forgery that they have ushered in. The challenge stems from the constantly evolving and cyclical method of deepfake creation.[43] The very process that programmers use to create deepfakes relies on incorporating algorithms designed to detect the subsets of data that do not match sample data sets provided to the discriminator model; this cycle's purpose is to root out inconsistencies.[44] This process therefore features a unique defense against programs that detect the frauds—any time a new method of determining whether a video is fake emerges, deepfake creators can use that to their advantage in the GAN cycle.[45]

For example, Associate Professor of Computer Science Siwei Lyu of the University at Albany conducted a study in 2018[46] on the then-current state of deepfake technology with the intent of attempting to pinpoint the reason that the fake videos "felt eerie to him, and not just because he knew they[] [had] been ginned up."[47] Professor Lyu identified one of the signs that a human likeness had been artificially created: there was something wrong with the way that the human depictions blinked.[48] The faces depicted in the deepfakes did not "open and close their eyes at the rates typical of actual humans" because the GAN model simply did not "*get* blinking" (at least not yet).[49]

manipulations in the 1990s with the current reality of software that makes "face-swapping simple for anyone regardless of skill or equipment").

[43] *See infra* Part I.A.

[44] Will Knight, *The US Military Is Funding an Effort to Catch Deepfakes and Other AI Trickery*, MIT TECH. REV. (May 23, 2018), https://www.technologyreview.com/s/611146/the-us-military-is-funding-an-effort-to-catch-deepfakes-and-other-ai-trickery [https://perma.cc/X2NP-AVKX].

[45] Nicholson, *supra* note 25 (comparing this process to the "game of cat and mouse" between a police officer learning to detect false notes and a counterfeiter improving her ability to pass false notes by learning the police officer's methods).

[46] Yuezun Li, Ming-Ching Chang & Siwei Lyu, *In Ictu Oculi: Exposing AI Generated Fake Face Videos by Detecting Eye Blinking*, ARXIV (June 11, 2018), https://arxiv.org/pdf/1806.02877v2.pdf [https://perma.cc/KB8N-YZGT].

[47] Sarah Scoles, *These New Tricks Can Outsmart Deepfake Videos—For Now*, WIRED (Oct. 17, 2018, 7:00 AM), https://www.wired.com/story/these-new-tricks-can-outsmart-deepfake-videosfor-now [https://perma.cc/3G9B-6JEB].

[48] *Id.*

[49] *Id.*

Professor Lyu's paper was a breakthrough in fake video detection by using forensic programs to catch "spontaneous and involuntary physiological activities such as breathing . . . and eye movement, [which] are oftentimes overlooked in the synthesis process of fake videos."[50] For the time being, Professor Lyu had struck a major victory against deepfake creators.

However, while Professor Lyu's success certainly challenged the forgers by rooting out the flaws in their product,[51] the victory was nonetheless muted by the very nature of the deepfake process. Not long after publishing the paper, Lyu's team began to receive anonymous emails that contained deepfake videos whose stars blinked more normally and therefore passed the detection tests his team had created.[52] The creators had incorporated a means of detection that the discriminator algorithm had previously not accounted for strongly enough and provided additional reference points for the algorithm to learn from (for example, pictures and videos of humans with their eyes closed, which were underrepresented in the sample data).[53] The discriminator then did a better job policing the generative model's fakes, essentially teaching the generative model how to overcome its prior weaknesses.[54] The short-lived success of the detection program actually made the forgery mechanism stronger.[55] The result is an "arms race between the creators and the detectors."[56]

Through a program called Media Forensics (MediFor), the Defense Advanced Research Project Agency has been following the challenge of deepfake emergence since even before the videos' namesake Reddit user popularized the concept in December 2017.[57] Among MediFor's lines of effort is an automated system designed to create an "integrity

[50] Li et al., *supra* note 46.

[51] *See supra* notes 47–49 and accompanying text.

[52] Scoles, *supra* note 47.

[53] *Id.*

[54] *Id.*

[55] The process is analogous to bacteria growing stronger by developing immunity to the antibiotics created to defeat them; each advancement in defeating bacteria produces strains of the bacteria naturally resistant to the antibiotic. *How Antibiotic Resistance Happens*, CTRS. FOR DISEASE CONTROL & PREVENTION (Feb. 10, 2020), https://www.cdc.gov/drugresistance/about/how-resistance-happens.html [https://perma.cc/VKS5-7ZZ6].

[56] Scoles, *supra* note 47.

[57] *Media Forensics (MediFor)*, DEF. ADVANCED RES. PROJECT AGENCY, https://www.darpa.mil/program/media-forensics [https://perma.cc/NK5A-XXVA]; *see also* Knight, *supra* note 44.

score" for an image or video, in which the content of the video is compared against a variety of external empirical facts to root out inconsistencies.[58] Efforts such as these will always be chasing the forgers, and their breakthroughs will always provide ammunition to the GAN models.[59] Every data point that gives up a video as fake (such as weather reports to cross-reference against the scene or incorrectly angled shadows that are incongruent with the position of the sun) is a source that deepfake creators can account for by tapping into those data streams for future videos.

C. *Fake Video's Significant Psychological Effects on Viewers*

Fraudulent evidence has always been a concern for courtroom integrity. Yet deepfakes raise an even greater level of concern due not only to their ability to seem real, but also to their impact on viewers. The threat that deepfakes pose to courtroom factfinding is not solely due to the high-fidelity human likenesses that are difficult to detect. The nature of viewing video elucidates psychological effects in which people actually believe that they remember things that they did not actually perceive.[60] This combination is extremely dangerous to witness reliability.

Although many conceive of human memories as an internal video playback system, various studies have shown critical vulnerabilities in

[58] Scoles, *supra* note 47. MediFor is attempting to create an integrity score for videos by layering several test models. *Id.* One model looks for certain background characteristics, such as background noise that is particular to a certain camera model. *Id.* The next looks at physical characteristics, such as whether shadows or reflections are consistent with the location of the light source. *Id.* The last is a "semantic level" model, which compares the video to context that the model knows to be true, such as whether the weather depicted matches the weather report for the date of the scene. *Id.* MediFor seeks to create prototype systems that can stack these levels into a quantifiable "integrity score." *Id.*

[59] For additional attempts to overcome this challenge, see Dr. Herb Lin's article, which suggests the possibility of using digital signatures as a strategy to authenticate digital recordings despite Canon and Nikon's failed attempts to overcome the technological challenge posed by would-be forgers. Herb Lin, *The Danger of Deepfakes: Responding to Bobby Chesney and Danielle Citron*, LAWFARE (Feb. 27, 2018, 7:00 AM), https://www.lawfareblog.com/danger-deepfakes-responding-bobby-chesney-and-danielle-citron [https://perma.cc/67T7-XPR2].

[60] Hadley Leggett, *Fake Video Can Convince Witnesses to Give False Testimony*, WIRED (Sept. 14, 2009, 6:02 PM), https://www.wired.com/2009/09/falsetestimony [https://perma.cc/M88G-8TKJ].

our ability to recall memories accurately.[61] Memory is more comparable to "putting puzzle pieces together than retrieving a video recording,"[62] and is therefore subject to a range of "potential mischief" from both internal and external sources.[63] There are a variety of psychological limitations on the accuracy of human memory; the most relevant to deepfakes is "suggestibility."[64] Suggestibility is a phenomenon that causes a person to implant memories as a result of leading questions, narratives, or visuals when attempting to recall a past experience.[65] Due to suggestibility, reconstruction of an experience in the context of prepared materials or leading questions intended to help tell a desired narrative "can cause the witness'[s] memory to change by unconsciously blending the actual fragments of memory of the event with information provided during the memory retrieval process."[66]

Video exacerbates suggestibility's effect on memory. In 2010, researchers at the University of Warwick conducted a study illustrating the psychological effect that video has on reconstructing personal observations.[67] The researchers placed sixty college students in a room to engage in a computerized gambling task.[68] Following completion of the task, researchers individually showed each subject digitally altered

[61] *See* Mark W. Bennett, *Unspringing the Witness Memory and Demeanor Trap: What Every Judge and Juror Needs to Know About Cognitive Psychology and Witness Credibility*, 64 AM. U. L. REV. 1331, 1335–37, 1352 (2015) (examining a host of challenges to accurate witness testimony and proposing a "Model Plain English Witness Credibility Instruction").

[62] Hal Arkowitz & Scott O. Lilienfeld, *Why Science Tells Us Not to Rely on Eyewitness Accounts*, SCI. AM. (Jan. 1, 2010), https://www.scientificamerican.com/article/do-the-eyes-have-it (quoting psychologist and memory researcher Professor Elizabeth F. Loftus).

[63] Bennett, *supra* note 61, at 1336.

[64] *See id.* at 1342–44 (citing DANIEL L. SCHACTER, THE SEVEN SINS OF MEMORY: HOW THE MIND FORGETS AND REMEMBERS 4 (2001)) (dividing the malfunctions of memory into seven categories: transience, absent-mindedness, blocking, misattribution, bias, persistence, and, most relevant here, suggestibility).

[65] SCHACTER, *supra* note 64, at 5.

[66] *See* Richard S. Schmechel et al., *Beyond the Ken? Testing Jurors' Understanding of Eyewitness Reliability Evidence*, 46 JURIMETRICS 177, 195 (2006) (presenting an independent study of the ability of potential jurors in the District of Columbia to understand limitations on the reliability of eyewitness identification under various strenuous circumstances).

[67] Kimberly A. Wade, Sarah L. Green & Robert A. Nash, *Can Fabricated Evidence Induce False Eyewitness Testimony?*, 24 APPLIED COGNITIVE PSYCHOL. 899, 900 (2010).

