DESIGNED DECEIT: WHEN DEEPFAKES LIE, CAN WE DISCERN THE TRUTH?

JOHN J. HEALY JR.[†]

[†] IP Litigation Associate, Ropes & Gray; J.D., 2024, Maurice A. Deane School of Law at Hofstra University; B.S. Industrial & Systems Engineering, 2021, The Ohio State University. I am grateful for the overwhelming support from Professors Ellen Yaroshefsky, Fred Klein, Eric Freedman, Irina Manta, and G. Alex Sinha. I credit this Article's success to their encouragement and guidance. Many thanks to the brilliant minds of the *Wayne Law Review* for their masterful efforts in perfecting this Article. The arguments and opinions expressed herein are solely my own and do not reflect the views of my associated law firm or educational institutions.

I. INTRODUCTION

Interrogation proceedings are an integral component of criminal prosecutions.[1] When a suspect freely and voluntarily confesses, their confession is admissible in court.[2] Generally, a confession is voluntary if a suspect confesses without the police inducing duress, fear, or compulsion.[3] The police, however, may lie during interrogation proceedings.[4] Courts have held that a confession obtained by means of fraud, deception, or trickery is admissible—hence, a confession secured by lying to the suspect about the state of the evidence against him, or even by means of the police disguising themselves as another prisoner, is permissible.[5]

Traditional means of lying to a suspect that encourage an admission have certainly been effective.[6] However, new technologies have the potential to heavily tilt the scales of equity in the government's favor.[7] Lying or pretending to be an inmate is one thing, but what if the police present a fake video or audio recording of a co-defendant confessing to the crime?[8] What if the police present a fake audio recording of the suspect's child or wife admitting that the suspect committed the crime?[9] Would it

---

1. *See* Markus M. Thielgen et al., *Police Officers' Interrogation Expertise and Major Objectives in Police Service and Training: A Comprehensive Overview of the Literature*, 13 FRONTIERS PSYCH., June 2022, at 1–3, https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2022.823179/full [https://perma.cc/5M8V-CB9T] (highlighting how skilled interrogations are essential for gathering information that can significantly impact the outcome of criminal prosecutions); *see also* HIGH VALUE DETAINEE INTERR. GRP, FED. BUREAU OF INV., INTERROGATION: A REVIEW OF THE SCIENCE HIG REPORT 5, 14–23 (2016), https://www.fbi.gov/file-repository/hig-report-interrogation-a-review-of-the-science-september-2016.pdf/view [https://perma.cc/6BVH-KQG4] (discussing how effective interrogation methods are vital to securing key evidence in criminal prosecutions).

2. *See, e.g.*, 18 U.S.C. § 3501(a).

3. Colorado v. Connelly, 479 U.S. 157, 162 (1986) (citing Culombe v. Connecticut, 367 U.S. 568, 599 (1961)).

4. Frazier v. Cupp, 394 U.S. 731, 739 (1969) (finding that deceptive police conduct alone does not render a confession involuntary in "viewing a totality of the circumstances").

5. *See, e.g.*, Michael J. Zydney Mannheimer, *Fraudulently Induced Confessions*, 96 NOTRE DAME L. REV. 799, 808 (2020).

6. *See* Laurie Magid, *Deceptive Police Interrogation Practices: How Far is too Far?*, 99 MICH. L. REV. 1168, 1176, 1186–87 (2001).

7. Gavin Oxburgh & Laura Farrugia, *New Technology to Improve Police Interviews*, NORTHUMBRIA UNIV. (May 10, 2023, 10:21 AM), https://newsroom.northumbria.ac.uk/pressreleases/new-technology-to-improve-police-interviews-3251892 [https://perma.cc/7EWA-95DD].

8. *See infra* Part II.

9. *See infra* Part II.

matter if the suspect had no recollection of the event?[10] Confessions of this magnitude are the exception today, but this rare confession may soon become mainstream thanks to the power of deepfake technology.[11]

Deepfakes, or synthetic manipulations of audio or visual media, have begun to distort reality.[12] The public has continuously failed to separate artificial media from reality, presenting the opportunity for malicious actors to capitalize on unsuspecting victims.[13] This is especially true if the police use this technology against the most vulnerable in our criminal legal system.[14] Deepfake technology can distort reality to the point where people cannot decipher real life from artificial manipulation.[15] In fact, scammers have already used deepfakes to trick people or companies into sending them money,[16] impersonate top corporate executives,[17]

---

10. *See infra* Part II.

11. *See* Roop Reddy, *24 Deepfake Statistics – Current Trends, Growth, and Popularity (December 2023)*, CONTENTDETECTOR.AI (Dec. 13, 2023), https://contentdetector.ai/articles/deepfake-statistics [https://perma.cc/YMJ7-4TJE]; *see also* DEP'T OF HOMELAND SEC., INCREASING THREAT OF DEEPFAKE IDENTITIES 6 (2022), https://www.dhs.gov/sites/default/files/publications/increasing_threats_of_deepfake_identities_0.pdf [https://perma.cc/XJ6N-PLRJ] (explaining that the key difference between contemporary media and deepfakes is that casual viewers may "easily detect fraudulent" contemporary media, but deepfakes "may allow an adversary interested in sowing misinformation or disinformation to leveral far more realistic image, video, audio, and text content").