[68] *Id.* at 901–02.

video depicting a co-subject cheating, when in fact none of the subjects had actually cheated.[69] Nearly half of the subjects were willing to testify that they had personally witnessed a co-subject cheating after seeing the fake video; only one in ten was willing to testify to the same effect after the researcher merely told the subject about the cheating, rather than showing the fake video evidence.[70]

Consequently, deepfakes can have a devastating effect on courtroom integrity. If a party submits a deepfake video to the court, its deceptive harm is not limited solely to the video itself. The lies embedded within a fake video cascade into other portions of the proceedings; viewing fake videos is likely to affect the testimony of witnesses concerning their recollection of events.[71] The legal standard to admit video evidence into a courtroom for a jury to see is unfortunately ill-equipped to address this level of risk.

II. AUTHENTICATING PHOTOGRAPHIC AND VIDEO EVIDENCE

While technology generally outpaces the law, it is imperative to discern whether the contemporary legal framework is sufficient to address the potential harm that technological advances present. Some scholars and commentators have grappled with the interplay of deepfakes with privacy law, First Amendment rights, and regulatory challenges.[72] Additionally, deepfakes bring the possibility of unprecedented levels of distrust in the government and other public institutions if videos emerge featuring public figures saying or doing things that never happened.[73] Among the challenges specific to trust in public

[69] *Id.* at 903–04.

[70] Leggett, *supra* note 60. "[R]esearchers emphasized that no one should testify unless they were 100 percent sure they had seen their partner cheat." *Id.*

[71] Wade et al., *supra* note 67, at 901, 904.

[72] *See, e.g.*, Daniel de Zayas, *Legal Means to Prosecute Actors Behind Deepfakes*, AM. U. NAT'L SECURITY L. BRIEF (Sept. 23, 2019), https://nationalsecuritylawbrief.com/2019/09/23/legal-means-to-prosecute-actors-behind-deepfakes [https://perma.cc/P89N-CFAD] (identifying pre-existing legislation to prosecute creators and distributors of deepfakes); Harris, *supra* note 11, at 107, 110–11 (examining the insufficiency of nonconsensual pornography laws to address the deepfake crisis); Spivak, *supra* note 11, at 358–62 (contrasting First Amendment jurisprudence on obscenities and child pornography with how a court would likely rule on deepfakes).

[73] *See* Bobby Chesney & Danielle Citron, *Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security*, 107 CALIF. L. REV. 1753, 1779 (2019) (illustrating how deepfakes may be used to harm society and cautioning that public institutions in which

institutions is that which courtrooms will face in light of the current standards used to admit digital photography and video as evidence.

Common law standards initially governed the admissibility of photographic and video evidence; the *McKeever* test, originally a standard for admitting audio recordings, stood as a model for admissibility for decades.[74] The *McKeever* test began as a strict standard, but it eventually became more flexible as photographic and video evidence became more common in courtrooms.[75] The *McKeever* test later gave way to the Federal Rules of Evidence, which codified the test's main components.[76] As states codified their own evidence standards based on the Federal Rules of Evidence, courts began to use two theories—the pictorial communication theory and the silent witness theory—to authenticate photographic and video evidence under Rule 901(b).[77] This Part discusses the history of the standard of admissibility of photographic and video evidence, two common theories under which courts admit such evidence, and the guide that Federal Rule of Evidence 602 provides for authenticating such evidence.

A. The Evolution of Photographic and Video Evidence Authentication

Suspicion of the susceptibility of photographic and video evidence[78] to modification or tampering is nothing new to courtrooms; courts have articulated their concerns over photographs and motion pictures since the invention of photography, and such concern continued as photography became more prevalent in society.[79] The modern standard

the public's trust may be eroded by deepfakes "includ[e] elected officials, appointed officials, judges, juries, legislators, staffers, and agencies").

[74] *See infra* notes 80–83 and accompanying text.

[75] *See infra* notes 80–86 and accompanying text.

[76] *See infra* notes 87–88 and accompanying text.

[77] *See infra* Part II.B.1–2.

[78] The Federal Rules of Evidence define a photograph as "a photographic image or its equivalent stored in any form." FED. R. EVID. 1001(c). Video is treated largely similarly to digital and traditional photography for authentication. *See, e.g.*, Linde v. Arab Bank, PLC, 97 F. Supp. 3d 287, 338 (E.D.N.Y. 2015) (authenticating videos "on the same principles as still photographs"), *vacated on other grounds*, 882 F.3d 314 (2d Cir. 2018). For the purposes of this Comment, photographic evidence generally refers to video evidence as well.

[79] *See, e.g.*, Cowley v. People, 83 N.Y. 464, 478 (1881) ("The portrait and the photograph may err, and so may the witness. That is an infirmity to which all human testimony is lamentably liable."); Gibson v. Gunn, 202 N.Y.S. 19, 20 (App. Div. 1923) (per curiam) (commenting that "moving pictures present a fertile field for exaggeration

for video authentication prior to admission initially mirrored the strict standards that courts used for sound recordings.[80] For decades, courts used the seven-part *McKeever* test[81] as the standard to admit sound recordings as evidence.[82] The *McKeever* test required the proponent to establish authenticity based on seven elements at a hearing prior to admission and was eventually expanded to include video evidence.[83]

As photographs, motion pictures, and recordings became more familiar and common in daily life, their use in court expanded.[84] Accordingly, courts loosened the *McKeever* test over time and eventually set it aside in favor of more lenient standards.[85] Interpreting the *McKeever* test as "a guide rather than a rule," and adopting more relaxed tests, courts determined that trial judges should have "wide latitude" to determine whether a proponent of recordings had laid a sufficient foundation for a reasonable jury to conclude that it was authentic.[86]

of any emotion or action" while separately considering the manipulative effect of its lack of relevance).

[80] Jill Witkowski, Note, *Can Juries Really Believe What They See? New Foundational Requirements for the Authentication of Digital Images*, 10 WASH. U. J.L. & POL'Y 267, 279 (2002).

[81] United States v. McKeever, 169 F. Supp. 426, 430 (S.D.N.Y. 1958) (requiring that the proponent show: "(1) That the recording device was capable of taking the conversation now offered in evidence. (2) That the operator of the device was competent to operate the device. (3) That the recording is authentic and correct. (4) That changes, additions or deletions have not been made in the recording. (5) That the recording has been preserved in a manner that is shown to the court. (6) That the speakers are identified. (7) That the conversation elicited was made voluntarily and in good faith, without any kind of inducement"), *rev'd on other grounds*, 271 F.2d 669 (2d Cir. 1959).

[82] Witkowski, *supra* note 80, at 276–77.

[83] *Id.* at 279 (citing *McKeever*, 169 F. Supp. at 374–75).

[84] *See* 2 KENNETH S. BROUN ET AL., MCCORMICK ON EVIDENCE § 215 (7th ed. 2013) (describing the different ways that photographs are used in courts). "As judges, counsel and the lay public have become accustomed to the prevalence of such recordings in court, their persuasive potential is both widely acknowledged and the subject of concern." *Id.* § 216, at 35.

[85] Witkowski, *supra* note 80, at 279 ("Over time, however, the courts replaced the strict foundational requirements concerning the process of taking motion pictures with the admission of witness testimony that the film was a fair and accurate representation of what actually happened."); *see also* EDWARD J. IMWINKELRIED, EVIDENTIARY FOUNDATIONS § 4.09[2] (9th ed. 2015) (stating that although "the courts were initially very conservative in their treatment of motion pictures," "[t]he law governing the admission of motion pictures has been liberalized in recent years").

[86] Witkowski, *supra* note 80, at 278; *see also* United States v. Branch, 970 F.2d 1368, 1371–72 (4th Cir. 1992) (finding the *McKeever* factors sufficient but not required to establish a foundation for authenticity); United States v. Biggins, 551 F.2d 64, 66–67

The authentication standard eventually transitioned from the common law to codification after Congress passed the Federal Rules of Evidence in 1975 after decades of study, delay, and deliberation.[87] The rules reflected the standards for admissibility of videos that courts had adopted since relaxing the *McKeever* test: relevance (codified in Rule 401), probative value balanced against undue prejudice (codified in Rule 403), and accuracy (codified in the sufficient to support a finding standard in Rule 901).[88] Forty-two states have adopted the Uniform Rules of Evidence (based on the Federal Rules of Evidence).[89]

The authenticity of evidence is ultimately a factual determination for the trier of fact (typically, but not necessarily, a jury) to evaluate.[90] However, before a court admits evidence for the jury to consider, the court "must determine whether its proponent has offered a satisfactory foundation from which the jury could reasonably find that the evidence is authentic."[91] The process by which a judge addresses proper foundation for authentication does not itself establish evidence as authentic; the jury is still responsible for the ultimate determination of authenticity and therefore credibility.[92]

(5th Cir. 1977) (holding that the court "neither adopt[ed] nor reject[ed] [the *McKeever* test] as a whole" and looking to four factors as a guideline that is not intended to "sacrifice[] [evidence] to a formalistic adherence to the standard [the court] establish[ed]").

[87] An Act to Establish Rules of Evidence for Certain Courts and Proceedings, Pub. L. No. 93-595, 88 Stat. 1926 (1975) (codified as amended at 28 U.S.C. §§ 2072–2074 (2018)). The Judicial Conference responsible for implementing the Rules Enabling Act of 1934 did not formally study a uniform evidence code until 1961 and finally submitted its proposed rules to Congress for approval in 1972. Paul R. Rice & Neals-Erik William Delker, *A Short History of Too Little Consequence*, 191 F.R.D. 678, 682–84 (2000).

[88] Witkowski, *supra* note 80, at 279–80; *see* FED. R. EVID. 401, 403, 901.

[89] GREGORY P. JOSEPH, MODERN VISUAL EVIDENCE § 1.02 (2005) (explaining that "[e]ven in states without codification, the courts frequently look to the Federal Rules for guidance, occasionally going so far as to adopt particular rules as a matter of decisional law. The Federal Rules of Evidence have thus come to set the standard of evidence law nationally, in the state as well as the federal courts").