12. Michael Hameleers et al., *Distorting the Truth Versus Blatant Lies: The Effects of Different Degrees of Deception in Domestic and Foreign Political Deepfakes*, COMPUTS. HUM. BEHAV., Mar. 2024, at 1, 10–12, https://www.sciencedirect.com/science/article/pii/S0747563223004478 [https://perma.cc/2N95-NQG2].

13. Janna Anderson & Lee Rainie, *As AI Spreads, Experts Predict the Best and Worst Changes in Digital Life by 2035*, PEW RSCH. CTR. (June 21, 2023), https://www.pewresearch.org/internet/2023/06/21/themes-the-most-harmful-or-menacing-changes-in-digital-life-that-are-likely-by-2035 [https://perma.cc/DK52-FFYX].

14. *See id.* at 117 ("If these tools are rolled out too quickly, the potential to harm vulnerable populations is greater.").

15. Adam Satariano & Paul Mozur, *The People Onscreen Are Fake. The Disinformation Is Real.*, N.Y. TIMES (Feb. 7, 2023), https://www.nytimes.com/2023/02/07/technology/artificial-intelligence-training-deepfake.html [https://perma.cc/7A2U-MVS5].

16. Wendy Hughes et al., *Deepfake Scammers Steal $25 Million From Company: 5 Ways You Can Avoid Being Victim to Latest AI Nightmare*, FISHER PHILLIPS (Feb. 9, 2024), https://www.fisherphillips.com/en/news-insights/5-ways-avoid-being-victim-to-latest-ai-nightmare.html [https://perma.cc/P7RK-MZVE]; Michael Atleson, *Chatbots, Deepfakes, and Voice Clones: AI Deception for Sale*, FED. TRADE COMM'N (Mar. 20, 2023), https://www.ftc.gov/business-guidance/blog/2023/03/chatbots-deepfakes-voice-clones-ai-deception-sale [https://perma.cc/J6JJ-S89B] (describing sales using deception through deepfakes).

17. Catherine Stupp, *Fraudsters Used AI to Mimic CEO's Voice in Unusual Cybercrime Case*, WALL ST. J. PRO (Aug. 30, 2019, 12:52 PM), https://www.wsj.com/articles/fraudsters-use-ai-to-mimic-ceos-voice-in-unusual-cybercrime-case-11567157402 [https://perma.cc/B64G-PPKD].

impersonate former presidents to deceive the public,[18] manipulate scientific images,[19] bully students with lewd images of themselves,[20] and spread misinformation during campaign advertising.[21] Deepfake videos have increased exponentially since 2019,[22] yet despite this increase, researchers estimate that only 29% of the global population is aware of the existence of deepfakes.[23]

Psychologists have conducted several studies which highlight the effectiveness of altered media and variables that increase their effectiveness.[24] The MIT Detect Deepfakes project recently conducted a study to determine the rate of successful detection of political deepfakes.[25] Interestingly, the study found that increasing the number of fabricated media modalities (such as audio and video) increases a participant's accuracy in detecting deepfakes.[26] However, the study concluded by recognizing that human discernment relies more on *how* manipulated

---

18. Jon Bateman, *Get Ready for Deepfakes to be Used in Financial Scams*, CARNEGIE ENDOWMENT FOR INT'L PEACE (Aug. 10, 2020), https://carnegieendowment.org/2020/08/10/get-ready-for-deepfakes-to-be-used-in-financial-scams-pub-82469 [https://perma.cc/V6GW-ZKFS].

19. Liansheng Wang et al., *Deepfakes: A New Threat to Image Fabrication in Scientific Publications?*, PATTERNS (May 13, 2022) at 1–3, https://www.sciencedirect.com/science/article/pii/S2666389922001015 [https://perma.cc/28Y6-TCT9] (manipulating scientific images); M. M. El-Gayar et al., *A Novel Approach for Detecting Deep Fake Videos Using Graph Neural Network*, J. BIG DATA (Feb. 1, 2024) at 21–25, https://journalofbigdata.springeropen.com/articles/10.1186/s40537-024-00884-y [https://perma.cc/7CAH-CVQV] (detailing an approach for detecting images).

20. Zoe Thomas, *AI Used to Spread Fake Nudes of Students at a New Jersey High School*, WALL ST. J.: PODCASTS (Nov. 7, 2023, 3:01 AM), https://www.wsj.com/podcasts/tech-news-briefing/ai-used-to-spread-fake-nudes-of-students-at-a-new-jersey-high-school/6bc2a269-9fc4-4e75-ab24-cac66674f687 [https://perma.cc/TPE3-N58S].