[90] United States v. Branch, 970 F.2d 1368, 1370 (4th Cir. 1992) (citing FED. R. EVID. 104 advisory committee's note to subdivision (b) ("If the evidence is not such as to allow a finding [that a jury could reasonably conclude authenticity], the judge withdraws the matter from their consideration.")).

[91] *Id. See generally* IMWINKELRIED, *supra* note 85, § 4.01[1] (outlining the procedure for authentication under Rule 901).

[92] *Branch*, 970 F.2d at 1370–71.

Rule 901(a) states that to establish a proper foundation for authentication evidence, "the proponent must produce evidence sufficient to support a finding that the item is what the proponent claims it is."[93] While Rule 901(a) is not particularly specific in its mandate, Rule 901(b) provides a variety of means through which a party can satisfy Rule 901(a), such as nonexpert opinions about handwriting or evidence derived from public records.[94] Rule 901(b), however, is not exhaustive; there are other means of satisfying Rule 901(a)'s sufficient evidence standard, such as through circumstantial evidence that provides indicia of authenticity.[95]

B. Theories of Authenticating Photographic and Video Evidence

In alignment with Rule 901(b)'s various means of authenticating evidence, courts typically admit photographic evidence under one of two theories: the "pictorial communication" theory and the "silent witness" theory.[96] Each theory utilizes a different sub-section of Rule 901(b) to meet Rule 901(a)'s sufficient evidence standard for authentication.[97]

The logic behind distinct foundational standards for the pictorial communication theory and silent witness theory hinges on the intended purpose of substantive as opposed to illustrative evidence. Substantive evidence provides an "independent probative value for proving a fact," such as a physical object recovered from a scene relevant to the case.[98] Illustrative evidence, on the other hand, accompanies witness testimony and is intended to "aid the trier [of fact] in understanding the witness's testimony."[99] The distinction is important but problematic in the context of photographs and videos because illustrative evidence often becomes substantive by showing the jury more than the witness can recollect or convey, thereby introducing

[93] FED. R. EVID. 901(a).

[94] FED. R. EVID. 901(b)(2), 901(b)(7).

[95] FED. R. EVID. 901 advisory committee's note to subdivision (b) ("The examples are not intended as an exclusive enumeration of allowable methods but are meant to guide and suggest, leaving room for growth and development in this area of law."); PAUL R. RICE & ROY A. KATRIEL, EVIDENCE: COMMON LAW AND FEDERAL RULES OF EVIDENCE § 7.02[A][3][a] (5th ed. 2005) ("[T]here are no inherent limitations on the means by which one can circumstantially authenticate a piece of evidence."); *see infra* Part III.C (recommending requiring the use of means other than Rule 901(b)(1) to establish a foundation for video evidence).

[96] 2 BROUN ET AL., *supra* note 84, § 215.

[97] *Id.*; *see infra* Parts II.B.1–2.

[98] 2 BROUN ET AL., *supra* note 84, § 212.

[99] *Id.*

independent, substantive evidence for which there is no foundation.[100] Nonetheless, the pictorial and silent witness theories derive their separate standards from the supposition that illustrative evidence is limited to the perceptions and recollections of the witness's testimony.[101]

1. The pictorial communication theory

Courts most commonly admit photographic evidence as illustrative evidence, intended to accompany a witness's testimony.[102] This application of photographic evidence is known as the pictorial communication theory, in which photographic evidence is intended to be viewed "merely as a graphic portrayal" to supplement a witness's oral testimony.[103] Under the pictorial communication theory, the typical means of establishing a foundation for authentication is Rule 901(b)(1), which provides that a "[w]itness with [k]nowledge" testify that an item is what it is claimed to be.[104]

Rule 901(b)(1)'s method for establishing an evidentiary foundation is nearly as vague as Rule 901(a)'s standard that it seeks to meet. Applying Rule 901(b)(1), a proponent establishes a foundation for photographic evidence if a witness testifies that the photograph is a "correct and accurate representation of relevant facts personally observed by the witness."[105] Courts commonly refer to this rule as the "fair and accurate portrayal" standard.[106]

The fair and accurate portrayal standard assumes that video is difficult to alter—the standard is rooted in an age of traditional film photography, prior to the advent of digital photography and other media.[107] Traditional photography differs from digital media (whether

[100] *Id.*; *see infra* Part III.A (arguing that the vast substantive detail that video conveys to a jury exceeds the illustrative intent of the pictorial communication theory).

[101] 2 BROUN ET AL., *supra* note 84, § 215.

[102] *Id.*

[103] *Id.*

[104] FED. R. EVID. 901(b)(1).

[105] 2 BROUN ET AL., *supra* note 84, § 215.

[106] *Id.* The fair and accurate portrayal standard was a common law standard prior to the adoption of the Federal Rules of Evidence, at which point it was incorporated into applying Rule 901(b)(1) to photographic evidence. *Id.* "Rule 901 is little more than a delineation of the methods of authentication that courts recognized under the common law." RICE & KATRIEL, *supra* note 95, § 7.02[B][1][a]; *see, e.g.*, Kooyumjian v. Stevens, 135 N.E.2d 146, 151 (Ill. App. Ct. 1956) (applying the common law principle of fair and accurate portrayal prior to the adoption of the Federal Rules of Evidence).

[107] Witkowski, *supra* note 80, at 282 & n.65.

still photography or video) in several ways.[108] The most relevant difference is that digital media stores individual pixels as data in an electronic file; there is no traditional original image that exists with, for example, older thirty-five millimeter film cameras.[109] Traditional film cameras capture light data as imprinted onto physical film, which can then be protected through a secure chain of custody.[110] Digital photography, however, as a "finite set of ones and zeroes," makes determining whether a digital photograph is an original or a copy nearly impossible.[111]

Additionally, because early digital photography featured lower initial image quality compared to film photography, its proponents commonly needed to enhance digital photographs to aid the trier of fact.[112] Thus, an abundance of cases have addressed the issue of non-insidious modifications of video, such as editing, enhancing, taping over, or curating certain portions of a longer video or recording.[113] In these commonplace instances, courts have required no more than satisfaction of the fair and accurate portrayal standard—or the "evidence as a process or system" standard if admitted under the silent witness theory[114]—to admit the recording.[115] For example, the Supreme

[108] *See id.* at 269–71 (outlining the digital image creation process in scientific detail concerning image compression and physical characteristics).

[109] *Id.* at 272–73.

[110] *Id.* at 268 n.3, 272.

[111] *Id.* at 272. *But see* John M. Facciola & Lindsey Barrett, *Law of the Foal: Careful Steps Towards Digital Competence in Proposed Rules 902(13) and 902(14)*, 1 GEO. L. TECH. REV. 6, 11–12 (2016) (explaining how iPhone software captures the date, time, and GPS coordinates of pictures as metadata while subsequently acknowledging the possibility that it could be altered); Lin, *supra* note 59 (suggesting the possibility of "digital signatures" to ensure image security).

[112] Witkowski, *supra* note 80, at 269 n.6, 271 n.16 (citing Herb Blitzer, *Creating the Digital Image SOP*, L. ENFORCEMENT TECH. 58–61 (June 2000), http://desksgt.com/Classes/Reading/digitalimagesop.pdf [https://perma.cc/S29H-EK9V]). "In general, both traditional photographs and digital images often need to be enhanced. Enhancing an image involves adjusting the contrast so that the picture is clearer." Witkowski, *supra* note 80, at 271 n.17.

[113] *See, e.g.*, United States v. Seifert, 445 F.3d 1043, 1045–46 (8th Cir. 2006) (admitting digitally enhanced surveillance tape after expert video analyst's testimony about each step of the digital enhancement process); United States v. Mills, 194 F.3d 1108, 1111–12 (10th Cir. 1999) (admitting an incomplete videotape where an officer responsible for filming testified as to authenticity of the tape and confirmed that, "except for the deleted portion, it accurately depicted the entire episode").

[114] *See infra* Part II.B.2 (addressing the more demanding requirement for the silent witness theory).

[115] FED. R. EVID. 901(b)(9).

Court of Arkansas drew a careful distinction between video that had been "enhanced" by adjusting the brightness and contrast of the video with that which was "altered," such as by changing the "face, features, or physique of someone not present in the original videotape."[116] The court dismissed the defendant's contention that the video had been manipulated by stating that the jury had ample opportunity to determine whether any alterations were present.[117] In these types of cases, courts address both whether the alteration process distorted the image such that the resulting product remains authentic as well as whether the curation conveys a message so different from the original that it is no longer "relevant" under Rule 403.[118] For both issues, courts envision having the "original" recording to reference against;[119] courts rarely consider the possibility of outright forgery when considering authentication standards for admission.[120] The rare cases when courts reject photographic evidence are when there is no authenticating witness or the witness expressly rejects the photograph as an accurate depiction.[121] This was the case in *United States v. Lawson*,[122] where the defendant offered photographs that were excluded from evidence because the only witness at trial testified that the photographs "did not accurately reflect what he saw."[123]

Because of this traditional framework, the fair and accurate portrayal standard is not a difficult hurdle to clear. A witness who testifies as to a photograph's or video's accuracy does not need to be the actual photographer or understand the process by which the originator created it.[124] The standard to establish a foundation is so minimal that issues

[116] Nooner v. State, 907 S.W.2d 677, 686 (Ark. 1995); *see also* Louis Vuitton S.A. v. Spencer Handbags Corp., 765 F.2d 966, 973–74 (2d Cir. 1985) (finding sufficient authentication in the absence of any accusations of inaccuracy or tampering).

[117] *Nooner*, 907 S.W.2d at 686.

[118] 2 BROUN ET AL., *supra* note 84, § 215.