21. Robert McMillan et al., *New Era of AI Deepfakes Complicates 2024 Elections*, WALL ST. J. (Feb. 15, 2024, 12:31 PM), https://www.wsj.com/tech/ai/new-era-of-ai-deepfakes-complicates-2024-elections-aa529b9e [https://perma.cc/9KBS-CARN]; Cristian Vaccari & Andrew Chadwick, *Deepfakes and Disinformation: Exploring the Impact of Synthetic Political Video on Deception, Uncertainty, and Trust in News*, SOC. MEDIA & SOC'Y (Feb. 19, 2020) at 1–2, 9, https://journals.sagepub.com/doi/full/10.1177/2056305120903408 [https://perma.cc/R9YC-H7V4].

22. Reddy, *supra* note 11.

23. *Id.* (referencing a survey conducted by iProov in 2022, which found that 29% of respondents knew what a deepfake was).

24. *See infra* Part II.B.2.a–b.

25. Matthew Groh et al., *Deepfake Detection by Human Crowds, Machines, and Machine-informed Crowds*, PNAS, 2022, at 8; Matthew Groh et al., *Human Detection of Political Speech Deepfakes Across Transcripts, Audio, and Video*, NATURE COMMC'N (Sept. 2, 2024) at 1–3, https://www.nature.com/articles/s41467-024-51998-z [https://perma.cc/ZLD5-B5LQ] [hereinafter Groh Political Speech].

26. Groh Political Speech, *supra* note 25, at 7.

media modalities portray something, rather than *what* the content is.[27] Additionally, a study conducted by scholars from the Center for Humans and Machines and the University of Amsterdam found that participants correctly spotted a deepfake 77.6% of the time on one video, but only 46.7% of the time on a different video, demonstrating a stark contrast between different manipulated videos from the same source.[28] The study also recognized the participant's tendency to presume that a video is authentic.[29] Scholars from the RAND Corporation, Pardee RAND Graduate School, Carnegie Mellon University, and the Challenger Center conducted a study that found between 27–50% of participants could not distinguish authentic videos from deepfake videos.[30] Hence, almost half the participants could not decipher reality from an artificial creation.[31] Lastly, a study conducted by the University College London found that 73% of people were able to detect audio deepfakes correctly, but subsequent training on how to spot deepfakes had no impact on a participant's accuracy.[32] The study hypothesized that as deepfakes improve, the detection rate will suffer, and no current method of training could ameliorate the advancements in this field.[33]

In all, these studies demonstrate the current gaps in deepfake technology that will soon be rectified to enhance a deepfake's effectiveness. Indeed, deepfake technology is so powerful that it has the potential to shift a once voluntary confession into an unconstitutionally elicited confession.[34] This Article argues that the presence of deepfake

---

27. *Id.* at 12.

28. Nils C. Köbis et al., *Fooled Twice: People Cannot Detect Deepfakes but Think They Can*, ISCIENCE (Nov. 19, 2021) at 2, 11, https://www.sciencedirect.com/science/article/pii/S2589004221013353 [https://perma.cc/JBA9-CY4K].

29. *Id.* at 8 ("[P]articipants are very conservative when reporting that a video is a deepfake, i.e., people have a tendency toward guessing authentic").

30. Christopher Doss et al., *Deepfakes and Scientific Knowledge Dissemination*, SCI. REPS (Aug. 18, 2023) at 3, https://www.nature.com/articles/s41598-023-39944-3 [https://perma.cc/AF6D-ETHF].

31. *Id.* at 2 ("[B]etween 27 percent to over half of survey responses were unable to correctly identify the authenticity of videos *regardless* of whether the video was authentic or a deepfake.") (emphasis in original).

32. Kimberly T. Mai et al., *Warning: Humans Cannot Reliably Detect Speech Deepfakes*, PLOS ONE (Aug. 2, 2023) at 8, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10395974/pdf/pone.0285333.pdf [https://perma.cc/X7G8-RFGJ].

33. *See id.* at 16 ("Humans can detect speech deepfakes, but not consistently. They tend to rely on naturalness to identify deepfakes regardless of language. As speech synthesis algorithms improve and become more natural, it will become more difficult for humans to catch speech deepfakes.").