[119] Witkowski, *supra* note 80, at 272.

[120] *Id.* at 285–86 (considering various reasons for the "infrequency of challenges to digital images," including general lack of awareness and a focus on editing, not forgery); *see infra* notes 201–05 and accompanying text.

[121] *See, e.g.*, United States v. Lawson, 494 F.3d 1046, 1052 (D.C. Cir. 2007) (determining that the trial court properly excluded photographs from evidence because they were not authenticated by the only witness familiar with the scene).

[122] 494 F.3d 1046 (D.C. Cir. 2007).

[123] *Id.* at 1052.

[124] *See, e.g.*, Kooyumjian v. Stevens, 135 N.E.2d 146, 151 (Ill. App. Ct. 1956) (admitting photographs when the authenticating witness did not know when the

concerning the possibility that the witness's fair and accurate testimony is "limited" or "defective" or that the witness is "otherwise unsure of his perceptions" are matters saved for the jury, to which the jury must assign weight to evaluate the evidence's credibility—not matters of admissibility with which the proponent of the evidence must grapple.[125] Instead, the standard imposes only a "sufficient to support a finding" requirement on the proponent.[126]

2. *The silent witness theory*

In addition to the pictorial communication theory, a party may also submit a photograph or video as substantive evidence—that is, the photograph or video is capable of standing on its own to convey what it depicts and, in turn, obviates the need for a witness.[127] Courts admit photographic evidence in this manner under the silent witness theory.[128] By treating evidence as a "potential independent source[] of substantive information for the trier of fact," the silent witness theory has stricter requirements for the admission of photographic evidence than the pictorial communication theory's requirements for admission.[129] Evidence admitted under the silent witness theory is generally subject to Rule 901(b)(9), which allows a proponent of evidence to establish a foundation for authentication by "describing a process or system and showing that it produces an accurate result."[130]

One of the most common examples of evidence admitted under the silent witness theory is security camera footage. Typically, when a party submits video from a closed-circuit television (CCTV) device at, for example, a bank or convenience store, a worker or expert will testify as to the reliability of the video and the process for maintaining an accurate

pictures were taken); State v. Pearson, 975 So. 2d 646, 655 (La. Ct. App. 2007) (finding proper establishment of foundation even though photographer did not testify).

[125] 31 CHARLES ALAN WRIGHT & ARTHUR R. MILLER, FEDERAL PRACTICE AND PROCEDURE, § 7106 (1st ed.), Westlaw (database updated Apr. 2020).

[126] *Id.*

[127] *See* 2 BROUN ET AL., *supra* note 84, § 216 ("[I]t is important for courts to acknowledge that films and videos are often not merely illustrative of a witness's testimony, but are potential independent sources of substantive information for the trier of fact."); *see also supra* note 89 (discussing similar authentication treatment for photographs and videos).

[128] 2 BROUN ET AL., *supra* note 84, § 216.

[129] *Id.*

[130] FED. R. EVID. 901(b)(9).

system.[131] For example, in *United States v. Rembert*,[132] the government offered no witnesses to testify that a CCTV video fairly and accurately depicted the scene; instead, a bank employee testified as to how the cameras were loaded, how the results were secured, and the internal metadata concerning the date and location of the filming.[133] Courts commonly accept details of this nature when the cameras are part of a regulated system that is maintained and operated according to accepted standards, such as those of a police department or bank security system.[134]

Because evidence admitted under the silent witness theory may stand alone as substantive evidence without accompanying witness testimony, courts generally only admit it when the device and process are set up and executed in a controlled environment. Courts have accepted testimony concerning the process and system as it applies to CCTV surveillance videos as described above, as well as x-ray photography and police footage.[135] However, digital photography that the general public personally creates falls largely beyond this threshold because it lacks a systematic and reliable scientific process and because the proponent cannot demonstrate a secure chain of custody.[136] For example, even though surveillance or police footage is digitally created, the chain of custody (generally secured through police channels) insulates the product from tampering, and therefore the footage may potentially stand on its own as substantive in ways that evidence admitted under the pictorial communication model theoretically could not.[137]

Over the past several decades, courts have begun to test the digital boundaries of the silent witness theory. For example, in an instance

[131] 2 BROUN ET AL., *supra* note 84, § 216.

[132] 863 F.2d 1023 (D.C. Cir. 1988).

[133] *Id.* at 1028 (rejecting a heightened standard of evidentiary authentication in criminal proceedings).

[134] 2 BROUN ET AL., *supra* note 84, § 215.

[135] *See, e.g.*, United States v. Stephens, 202 F. Supp. 2d 1361, 1368 (N.D. Ga. 2002) (admitting surveillance video under the silent witness theory after police official testified as to the process and "general reliability of the entire system"); Woodward v. State, 123 So. 3d 989, 1027 (Ala. Crim. App. 2011) (applying the silent witness theory and upholding the validity of video footage from a patrol car as a sufficiently reliable mechanism capable of accurately recording a criminal shooting); People v. Bowley, 382 P.2d 591, 595 (Cal. 1963) (holding x-ray imaging admissible under the silent witness theory since no one can testify to the accuracy of an image, because it is not possible to directly observe the inside of a body).

[136] 2 BROUN ET AL., *supra* note 84, § 215.

[137] *Id.*

where a police officer took a cell phone video recording of a CCTV surveillance video system of a convenience store, the state failed to establish a foundation when the police officer testified that his video was a fair and accurate portrayal of what the CCTV depicted.[138] The officer's fair and accurate portrayal testimony was insufficient where he could only speak to his knowledge of the depiction of his cell phone tape; in other words, the officer had no more personal knowledge that the video of the scene of the crime was a fair and accurate portrayal than anyone else.[139] Without Rule 901(b)(9) evidence concerning the reliability of the CCTV itself, the recording was inadmissible.[140] Cell phone videos present particularly unique challenges in the silent witness theory context because of the lack of reliability concerning the process and preparation of such videos. Courts have distinguished video recordings originating from cameras worn by an undercover police officer and prepared by state officials from videos taken by an undercover officer with a cell phone in otherwise the same context.[141] In *McFall v. State*,[142] the Court of Appeals of Indiana addressed this very issue when the prosecution introduced evidence of a controlled drug buy using video from a confidential informant's cell phone.[143] Whereas normally police officers equip an informant with government owned and managed recording equipment and secure it from the informant following an operation, here the detective did not exercise control over the informant's cell phone and filming process throughout the operation.[144] The prosecution therefore could not attest to the accuracy of a process or system under Rule 901(b)(9) because the informant's personal phone was not subject to the same standard operating procedures and chain of custody that the police use for typical surveillance equipment.[145]

[138] State v. Moore, 803 S.E.2d 196, 210 (N.C. Ct. App. 2017).

[139] *Id.* ("No witness was asked whether the video accurately depicted events that he had observed, and no testimony was offered on the subject.").

[140] *Id.*

[141] McFall v. State, 71 N.E.3d 383, 388 (Ind. Ct. App. 2017).

[142] 71 N.E.3d 383 (Ind. Ct. App. 2017).

[143] *Id.* at 388–89 (rejecting the trial court's admission of the confidential informant's cell phone footage under the silent witness theory but ultimately rendering the error harmless because the defendant "identified herself in the videos . . . and acknowledged that the events depicted in them occurred on [the date in question]").

[144] *Id.* at 388.

[145] *Id.*; FED. R. EVID. 901(b)(9). However, had the government presented an authentication witness with personal knowledge of the depiction itself, the court could have admitted the video under Rule 901(b)(1). *See* United States v. Richardson, 562

These cases demonstrate courts' acknowledgement of the risk that digital photography poses and their hesitance to incorporate it into the silent witness theory without an authenticating witness. Despite these risks, courts have refused to incorporate any changes to the pictorial communication standard when it comes to digital photography.[146]

C. Rule 602 Caselaw Establishes a Baseline for Distinguishing Personal Knowledge from Speculation and Logically Applies to Rule 901(b)(1) Witnesses

Normally, a judge will not exclude an eyewitness if her memory or perception is limited; as long as the testimony could assist a reasonable trier of fact in establishing the facts, the court will allow the witness to testify.[147] However, a judge has discretion to exclude evidence (prior to its admission) when a witness's personal knowledge is particularly uncertain or unreliable or when there is not enough evidence that a reasonable juror could give some weight to the testimony.[148] For example, in *Nolin v. Douglas County*,[149] the judge did not admit a document when the witness stated that he was only "somewhat familiar with the document."[150] Thus, judges must walk a fine line between the minimum amount of personal knowledge required to testify and imperfect knowledge that crosses the threshold into speculation.

This fine line determines whether a witness has the requisite personal knowledge to testify to the fair and accurate portrayal standard to establish a foundation of authenticity under Rule 901(b)(1). Since Rule 901(b)(1) does not specifically define knowledge, other sections of the

F.2d 476, 479 (7th Cir. 1977) (admitting bank surveillance film despite the prosecution's inability to meet the Rule 901(b)(9) standard due to lack of secure chain of custody because eyewitnesses testified to the fair and accurate standard under Rule 901(b)(1)).

[146] *See, e.g.*, Owens v. State, 214 S.W.3d 849, 854 (Ark. 2005) ("[W]e do not agree that this court should impose a higher burden of proof for the admissibility of digital photographs merely because digital images are easier to manipulate.").

[147] *See supra* notes 125–26 and accompanying text.

[148] 27 CHARLES ALAN WRIGHT & ARTHUR R. MILLER, FEDERAL PRACTICE AND PROCEDURE, § 6027 (2d ed.), Westlaw (database updated Apr. 2020).

[149] 903 F.2d 1546 (11th Cir. 1990).

[150] *Id.* at 1552. Rule 901(b)(1) is not specific to video, and in this case the witness's uncertainty applies more broadly to a knowledgeable witness rather than the "fair and accurate" standard for photographs. *See, e.g.*, United States v. Crute, 238 F. App'x 903, 905–06 (3d Cir. 2007) (authenticating vehicle registration records through a knowledgeable witness); Kruse v. Hawai'i, 857 F. Supp. 741, 745–46 n.5 (D. Haw. 1994) (authenticating hospital records), *aff'd*, 68 F.3d 331 (9th Cir. 1995).

Federal Rules of Evidence are instructive.[151] The most relevant section in this context is Rule 602, which requires witnesses to have personal knowledge of the matters about which they testify. Rule 702 allows expert testimony based on "scientific, technical, or other specialized knowledge;"[152] because most witnesses with potential fair and accurate portrayal testimony will not have such expertise, Rule 602's personal knowledge requirement is a more appropriate standard for knowledge than Rule 702 in this context.

Rule 602 requires that a witness have personal knowledge of the matter about which she is testifying for the testimony to be relevant.[153] Because other subdivisions of Rule 901(b) describe means of authentication based on either personal or specialized knowledge,[154] Rule 602 and its associated caselaw applies to Rule 901(b)(1) by logical extension despite the lack of a definition of knowledge in Rule 901(b)(1) itself. Thus, examining the Rule 602 standard for personal knowledge helps articulate the requirement for whether a witness testifying to the fair and accurate portrayal standard has the requisite personal knowledge for Rule 901(b)(1). The Rule 602 standard helps define the line between personal knowledge shortcomings that pass the foundational requirements for a jury to consider and those that the court rejects at the foundational stage as speculative, as was the case in *Nolin*.[155]

In applying Rule 602 for determining personal knowledge, courts have long resisted refusing to allow a witness to testify merely because the court believes the witness to be obviously mistaken or dishonest.[156] The only appropriate circumstance for a court to reject a witness's testimony is when no reasonable trier of fact could believe that a witness

[151] *See* 31 WRIGHT & MILLER, *supra* note 125, § 7106 ("The fact that Rule 901(b)(1) uses the word 'knowledge' without restrictions or modifiers suggests that authentication testimony may be based on knowledge of the sort described by either Rule 602 or Rule 702.").

[152] FED. R. EVID. 702.

[153] FED. R. EVID. 602. The witness must demonstrate personal knowledge on the matter by a preponderance of the evidence under Rule 104(a). Miller v. Keating, 754 F.2d 507, 511 (3d Cir. 1985).

[154] 31 WRIGHT & MILLER, *supra* note 125, § 7106 (referring to FED. R. EVID. 901(b)(5) (stating that voice may be identified based on a witness hearing it "firsthand") and FED. R. EVID. 901(b)(3) (explaining authentication through expert's comparison with specimen authenticated by another)).

[155] Nolin v. Douglas Cty., 903 F.2d 1546, 1552 (11th Cir. 1990).

[156] EDMUND M. MORGAN, BASIC PROBLEMS OF STATE AND FEDERAL EVIDENCE 53–54 (Jack B. Weinstein, 5th ed. 1976).

perceived what she claims.[157] Courts' inclination is for the jury, as the trier of fact, to assign weight to testimony in accordance with its perception of the witness's reliability and other factors to aid in its judgment.[158] Personal knowledge of objects or events under Rule 602 is comprised of four elements: "(1) sensory perception; (2) comprehension about what was perceived; (3) present recollection; and (4) the ability to testify about what was perceived."[159] Each of these four elements is required for a judge to allow a jury to hear a witness's testimony.[160]

The first requirement for personal knowledge under Rule 602 is sensory perception, which courts commonly label "observation."[161] Although this shorthand most immediately invokes sight, sensory perception may be based on any of the five senses.[162] To satisfy the sensory perception element, the witness must have the ability to perceive and must in fact have perceived what she is testifying to; the witness's ability, however, may be limited, or even minimal.[163] Courts have long recognized that the personal knowledge standard to admit a witness's testimony does not require positive or absolute certainty.[164]

For a court to exclude a witness for lack of sensory perception, the witness must have not been able to perceive relevant facts directly. For

[157] *Id.*

[158] *See* 27 WRIGHT & MILLER, *supra* note 148, § 6027 (summarizing the threshold for satisfying Rule 602 by noting that "[t]he judge should allow the testimony to go to the jury unless the judge concludes the foundation for personal knowledge is so weak that the testimony will be a waste of time").

[159] Keiser v. Borough of Carlisle, No. 1:15-CV-450, 2017 WL 4075057, at *5 (M.D. Pa. Sept. 14, 2017); *see also* 2 JOHN HENRY WIGMORE, EVIDENCE IN TRIALS AT COMMON LAW § 478 (James H. Chadbourn ed., 1979) (generally outlining observation/perception, recollection, and communication as requirements for testimonial assertions).

[160] *Keiser*, 2017 WL 4075057, at *5.

[161] 27 WRIGHT & MILLER, *supra* note 148, § 6023.

[162] *See* Fox v. Order of United Commercial Travelers of Am., 192 F.2d 844, 846 (5th Cir. 1951) ("A witness may testify to what he hears, feels, tastes, and smells, as well as to what he sees, and regardless of whether he sees anything.").

[163] *See, e.g.*, Auerbach v. United States, 136 F.2d 882, 885 (6th Cir. 1943) (witness identified defendant's voice "to the best of his belief" and acknowledged that "it was possible that he could be mistaken").

[164] *See, e.g.*, United States v. Hickey, 917 F.2d 901, 904–05 (6th Cir. 1990) ("Despite the fact that . . . [the witness's] perception was sometimes impaired, a reasonable or rational juror could believe that [the witness] . . . perceived the course of events to which [he] testified."); United States v. Evans, 484 F.2d 1178, 1181 (2d Cir. 1973) (applying personal knowledge standards to the competency of a witness to make an in-court identification).

example, in *State v. Tutt*,[165] when "it was dark, [and the witness] could[] n[o]t make out exactly what was happening," the court precluded the witness from testifying because of an inability to visually perceive what she purported to testify to.[166] Similarly, in *McCrary-El v. Shaw*,[167] the Eighth Circuit affirmed the trial court's exclusion of the deposition of a witness who claimed to have seen a confrontation between the defendant and several correctional officers from an adjoining jail cell.[168] The court reviewed a diagram of the jail layout and found that no reasonable person could conclude that the witness could see anything of relevance.[169] As these cases demonstrate, the personal knowledge standard allows a witness's limitations and gaps in perception but not a complete inability to perceive.[170]

The second element of personal knowledge is recollection, which, like sensory perception, does not need to be perfect to satisfy the test. Of course, no human memory is flawless. Incomplete or limited memory is usually sufficient to satisfy this requirement and is generally a matter to which a trier of fact must assign weight.[171] For example, in *United States v. Sinclair*,[172] the court admitted the testimony of a drug user despite allegations of a "clouded memory," relying on its confidence in the jury's traditional role of determining witness credibility.[173]

There is, however, an important line that a witness crosses with too many memory or perception gaps; eventually, the witness can only convey the testimony coherently by filling the gaps with hearsay or speculation.[174] Witnesses commonly attach caveats to the accuracy of their memory, such as "I believe," "to the best of my recollection," or "I cannot be positive, but I think."[175] The critical threshold, which the

[165] 622 A.2d 459 (R.I. 1993).

[166] *Id.* at 462.

[167] 992 F.2d 809 (8th Cir. 1993).

[168] *Id.* at 811.

[169] *Id.*

[170] 27 WRIGHT & MILLER, *supra* note 148, § 6023.

[171] Tippens v. Celotex Corp., 805 F.2d 949, 953–54 (11th Cir. 1986).

[172] 109 F.3d 1527 (10th Cir. 1997).

[173] *Id.* at 1537.

[174] *See* 2 WIGMORE, *supra* note 159, § 659 ("[T]he law may reject testimony which appears to be founded on data so scanty that the witness'[s] alleged inferences from them may at once be pronounced . . . extreme.").

[175] *See* Mason Ladd, *Expert and Other Opinion Testimony*, 40 MINN. L. REV. 437, 437, 440 (1956) (examining the evolution of Minnesota's opinion standards for both lay and expert witnesses against the Uniform Code of Evidence prior to the adoption of the Federal Rules of Evidence).

trial judge wields tremendous latitude in determining, is where the witness can only convey the narrative of her testimony by filling relevant gaps with speculation.[176] At this point, it is proper for a judge to exclude the testimony as speculative.[177] The speculation threshold is similar for the recollection and perception components of personal knowledge. The witness in *McCrary-El* could not convey a complete narrative without speculation because he could not perceive key elements of the story due to his lack of vantage point from which to observe the relevant events;[178] the *Sinclair* witness, on the other hand, could convey a complete story, even if the opposing party called his ability to recall into question, because he was able to perceive to the subject of his testimony.[179] The key element that distinguishes these cases is whether the ability to perceive or remember is essentially nonexistent or merely limited, distorted, or otherwise imperfect.

Rule 602's third element is comprehension. Even when a witness perceives an event through direct sensory perception, she must still comprehend what she sees to have personal knowledge to testify on the matter.[180] Again, a witness's comprehension does not need to be perfect. For example, a court may admit a child's testimony, even if she did not fully understand what was happening, so long as the other elements are met.[181] A witness's comprehension of her perceptions will never be without inference, as a natural degree of inference is always present in human comprehension.[182] To understand sensory perceptions, a person has no choice but to connect those perceptions to past experiences and draw inferences about what she perceives.[183] Ultimately, the judge controls the amount of latitude to grant to a

[176] 2 WIGMORE, *supra* note 159, § 659.

[177] *Id.*

[178] McCrary-El v. Shaw, 992 F.2d 809, 811 (8th Cir. 1993).

[179] United States v. Sinclair, 109 F.3d 1527, 1536–37 (10th Cir. 1997).