34. *See* Hameleers, *supra* note 12, at 10 ("The finding also supports the truth-default-theory. . . . As long as deceptive information is similar to reality and the familiar viewpoints of a political actor, deception may not be triggered.").

technology alone should render a confession unconstitutional because psychological studies demonstrate a high risk of false confessions when presented with fake or misleading evidence.[35] While the psychological studies available do not concern deepfakes specifically, the studies referenced in this Article concern what deepfakes hope to become: a fake or misleading demonstrative that is seemingly real to the naked eye.[36]

Part II provides a background on police interrogation techniques, deepfake technology, and the cross-over between the two.[37] Part III discusses the constitutionality of police lying during interrogation proceedings and when a police interrogation violates the Constitution.[38] Part IV introduces a two-part solution for addressing deepfake technology and other future technologies.[39] Lastly, Part V concludes by calling for a categorical ban on deepfake technology during interrogation proceedings.[40]

II. BACKGROUND

This Article begins with three hypotheticals to illustrate the issues presented with police interrogations and deepfake technology. The first hypothetical is as follows:

Prosecutors accuse two co-conspirators of murder. The police videotape the interrogation of the first conspirator through a camera in the back corner of the room. Before interrogating the second conspirator, the police manipulate the audio of the interrogation so it appears that the first conspirator proclaims the second conspirator "planned the entire operation and actually pulled the trigger." The police likewise alter the video to match the now-fabricated audio—hence, the lips of the first conspirator move in sync with the fabricated audio. Now, the police present an entirely new interrogation to the second conspirator. When the police hit play on

35. *Compare* Frazier v. Cupp, 394 U.S. 731, 739 (1969) (holding that misrepresentation alone is insufficient to render a confession inadmissible), *and* Miranda v. Arizona, 384 U.S. 436, 455–56 (1966) (detailing the constitutional concerns with false confessions based on coercive interrogation proceedings), *with* Hameleers, *supra* note 12, at 3, 12 (describing that misinformation created through deepfakes impacts the beliefs held by individuals).

36. *What the Heck is a Deepfake?*, UNIV. OF VA., https://security.virginia.edu/deepfakes [https://perma.cc/KRH6-J799] (last visited Apr. 16, 2024).

37. *See infra* Part II.A–B.

38. *See infra* Part III.

39. *See infra* Part IV.

40. *See infra* Part V.

the video, the second conspirator witnesses a seemingly real video of the first conspirator betraying him.[41]

The second hypothetical is as follows:

The police interrogate a suspect accused of a domestic assault crime. The police introduce an audio recording from a hidden microphone on the officer's body. The audio recording is the suspect's son, who is crying and screaming that "daddy kept hitting mommy." Unbeknownst to the suspect, the police manipulated this audio recording using deepfake technology. The original audio recording was the suspect's son saying how much "daddy loved mommy."[42]

Lastly, the third hypothetical is as follows:

The police arrest a suspect for murder. The suspect claims that he passed out before the victim died due to inebriation on the night in question. The police introduce video evidence of an altercation at a bar that both the suspect and the victim were at on the night of the incident. The video demonstrates the suspect breaking a glass bottle on the bar top and slicing the victim's throat before running away. This incident never happened though—the police created it by deepfake technology.[43]

---

41. *See, e.g.*, Frazier v. Cupp, 394 U.S. 731, 737–39 (1969). This hypothetical was inspired by *Frazier*, which stands for the general notion that misrepresenting what a co-defendant has said is admissible. *Id.* at 739.

42. *See, e.g.*, Lynumn v. Illinois, 372 U.S. 528, 534 (1963); United States v. Tingle, 658 F.2d 1332, 1336 (9th Cir. 1981). This hypothetical was inspired by *Lynumn* and *Tingle* because while the hypothetical does not involve any threats, the inclusion of family matters seemed to strike a nerve in the Court's voluntariness inquiry. *Lynumn*, 372 U.S. at 534; *see also* United States v. Syslo, 303 F.3d 860, 867 (8th Cir. 2002) (finding that the district court did not err because it "reasonably concluded that the children's presence at the station did not begin to exert any coercive influence on [defendant] until she realized that they would not be able to be picked up by a relative.").

43. *See, e.g.*, Tankleff v. Senkowski, 135 F.3d 235, 240 (2d Cir. 1998). This hypothetical was inspired by the story of Martin Tankleff, who at seventeen years old was arrested for the murder of his parents. *Id.* Issues with *Miranda* aside, the pertinent facts of *Tankleff* for this Article are as follows: Tankleff woke up to find his mother dead and his father unconscious. *Id.* Tankleff dialed 9-1-1 and proclaimed that someone had killed his parents; he placed the blame on his father's business partner who had owed him substantial amounts of money. *Id.* Homicide detectives interviewed Tankleff and recognized inconsistencies in his accounts of the events of that morning. *Id.* This prompted the police to take Tankleff to headquarters for further questioning. *Id.* At this point, the police considered Tankleff a suspect. *Id.* Tankleff was interviewed for two hours when police questioning became "increasingly hostile." *Id.* During the interrogation, one officer left the room because he "receiv[ed] a telephone call." *Id.* at 241. The officer spoke loud enough just outside the interrogation room for Tankleff to overhear him. *Id.* The officer returned to the interview room and stated that he spoke with a detective overseeing Tankleff's father at the hospital. *Id.* The officer stated that "the doctors had pumped [Tankleff's father] full of adrenaline, that he had come out of the coma, and that he had accused his son" of the murder. *Id.* However, this story was not true—Tankleff's father remained in a coma until his death a few weeks later. *Id.* He never woke up or accused his son of a crime. *Id.*