[180] 27 WRIGHT & MILLER, *supra* note 148, § 6023.

[181] *See* Sauer v. Exelon Generation Co., 280 F.R.D. 404, 405, 407 (N.D. Ill. 2012) (refusing to exclude deposition of a child because she had "difficulty . . . remembering, communicating and understanding").

[182] 27 WRIGHT & MILLER, *supra* note 148, § 6023.

[183] *Id.*; *see also* United States v. Joy, 192 F.3d 761, 767 (7th Cir. 1999) ("Because most knowledge is inferential, personal knowledge includes opinions and inferences grounded in observations or other first-hand experiences."). Humans must use some inferences to make sense of their world, or else testimony would consist only of a "description of the chemical and electrical effects of perception on the witness'[s] brain," which no witness is consciously aware of. 27 WRIGHT & MILLER, *supra* note 148, § 6023.

witness by either requiring more literal perceptions or allowing more inferences to describe the events that the witness perceived.[184]

The final element is the ability to testify based on the first three components. This is closely related to the third element of comprehension, but refers to the witness's comprehension at the time of testimony rather than at the time of perception.[185] For example, when a witness has been hypnotized to refresh her memory or has suffered a brain injury since the event at issue, she may no longer be able to comprehend the line of questioning or her perceptions of the event, even though she understood the event at the time she perceived it.[186] If she is not able to comprehend at the time of questioning, she cannot satisfy the personal knowledge requirement.[187]

The personal knowledge standard from Rule 602 direct testimony helps illustrate the knowledge required to meet the knowledge standard of Rule 901(b)(1). Thus, a witness must meet Rule 602's personal knowledge elements to testify as to whether photographic evidence is a fair and accurate portrayal.[188] To have the requisite knowledge, the witness must base her fair and accurate portrayal judgment on the direct use of her own senses, must have comprehended what she perceived at the time as well as at the time of her testimony, and must have a recollection of that prior perception. The witness is, of course, entitled to an imperfect memory as well as limitations in perception.[189]

III. AUTHENTICATING WITNESSES CAN NO LONGER RELIABLY TESTIFY TO THE FAIR AND ACCURATE PORTRAYAL STANDARD TO AUTHENTICATE PHOTOGRAPHIC EVIDENCE

Over the past twenty-five years, several scholars have noted the risk that evidentiary standards are too low to address advances in digital photography,[190] but they have made little progress in motivating any

[184] *See* Visser v. Packer Eng'g Assocs., 924 F.2d 655, 659 (7th Cir. 1991) (stating that personal knowledge includes inferences and some opinions because "all knowledge is inferential"). The balance between pure sensory perception and the inferences required to comprehend what a person perceives can reach esoteric levels beyond the intent of both the personal knowledge standard and this Comment.

[185] 27 WRIGHT & MILLER, *supra* note 148, § 6023.

[186] *Id.*

[187] *Id.*

[188] FED. R. EVID. 602, 901.

[189] *See supra* notes 163–70 and accompanying text.

[190] *See* Witkowski, *supra* note 80, at 285–87 (arguing in 2002 that the standard to admit digital images was insufficient); *see also* Sharon Panian, *Truth, Lies, and Videotape:*

changes to the standards.[191] Two factors have historically mitigated the impact of such a low bar: first, the court could rely on expert witnesses to assist with authenticity determinations, and second, it was still extremely difficult to create high quality fake video. The dawn of the deepfakes era brings this deficiency to the forefront with a new sense of urgency.[192] The proliferation of deepfakes technology renders obsolete the assumptions upon which the fair and accurate portrayal test relies; witnesses can no longer meet the fair and accurate portrayal standard within the legal standard of personal knowledge required to authenticate video evidence.

The unworkability of the fair and accurate portrayal standard is born out of a convergence of several factors. Deepfakes vastly increase the likelihood that authenticating witnesses will be unable to identify material changes from the actual scene that the video depicts.[193] Moreover, fake video is more likely to corrupt an authenticating witness's memories to lead her to actually recall the falsehoods that the video depicts.[194] The authenticating witness's inability to detect alterations from what she observed and the possibility of false memories leads to a complete inability for the witness to attest to a video as a fair and accurate depiction. The only way to attest that a video is a fair and accurate portrayal is by speculating on vast amounts of detail which, critically, witnesses are likely to believe as their own memory when the court shows them a fake video.[195] When combined with the disconnect inherent in the pictorial communication theory,[196] the result is a high probability of the court presenting to a jury fraudulent substantive evidence that has been authenticated by a witness without proper personal knowledge.

Are Current Federal Rules of Evidence Adequate?, 21 SW. U. L. REV. 1199, 1205–14 (1992) (highlighting common distortion problems with misleading computer graphics and edited video tapes).

[191] *See, e.g.*, Owens v. State, 214 S.W.3d 849, 854 (Ark. 2005) (refusing to alter the standard for digital photographs).

[192] *See supra* Part I (explaining the believability of human likenesses deepfakes and the unique detection challenges that make expert witness authentication more challenging for deepfakes than for other means of fraud).

[193] *See supra* Part I.A (examining the alarming level of precision and realism that programmers using generative adversarial networks can create).

[194] *See supra* Part I.C (exploring the tremendous effect that video has on human recollection by creating suggestive false memories).

[195] *See* Wade et al., *supra* note 67, at 899.

[196] *See infra* Part III.A (arguing that illustrative evidence often conveys substantive effects beyond the scope of the pictorial communication theory).

A. Muddled Theories: Video Causes Pictorial Communication Evidence to Leech into Substantive Evidence

The standard for admitting photographic evidence without an accompanying witness is far more comprehensive than when a witness is available to testify that the visual is a fair and accurate depiction.[197] However, the natural result of society's familiarization with and trust in photography and video recordings is that illustrative evidence's impact perpetually bleeds over into substantive effect; scholars have articulated this concern for some time, yet the problem remains.[198] Under the pictorial communication theory, photographic evidence should, strictly speaking, "illustrate[] the witness'[s] testimony, . . . add[ing] nothing further."[199] But this belies the natural human experience of consuming photographic evidence—such evidence conveys more information to the trier of fact than the witness could possibly have seen or heard but also may not have picked up every detail that the witness actually perceived. This dilemma is both technical, in the sense that a photograph is "not a replication but a representation, a constructed—and hence fallible—image,"[200] and experiential, in that the witness could not possibly recollect every single detail a recording conveys and simultaneously may very well recall information that the recording device did not capture. Courts have acknowledged the risk inherent in "[t]he masking of the substantive effect of photographs under the rubric of 'illustrative evidence'" as lacking "conceptual honesty."[201] The resulting effect is

[197] *See supra* Part II.B.1–2 (comparing the legal standard for the pictorial communication theory and silent witness theory).

[198] 2 BROUN ET AL., *supra* note 84, § 215; *see also* Robert D. Brain & Daniel J. Broderick, *The Derivative Relevance of Demonstrative Evidence: Charting Its Proper Evidentiary Status*, 25 U.C. DAVIS L. REV. 957, 998, 1018 (1992) (examining the evolving effect of demonstrative evidence and proposing modifications to Rule 401 relevance standards).

[199] Jessica M. Silbey, *Judges as Film Critics: New Approaches to Filmic Evidence*, 37 U. MICH. J.L. REFORM 493, 500–03 (2004) (highlighting the "jurisprudential anxieties" inherent in mischaracterizing demonstrative evidence).

[200] *See* Jennifer L. Mnookin, *The Image of Truth: Photographic Evidence and the Power of Analogy*, 10 YALE J.L. & HUMAN. 1, 7, 23 (1998) (examining the "kaleidoscopic understandings of the meaning of photographic evidence" and judicial attempts to govern evolving visual technology).

[201] 2 BROUN ET AL., *supra* note 84, § 215; *see also* JOSEPH, *supra* note 89, § 5.02[1][c] (analyzing the terminology separating the pictorial communication and silent witness theory as "unfortunate" because photographic evidence "introduced by means of the

that photography admitted under the low standard of the pictorial communication theory can easily have the practical effect of substantive evidence as if admitted under the silent witness theory but without meeting the more stringent requirements of Rule 901(b)(9).[202] This occurrence is rooted in the judicial system's confidence in the reliability of the photographic process, despite the fact that film theory teaches camera operators how to deliberately invoke reactions through a host of techniques.[203]

Although courts acknowledge an underlying risk to digital photography and video, the fair and accurate portrayal standard has nonetheless been seemingly immune to reconsideration. The Supreme Court of Arkansas, for example, has recognized the risk that it is easier to manipulate digital, rather than traditional, images yet it refused to impose a higher burden of proof for their admissibility when a defendant challenged the admission of surveillance video under the fair and accurate portrayal standard.[204] The lack of evolution of the admissibility standard is in part because challenges to the veracity of digital images are rare.[205] This is likely attributable to the legal community's lack of awareness of the risk inherent in digital images compared to older technology.[206] Additionally, when a party does challenge a digital image, the challenge typically addresses an overt enhancement of the image rather than the image's authenticity.[207] Moreover, courts may fear that elevating admission standards for photographic evidence will stifle the efforts of law enforcement, whose use of digital equipment during crime scene investigation has become commonplace,[208] or will slow the trend towards the convenience that comes with increased computer use in litigation.[209]

fair-and-accurate standard need not be given merely illustrative effect but may be, and often are, entitled to be given substantive effect").

[202] Silbey, *supra* note 199, at 531. "[C]ourts in their rulings admitting or excluding filmic evidence frequently evaluate film as a demonstrative aid only to later marshal the film toward substantive ends." *Id.*

[203] *See id.* at 531–32 (noting the judicial system's confidence in the transparency of film despite "a century of film theory and history teaching the opposite"); *see also id.* at 548–49 (finding that camera operators make "each spectator feel as if he or she is an eyewitness, despite that impossibility").