Assume the police properly administrated *Miranda* warnings and otherwise complied with the law.[44] Additionally, assume all three hypotheticals result in the suspect confessing to the crime. Irrespective of whether the confession was false—meaning the suspect did not actually commit the crime—are these confessions voluntarily given under Fifth and Fourteenth Amendment jurisprudence?[45] The answer to this question rests on how a prospective court would define deepfake technology—are deepfakes just a better lie, or a psychologically coercive tool?[46]

This Article argues that deepfakes are undoubtedly a psychologically coercive tool when compared to traditional means of lying. Using the first hypothetical as an example, the traditional method of lying to a conspirator is *telling* the conspirator what his co-conspirator said—an out of "investigation room" statement offered for the truth. This statement is akin to hearsay, the weakest form of evidence when compared to demonstrative evidence, such as a video or audio recording.[47] A demonstrative lie would

---

Nevertheless, this fake phone call, followed by more questions, led Tankleff to challenge his recollection: Tankleff said, "[c]ould I have blacked out and done it?" *Id.* The police encouraged Tankleff to further describe his thoughts, where he asked if he had been "possessed" and that "[i]t's coming to me" before telling the officers how he had committed the acts. *Id.* Tankleff's confession was never suppressed, the New York courts upheld his conviction, and the Second Circuit denied his writ for habeas corpus, finding the confession was given voluntarily. *Id.* at 241, 246. Tankleff was exonerated in 2008 after evidence surfaced of the true perpetrators. Maurice Possley, *Exoneration of Martin Tankleff*, NAT'L REGISTRY OF EXONERATIONS, https://www.law.umich.edu/special/exoneration/Pages/casedetail.aspx?caseid=3675 [https://perma.cc/W9NP-WRE2] (last updated Sept. 13, 2022). Tankleff was imprisoned for eighteen years based off this false confession. *Id.*

44. *See* Miranda v. Arizona, 384 U.S. 436, 444 (1966).

45. *Id.*

46. Haynes v. Washington, 373 U.S. 503, 515 (1963) (discussing the issue of psychologically coerced confessions under the Due Process Clause). The court explained:

> The line between proper and permissible police conduct and techniques and methods offensive to due process is, at best, a difficult one to draw, particularly in cases such as this where it is necessary to make fine judgments as to the effect of psychologically coercive pressures and inducements on the mind and will of an accused. But we cannot escape the demands of judging or of making the difficult appraisals inherent in determining whether constitutional rights have been violated. We are here impelled to the conclusion, from all of the facts presented, that the bounds of due process have been exceeded.

Id.; *see also* Jimmie E. Tinsley, *Involuntary Confession: Psychological Coercion*, 22 AM. PROOF OF FACTS JUR. 2D §§ 10–26 (1980).

47. George F. James, *Role of Hearsay in a Rational Scheme of Evidence*, 34 ILL. L. REV. 788, 791 (1940); Ervin A. Gonzalez & Kyle B. Teal, *No Ideas but in Things: A Practitioner's Look at Demonstrative Evidence*, FLA. BAR J. (Dec. 2015), https://www.floridabar.org/the-florida-bar-journal/no-ideas-but-in-things-a-practitioners-look-at-demonstrative-evidence [https://perma.cc/Y6F4-G4ZP]; Mary Quinn Cooper, *The Use of Demonstrative Exhibits at Trial*, 34 TULSA L. REV. 567, 568 (1999) ("One advantage

trump a comparatively powerless hearsay lie, especially if the demonstrative lie looks and sounds like reality. In the second hypothetical, the police may seldom convince a suspect by relaying what the suspect's child said or did; however, if the suspect hears their child's voice (or a very realistic clone of their child's voice), they might react differently.[48] Even in the third hypothetical based in part on Martin Tankleff, imagine if Tankleff had *heard* or *seen* his father pleading that Tankleff had killed him—if the police's lie alone led Tankleff to question his memory and confess, certainly a demonstration of the lie could broker more.[49]

With this foundation laid, Subpart A will explore the existing tools employed, or actively forbidden, by police departments during interrogations.[50] Subpart B briefly explains the underlying technology of deepfakes to better illustrate how the above hypotheticals are plausible.[51] Subpart C highlights how police may use deepfakes, including how they might collect the mass amount of data required to create a deepfake.[52]

A. Police Interrogations

"[O]ur accusatory system of criminal justice demands that the government seeking to punish an individual produce the evidence against him by its own independent labors, rather than by the cruel, simple expedient of compelling it from his own mouth."[53] This principle, however, has not stopped the police from attempting to procure a confession from an alleged suspect. In the early 20th century, the Wickersham Reports and numerous other literature uncovered the severity of the "third degree," or the police use of violence to extract information or coerce a confession from criminal suspects.[54] The Wickersham Reports

---

of presenting demonstrative evidence to a jury is it focuses the jury's attention in a way oral testimony alone simply cannot.").