[204] Owens v. State, 214 S.W.3d 849, 854 (Ark. 2005).

[205] Witkowski, *supra* note 80, at 285.

[206] *Id.* at 286.

[207] *Id.*

[208] *Id.* at 286–87, 287 n.87.

[209] *Id.* at 287.

The consequences of the natural bleed over from illustrative to substantive evidence is that digital photography and video, admitted under the easily satisfied standard of Rule 901(b)(1), tend to convey substantive fact far beyond what the legal standard assumes or intends.

B. Witnesses Can No Longer Meet the Personal Knowledge Standard of Rule 602 to Attest to Photographic Evidence as a Fair and Accurate Depiction of a Scene

Establishing a foundation for admitting photographic evidence under the pictorial communication theory requires witness-with-knowledge testimony that the photograph or video is a fair and accurate depiction of the scene that it illustrates; to attest to this standard, a witness must be able to satisfy Rule 602's personal knowledge requirement.[210] Because witnesses are unable to perceive alterations or fabrications in deepfake videos, they can no longer determine whether the video's depiction is a fair and accurate portrayal of their memory. Using the personal knowledge standard articulated in Rule 602 caselaw, witnesses will commonly fail the recollection element of personal knowledge that a video is a fair and accurate portrayal.[211] The personal knowledge standard allows for significant gaps in the ability to recollect, but it does not permit gaps so central to the testimony that the testimony crosses the threshold into speculation.[212] Because witnesses cannot possibly recall all of the detail conveyed in a photograph or video, their limitations are likely to go beyond fuzziness or uncertainty and become speculative.

The underlying problems with the fair and accurate standard did not emerge with the invention of deepfakes; these problems have existed ever since photoshop became a commonly used verb.[213] Rather, deepfakes critically reduce the already limited effectiveness of authentication witnesses. Deepfakes exacerbate the inability of witnesses to determine their own recollection limitations and communicate the extent to which their limitations affect their ability to attest to the fair and

[210] *See supra* Parts II.B.1, II.C.

[211] *See supra* text accompanying notes 172–73; *supra* notes 67, 71, 192 and accompanying text.

[212] *See supra* notes 174–79 and accompanying text.

[213] Concerns over exaggerations or distortions in digital photography have been articulated in response to the emergence of photoshop. Witkowski, *supra* note 80, at 283–87. The potential for manipulation skyrockets with the ability to realistically create human likenesses in videos, not just still photography. *See supra* Part I.A (addressing the fidelity of human likenesses in deepfakes).

accurate portrayal standard. Deepfakes' lifelike fidelity reduces the likelihood that authentication witnesses will reliably rise to the task of stating either that something looks different from the way they remember it or that they do not recall it at all; the visuals are too convincing and too likely to take advantage of the suggestibility flaw inherent in our memories.[214] Thus, the speculation that occurs in blanketing the entire depiction as fair and accurate crosses the threshold of acceptable gap filling.[215]

Even a well-intentioned witness with no intention of deceiving the court will be unable to meet the threshold. The following example is illustrative. A criminal defendant offers a video made using a commercial iPhone. It depicts the defendant at an event with a date and location known to the public, such as a concert or other public event, thus providing an alibi. A witness who was at the event may recognize a variety of features that are true: the concert stage, the events transpiring in the background, or other individuals present. But the witness will not be able to discern small changes that are undetectable to her, such as the insertion of the defendant's likeness onto another individual who was actually present at the event. The proponent of the evidence cannot ask whether the witness recalls every detail in the video—the amount of detail makes the task inconceivable for both the proponent and witness. Instead, the proponent asks the witness to testify whether the picture is a fair and accurate portrayal of the scene that she remembers. Following the witness's fair and accurate portrayal testimony, the jury will see evidence with small but significant alterations.[216]

The blending of pictorial communication and silent witness theories sheds light on why a witness in this context can no longer meet the recollection element of personal knowledge.[217] The witness here likely has a variety of memories from the event depicted. She may remember which speaker or entertainer the event featured, some details on how the event was laid out, or what the stage looked like. By recalling any of these factors, she likely feels comfortable attesting to the video as a fair and accurate portrayal of the scene. If asked to testify whether she

[214] *See* Bennett, *supra* note 61, at 1335–37 (describing how human memory is imperfect and susceptible to suggestions); Wade et al., *supra* note 67, at 899 (using fabricated video in a psychological study to demonstrate witness suggestibility).

[215] *See supra* notes 174–79 and accompanying text.

[216] Witkowski, *supra* note 80, at 282 n.65.

[217] *See supra* Part III.A.

remembers these specifics, she certainly passes the personal knowledge standard for any of them, even if she expresses some uncertainty.[218] However, if asked specifically whether she saw the defendant at the event, the witness may have no recollection. Nonetheless, the witness testifies that the entirety of the scene is a fair and accurate depiction of her memory. Despite the proponent offering the video under the pictorial communication theory, the jury sees all of the surrounding details encompassed by the video, whether the witness recalled them or not. The witness cannot possibly recollect that volume of detail if the court (unrealistically) examined her recollection of each and every individual detail of the video. The only way for the witness to testify that the video is a fair and accurate portrayal is speculation because of the likelihood that the witness cannot detect whether changes have been made. Of course, she may specifically state that she remembers a manipulated part of the video and identify it, but in doing so, she has authenticated substantive facts that she did not actually remember and likely has no reason to suspect that there were any limitations on her fair and accurate assessment.[219]

Witnesses' inability to perceive changes in fake video is twofold: not only are witnesses unlikely to be able to perceive changes, but they are also willing to affirmatively remember portrayals in video that were altered and did not actually take place.[220] Critically, a witness's inability to perceive changes in the depiction does not reflect in her understanding of her own perceptions. The psychological suggestibility that fake video has on memory[221] warps the reliability of an authenticating witness. Professor Kimberly Wade's psychological study is a convincing demonstration that photographs and video are powerful tools to refresh a witness's memory, even when the memory that the imagery invokes never happened.[222] The suggestibility problem inherent in fake video and fake narratives vastly increases the likelihood that a witness believes that she has the personal knowledge to authenticate a video, even absent any intended deception by the witness.[223]

[218] *See* Ladd, *supra* note 175 and accompanying text (noting that statements like "I think" do not render witness testimony excludable).

[219] *See* Wade et al., *supra* note 67, at 904–06 (illustrating how few witnesses suspect that video evidence may be doctored).

[220] *Id.*

[221] *See supra* Part I.C.

[222] *See* Wade et al., *supra* note 67, at 904–06.

[223] Bennett, *supra* note 61, at 1357–58. One study showed that human memory is so susceptible to the effect of suggestibility that even using "the" instead of "a" as a

Suggestibility along with the precision of deepfakes pose both technological and psychological restraints to a witness's determination of fair and accurate portrayal. Such testimony does not merely represent a limitation on the witness's recollection capability when attesting to the authenticity of a video—it represents a complete inability to make the determination of her personal knowledge, which pushes a witness's personal knowledge past the level of uncertainty normally allowed to establish a foundation. The previous example of the alibi video is distinct from examples in which witnesses were unsure of their perceptions, had limited sensory perceptions available, or had incomplete information.[224] In each of these scenarios, there was some ability for the witness to recognize and articulate the limitations of her personal knowledge, whether in perception or recollection.[225] Here, however, a witness can only label the entire video or scene as a fair and accurate depiction by using those facts that she does recall from the scene and augmenting them with speculation. This is especially dangerous when combined with the psychological effect of suggestibility that is especially strong with video—the witness is likely to convey inherently speculative fair and accurate depiction testimony confidently and without doubts as to her recollection capability.[226]

This analysis does not characterize the evidence's probative value itself to be speculative—the video, if authenticated, may be highly probative or speculative in its own right depending on what it depicts and what its proponent intends to demonstrate to the jury.[227] Here, the witness's fair and accurate portrayal testimony, not the evidence, becomes speculative—that is, it has little probative value on establishing the foundation for authentication.

At first glance, this characterization of the witness's fair and accurate portrayal testimony seems to fly in the face of the strong tradition of a minimalist standard in which the proponent does not need to

definite article in a question dramatically affects the witness's likelihood of recalling seeing an object. *See id.* at 1357 n.144 (citing Elizabeth F. Loftus & Guido Zanni, *Eyewitness Testimony: The Influence of the Wording of a Question*, 5 BULL. PSYCHONOMIC SOC'Y 86, 87–88 (1975) (showing a significant increase in the percentage of affirmative test subject responses to the questions "Did you see *the* [object]?" and "Did you see *a[n]* [object]?")).

[224] *See supra* notes 165–79 and accompanying text.

[225] *See supra* notes 165–79 and accompanying text.

[226] *See supra* notes 67–71 and accompanying text.

[227] *See* FED. R. EVID. 403 (requiring the court to balance relevancy against the risk of undue prejudice).

eliminate "all possibilities inconsistent with authenticity, or to prove beyond any doubt that the evidence is what it purports to be."[228] However, there is an important distinction from the long list of examples of courts admitting testimony based on shaky memories and imperfect observations[229]: in each of these examples, the witness can qualify her imperfect memories by articulating the degree of limitation or imperfection. She can communicate how "positive" or not she is, or how well she was able to perceive the facts by explaining, for example, how dark it was, how far she could see, or whether she could make out facial features. She could also describe her vantage point and identify physical or environmental limitations. Ultimately, these examples all provide a minimal articulable basis for recollection and perception as a foundation. The deepfakes problem, compounding pre-existing issues with digital photography, creates an authenticating witness who cannot articulate her level of confidence or capability when it comes to labeling an entire video sequence as fair and accurate; the potential forgeries are too high quality,[230] and psychological factors create a sense of certainty that does not reflect the true degree of speculation.[231] The result is that the only way for a witness to testify that a video sequence is a fair and accurate depiction is by augmenting her memory with speculation, even if she do not realize she is doing it.[232]

C. *Digital Photographic Evidence Warrants a More Stringent Means of Authentication*

Because witnesses will no longer be able to meet the legacy standard of Rule 901(b)(1)'s knowledgeable witness by attesting that a video is a fair and accurate portrayal, courts need to look elsewhere for a sufficient finding that photographic evidence is what its proponent claims it is. This new standard does not necessarily replace Rule 901(b)(1), which is still applicable for a variety of other forms of evidence.[233] Instead, a proposed new section would specifically govern

[228] United States v. Gagliardi, 506 F.3d 140, 151 (2d Cir. 2007) (quoting United States v. Pluta, 176 F.3d 43, 49 (2d Cir. 1999)); *see also supra* notes 124–26 and accompanying text.

[229] *See supra* notes 165–73 and accompanying text.

[230] *See supra* Part I.A.

[231] *See supra* Part I.C.

[232] *See* Wade et al., *supra* note 67, at 899–900.

[233] For example, documents that are not self-authenticating may still fall under the knowledgeable witness standard of Rule 901(b)(1) or other means of Rule 901(b).

the unique challenges that digital photography in the modern age present:

> (Proposed New) Rule 901(b)(11): Before a court admits photographic evidence under this rule, a party may request a hearing requiring the proponent to corroborate the source of information by additional sources.

As mentioned earlier, the processes offered in Rule 901(b) to establish a foundation for authentication are not exhaustive; Rule 901(b)(1) or 901(b)(9) are not the exclusive options.[234] A proponent may also use circumstantial evidence to establish a foundation for authentication without adhering to one of the processes enumerated in Rule 901(b).[235] This new rule essentially codifies an existing means of authentication and requires it for photographic evidence. Thus, even if the proponent cannot produce a witness with personal knowledge, methods of proving authenticity "can be infinite in variety, limited only by the circumstances pertaining in the particular case."[236]

A Rule 901(b)(11) hearing would consider authentication factors beyond the bare bones requirement of 901(b)(1). A starting point for elements for the court to consider at this stage is the presence of additional corroborating evidence, as the court would consider in instances where the proponent establishes its foundation outside of the traditional 901(b)(1) or 901(b)(9) paths.[237]

Returning to the example above, if the government called for a Rule 901(b)(11) hearing to challenge the alibi video that the defendant submitted, the court would require more than a knowledgeable witness to establish a foundation for authenticity. For example, if the proponent offered a ticket stub or other circumstantial evidence of the defendant's attendance, it would corroborate the authenticity of the video. The proposed rule would not rule out the utility of the witness through direct testimony. If a witness testified that she personally saw the defendant at the alibi event (a specific observation that a witness is far more likely to recollect concretely) as opposed to whether the

[234] RICE & KATRIEL, *supra* note 95, § 7.02[A][3][a].

[235] *Id.*

[236] § 5 JACK B. WEINSTEIN & MARGARET A. BERGER, WEINSTEIN'S FEDERAL EVIDENCE § 901.03 (Mark S. Brodin & Matthew Bender eds., 2d ed. 2020), LexisNexis.

[237] *See* Marie-Helen Maras & Alex Alexandrou, *Determining Authenticity of Video Evidence in the Age of Artificial Intelligence and in the Wake of Deepfake Videos*, 23 INT'L J. EVIDENCE & PROOF 255, 258–59 (2019) (recommending requiring corroborating evidence to authenticate video in an article addressing deepfakes in the context of probative value rather than personal knowledge).

entire scene is a fair and accurate portrayal, then the witness would easily meet the personal knowledge standard.

D. Increased Scrutiny Prior to Admission is Worth Risking Excluding Relevant Evidence Because of the Heightened Risk of Jury Prejudice Associated with Photographic Evidence

The alibi example may seem redundant; if there was a witness to corroborate a defendant's alibi, then why does the defendant need the video in the first place? The more troublesome instance is where the video is the only source of evidence concerning the alibi, whether because the videographer is unavailable for some reason or is a criminal defendant herself and unwilling to testify at trial.[238] The proposed rule likely poses a threat to the volume of digital media submitted in court. The immediate counterargument to address elevated foundational standards for authentication of photographic evidence is that a jury can consider these factors at trial just the same as the court can in a preliminary hearing. After all, nearly all forms of evidence, from written documents to oral assertions, are vulnerable to the potential for fraud; the system depends on a jury (with help, if necessary, from expert witnesses) to assign weight to evidence based on credibility and relevance.

But the heightened risk of forgery inherent in deepfakes warrants heightened admission standards. Photography, and to a greater extent, video, have a stronger effect than other forms of evidence; they cannot be so easily dismissed once seen.[239] While the emotional power of photographic, and especially video, evidence is generally thought of as an issue of probative value for courts to consider under Rule 403—such as when evidence is relevant but contains extremely graphic content that renders it unduly prejudicial—suggestibility is not the type of emotion that typically factors into the Rule 403 calculus.[240]

In the context of questionable or competing forms of evidence, juries have a tendency to cast aside other, less interesting forms of evidence when presented with the ease and convincing nature of

[238] *See* U.S. CONST. amend. V.

[239] Mnookin, *supra* note 200, at 2–3. "The photograph, in particular, has long been perceived to have a special power of persuasion, grounded both in the lifelike quality of its depictions and in its claim to mechanical objectivity. Seeing a photograph almost functions as a substitute for seeing the real thing." *Id.* at 1–2 (footnotes omitted); *see also* Wade et al., *supra* note 67, at 899.

[240] 2 BROUN ET AL., *supra* note 84, § 215.

viewing photographic evidence.[241] Juries are also remarkably poor at adhering to limiting instructions[242] or even admonishments to disregard inadmissible evidence.[243] In fact, they also "paradoxically pay *greater* attention to information ruled inadmissible than if the judge had not drawn attention to the admissibility of the information and simply allowed it into evidence."[244] Thus, even if a party casts doubt on the authenticity of photographic or video evidence, once the court admits it, the vivid images remain in a jury's mind. By virtue of video's emotional effect and the tendency to prioritize it above other forms of evidence, the risk of waiting for a jury to consider initial corroborating evidence concerning a video's authenticity justifies the court's consideration of these factors prior to admission.[245]

A preliminary hearing to consider circumstantial authentication factors does not solve the deepfakes evidentiary crisis—but it does mitigate it. The proposed standard for establishing a foundation would still be limited and does not render photographic evidence forgery-proof; a jury still ultimately determines credibility and weight of the evidence that is admitted. Because of the challenges in creating effective detection measures (and the especially worrisome challenge that such measures will improve the forgery process),[246] regulation and potential criminal solutions are in order to address deepfakes on a larger scale and stem their potential entry into the courtroom.[247] Until then, a preliminary hearing process would bolster the confidence in video evidence for a jury to consider, rather than allowing all photographic

[241] *See* Cynthia A.R. Woollacott, *Evaluating Video Evidence*, 14 L.A. LAW. 24, 25 (1991) (considering various issues balancing probative value against the risk of undue prejudice); *see also* Thomas v. C.G. Tate Constr. Co., 465 F. Supp. 566, 571 (D.S.C. 1979) (articulating the court's concern over video's "dominating effect [that] will distract the jury from its proper consideration of other issues they will be called on to decide" because of how the video will "stand out in the minds of the jury").

[242] *See, e.g.*, Panian, *supra* note 190, at 1215 (citing HARRY KALVEN, JR. & HANS ZEISEL, THE AMERICAN JURY, 417–27 (1971) (discussing the tendency of juries to stray from judicial instructions)).

[243] *See* 1 JOEL D. LIEBERMAN & DANIEL A. KRAUSS, JURY PSYCHOLOGY: SOCIAL ASPECTS OF TRIAL PROCESSES 75–89 (2009) (examining juries' difficulty with limiting instructions and analyzing the "backfire effect" of inadmissible evidence already seen by juries).

[244] *Id.* at 79.

[245] *See* Woollacott, *supra* note 241, at 25.

[246] *See supra* Part I.B.

[247] *See, e.g.*, Caldera, *supra* note 11, at 177–78; de Zayas, *supra* note 72; Harris, *supra* note 11, at 102–03; Spivak, *supra* note 11, at 340–41.

evidence to pass the foundational stage with a testimonial witness who lacks the requisite personal knowledge to attest to the evidence's validity.

CONCLUSION

The age of machine learning has contributed to human achievements and triumphs equaled only by the risk that it creates when placed in the wrong hands.[248] Unfortunately for our trusting eyes, this atom cannot be unsplit, and artificial intelligence-enabled video creation is likely here to stay. As regulators scramble to address the risks posed by fake video created through GAN techniques, the legal standard for authentication of video evidence has fallen behind; evidentiary standards need to evolve to accommodate our changing world. The result will likely be a reduction in the reliance on photographic evidence in court after nearly a century of the steady rise in confidence and reliance upon photographic evidence to capture moments lost to human memory.

[248] *See* Charles Towers-Clark, *AI Diagnosis Tool Bridges the Gap Between Doctors and Patients*, FORBES (Feb. 13, 2019, 12:03 PM), https://www.forbes.com/sites/charlestowersclark/2019/02/13/ai-diagnosis-tool-bridges-the-gap-between-doctors-and-patients [https://perma.cc/E9L2-UM8A]; James Vincent, *Twitter Taught Microsoft's AI Chatbot to Be a Racist Asshole in Less than a Day*, VERGE (Mar. 24, 2016, 6:43 AM), https://www.theverge.com/2016/3/24/11297050/tay-microsoft-chatbot-racist [https://perma.cc/XA23-Y9XL] (describing how Twitter users manipulated a publicly accessible artificial intelligence chatbot).