48. *See, e.g.*, United States v. Tingle, 658 F.2d 1332, 1336 (9th Cir. 1981) (finding a confession involuntary where agents made the defendant fearful that she would not see her child for a long time); United States v. Syslo, 303 F.3d 860, 867 (8th Cir. 2002) (finding that a child's presence at the police station with no one to watch the child exerted coercive influence on the defendant).

49. *See* Possley, *supra* note 43 and accompanying text.

50. *See infra* Part II.A.

51. *See infra* Part II.B.

52. *See infra* Part II.C.

53. Miranda v. Arizona, 384 U.S. 436, 460 (1966).

54. Saul M. Kassin et al., *Police-Induced Confessions: Risk Factors and Recommendations*, 34 L. & HUM. BEHAV. 3, 4 (2010); NATION'L COMM'N ON L. OBSERVANCE & ENF'T, REPORT ON LAWLESSNESS IN LAW ENFORCEMENT (1931), https://www.ojp.gov/pdffiles1/Digitization/44549NCJRS.pdf [https://perma.cc/GDW4-49J4].

led to a general distrust of the police,[55] but the police have since replaced these outdated interrogation tactics with several artfully crafted manuals that painstakingly detail the most effective methods for procuring a confession.[56] The most popular of these manuals include *Criminal Interrogation and Confessions* by Fred Inbau (hereinafter, "the Inbau Manual"),[57] *Fundamentals of Criminal Investigation* by Charles O'Hara,[58] and *The Confession: Interrogation and Criminal Profiles for Police Officers* by John Macdonald and David Michaud (hereinafter, "the Macdonald Manual").[59]

The Inbau Manual touts the effectiveness of trickery and deceit.[60] It begins by distinguishing an "interview" from an "interrogation."[61] An interview is a non-accusatory interaction for the purpose of gathering information.[62] An interviewer can conduct the interview in any environment, so long as the questions and answers flow freely and remain unrestricted.[63] The more free flowing the interview, the better—this will allow the investigator to explore unanticipated areas of information for further probing.[64] An interrogation, however, is an accusatory interaction for the purpose of "learn[ing] the truth."[65] Interrogators must conduct interrogations in a controlled area instead of an open environment because the "persuasive tactics" employed demand a private environment "free from distractions."[66]

Before the investigator conducts an interrogation, the Inbau Manual mandates that the investigator should have some reasonable suspicion of

---

55. *Third Degree Lite: The Abuse of Confessions*, CRIME REPORT (Sept. 7, 2017), https://thecrimereport.org/2017/09/07/the-third-degree-lite-the-abuse-of-confessions [https://perma.cc/FH5D-YKMA].

56. *See, e.g.*, Miriam S. Gohara, *A Lie for a Lie: False Confessions and the Case for Reconsidering the Legality of Deceptive Interrogation Techniques*, 33 FORDHAM URB. L.J. 791, 808–16 (2006) (detailing several deceptive police interrogation techniques).

57. *Id.* at 807–14 (detailing the Inbau Manuals); *see also* FRED E. INBAU ET AL., CRIMINAL INTERROGATION AND CONFESSIONS (5th ed. 2013) (ebook).

58. Gohara, *supra* note 56, at 814 (detailing the O'Hara Manuals); CHARLES E. O'HARA, FUNDAMENTALS OF CRIMINAL INVESTIGATION (1st ed. 1956).

59. Gohara, *supra* note 56, at 814–15 (detailing the MacDonald Manuals); JOHN M. MACDONALD & DAVID L. MICHAUD, THE CONFESSION: INTERROGATION AND CRIMINAL PROFILES FOR POLICE OFFICERS (1987).

60. *See* Gohara, *supra* note 56, at 807–14 (detailing the Inbau Manuals); *see also* Inbau, *supra* note 57.

61. Inbau, *supra* note 57, at 32–40.

62. *Id.* at 32.

63. *Id.* at 34.

64. *Id.*

65. *Id.* at 36.

66. INBAU, *supra* note 57, at 37.

guilt.[67] To achieve this threshold suspicion, the Inbau Manual suggests that the police first conduct an interview to lay the foundation for an inevitable interrogation.[68] This will allow the police to obtain more information, establish a level of rapport, and gain "a psychological advantage" over the suspect.[69]

The Inbau Manual promotes the "Reid Nine Steps of Interrogation" technique, in part because it "is apt to make an innocent person confess and that all the steps are legally as well as morally justifiable."[70] The nine steps are as follow:

Step #1: The investigator should begin with "a direct, positively presented confrontation of the suspect with a statement that he is considered to be the person who committed the offense."[71] The purpose of this step is for the investigator to analyze the suspect: what verbal or nonverbals responses did they give to questions? Are they combative? What is their body language?[72] After a thorough analysis, the Inbau Manual encourages the investigator to push back on any denial of guilt, and remind the suspect of the importance of telling the truth.[73]

Step #2: Also called the "Interrogation Theme," the investigator postulates why the suspect may have committed the crime.[74] This gives the suspect a "moral excuse for having committed the offense."[75] The idea here is that:

> If a suspect seems to listen attentively to the suggested "theme," or seems to be deliberating about it, even for a short period of time, that reaction is strongly suggestive of guilt. If the suspect expresses resentment over the mere submission of such a suggestion, this reaction may be indicative of innocence.[76]

The Interrogation Theme can include pinning the moral blame onto another person, such as the victim or an accomplice,[77] or providing a tangible reason for the offense, such as a financial hardship.[78] For example,

---

67. *Id.* at 38–39.
68. *Id.* at 38–40.
69. *Id.* at 40.
70. *Id.* at 425.
71. INBAU, *supra* note 57, at 426.
72. *Id.*
73. *Id.*
74. *Id.*
75. *Id.*
76. INBAU, *supra* note 57, at 427.
77. *Id.*
78. *See* People v. Thomas, 8 N.E.3d 308, 311–13 (N.Y. 2014). In *Thomas*, the police suspected that the defendant inflicted traumatic head injuries on his infant child. *Id.* at 311.

in *People v. Thomas*, the police diminished the moral blame attributed to an individual suspected of battering his infant child by reminding the suspect that his wife and in-laws had greatly stressed and angered the suspect on the day of the incident.[79]

Step #3: After establishing the Interrogation Theme, the investigator should prepare for the suspect to deny guilt.[80] The purpose of this step is to "discourage[e] the suspect's repetition or elaboration of the denial [of guilt] and return[] to the moral excuse theme" of Step #2.[81] According to the Inbau Manual, "[a]n innocent person will not allow such denials to be cut off" but will attempt to regain control over the situation "rather than to submit passively to continued interrogation," while a guilty person "will cease to voice a denial, or else the denials will become weaker, and he will submit to the investigator's return to a theme."[82]

Step #4: The initial wave of denying guilt is followed by excuses; the investigator should prepare to counter any excuse proffered by the suspect that attempts to explain how they could not have committed the crime.[83]

The "Interrogative Theme" was that the defendant was stressed with his relationships at home and accidentally inflicted injuries on his infant child. *Id.* at 312–13. The defendant's infant child was pronounced brain dead prior to the interrogation, but the police assured the defendant that he could save his infant child if he told the police how his child was injured. *Id.* at 311. The questioning by the police proceeded as follows:

> SERGEANT MASON: The doctors need to know this. Do you want to save your baby's life, all right? Do you want to save your baby's life or do you want your baby to die tonight?
>
> [DEFENDANT]: No, I want to save his life.
>
> SERGEANT MASON: Are you sure about that? Because you don't seem like you want to save your baby's life right now. You seem like you're beating around the bush with me.
>
> [DEFENDANT]: I'm not lying.
>
> SERGEANT MASON: You better find that memory right now Adrian, you've got to find that memory. This is important for your son's life man. You know what happens when you find that memory? Maybe if we get this information, okay, maybe he's able to save your son's life. Maybe your wife forgives you for what happened. Maybe your family lives happier ever after. But you know what, if you can't find that memory and those doctors can't save your son's life, then what kind of future are you going to have? Where's it going to go? What's going to happen if Matthew dies in that hospital tonight, man?

Id. at 311–12.

79. *Id.* at 312–13.

80. INBAU, *supra* note 57, at 427.

81. *Id.*

82. *Id.*

83. *Id.* at 427–28; *see also* People v. Thomas, 8 N.E.3d 308, 312 (N.Y. 2014). Four hours into the interrogation, and framing the theme as "saving his child's life," the defendant claimed he "accidentally dropped" his infant child five or six inches from the ground. *Id.* at 312. The defendant was then confronted with another officer, who claimed to have experience with head trauma from Operation Desert Storm and accused the

The suspect may offer "economic, religious, or moral reasons for not committing the crime," but according to the Inbau Manual, such excuses are "normally offered only by the guilty suspect."[84]

Step #5: As the investigator becomes dissuaded from the suspect's weak excuses, the suspect is "likely to mentally withdraw and "tune out" the investigator's theme."[85] This provides the investigator a prime opportunity to regain the suspect's attention.[86] Hence, while the suspect becomes fixated on the investigator's questions again, the investigator should elicit feelings of heightened sincerity—this may be achieved by the investigator moving closer to the suspect and maintaining direct eye contact.[87]

Step #6: At this point, the suspect is mentally weighing the possible benefits of telling the truth.[88] The investigator should pay close attention to the suspect's newly passive mood, generally reflected by a change in the suspect's nonverbal behavior; this can include "tears, a collapsed posture, [and] eyes drawn to the floor."[89]

Step #7: Once the suspect's mood becomes somber, the investigator must present an alternate line of questioning that innocuously suggests some aspects of the crime.[90] This "alternative question" should give the suspect "a choice between two explanations for possible commission of the crime."[91] The purpose behind these questions is to make it easier for the suspect to tell the truth.[92] As an example, if the investigator asks a suspected thief, "[d]id you blow that money on booze, drugs, and women and party with it, or did you need it to help out your family," any answer to the alternative becomes tantamount to a confession.[93] In short, it gives the suspect an opportunity to "tell the truth while saving face."[94]

---

defendant of lying because the doctors said the infant child's injuries "could only have resulted from a far greater application of force than defendant had described . . . comparable to those that would have been sustained by a passenger in a high-speed car collision." *Id.* at 312. The initial officer told the defendant he felt "betrayed" but proffers that perhaps the defendant "had been depressed and emotionally overwhelmed after having been berated by his wife over his chronic unemployment and that, out of frustration, he had, without intending to harm the infant, responded to his crying by throwing him from above his head onto a low-lying mattress." *Id.* at 312.

84. INBAU, *supra* note 57, at 428.
85. *Id.*
86. *Id.*
87. *Id.* at 428–29.
88. *Id.* at 429.
89. INBAU, *supra* note 57, at 429.
90. *Id.* at 429.
91. *Id.* at 682.
92. *Id.*
93. *Id.*
94. INBAU, *supra* note 57, at 682.

Step #8: After the investigator leads the suspect down this alternative line of questioning, the suspect will begin to orally relay various details about the offense.[95] This step will establish legal guilt, but the investigator must be patient.[96]

Step #9: Lastly, the confession—at this stage, the investigator must now convert the oral confession from the suspect into a fixed tangible medium.[97]

At face value, the Inbau Manual is seemingly innocent. But the devil is in the details, especially when analyzing the questioning in *People v. Thomas*.[98] The Inbau Manual suggests methods of trickery,[99] deception,[100] and making potential empty promises.[101] Indeed, Step #2 and the creation of an Interrogation Theme provides the police with great latitude to

---

95. *Id.* at 429; *see also* People v. Thomas, 8 N.E.3d 308, 312 (N.Y. 2014). After suggesting the defendant threw his infant child on a mattress without intending to harm him, the police asked the defendant to "demonstrate with a clipboard how he threw the child down on the mattress." *Id.* at 312. The officer instructed:

> Move that chair out of the way. Here hold that like you hold the baby. Turn around, look at me. Now here's the bed right here, all right. Now like I said, the doctor said that this injury is consistent with a 60 mile per hour vehicle crash, all right, all right. That means it was a very severe acceleration. It means he was going fast and stopped suddenly, all right, so think about that. Don't try to downplay this and make like it's not as severe as it is. Because [we] both know now you are finally starting to be honest, okay, all right. Maybe this other stuff you said is the truth.
>
> [DEFENDANT]: That is.
>
> SERGEANT MASON: For what the information that I need to know we both know now you are starting to finally be honest with that, all right. Hold that like you hold that baby, okay and start thinking about them negative things that your wife said to you, all right, start thinking about them kids crying all day and all night in your ear, your mother-in-law nagging you and your wife calling you a loser, all right, and let that aggression build up and show me how you threw Matthew on you bed, all right. Don't try to sugar coat it and make it like it wasn't that bad. Show me how hard you threw him on that bed.

Id. at 312–13. The enactment was captured on an interrogation video, where the defendant demonstrated throwing his infant child onto the mattress several times prior to the infant child's hospitalization. *Id.* at 313. The confession was deemed involuntary under N.Y. CRIM. PROC. LAW § 60.45(2)(b)(i). *Id.* at 316; *see also infra* note 268 and accompanying text.

96. INBAU, *supra* note 57, at 706.

97. *Id.* at 429–30.

98. Thomas, 8 N.E.3d 308 (N.Y. 2014); *see supra* notes 78, 83, 95 and accompanying text.

99. INBAU, *supra* note 57, at 569–79 (describing a tactic where the investigator pins the co-defendants against each other by suggesting the other confessed).

100. *Id.* at 564–69 (describing a tactic where the investigator misleads the suspect to believe the evidence demonstrates that the suspect committed the crime).

101. *Id.* at 562–63 (describing a tactic where the investigator should describe the benefits of telling the truth short of promising leniency).

deceive a suspect.[102] Fred Inbau in other scholarship states that he is "unalterably opposed to the use of any interrogation tactic or technique that is apt to make an innocent person confess . . . but [he does] approve of such psychological tactics and techniques as trickery and deceit" because they are "frequently necessary in order to secure incriminating information from the guilty, or investigative leads from otherwise uncooperative witnesses or informants."[103] In fact, Inbau proclaims that police interrogations are an absolute necessity for three reasons: (1) many criminal cases cannot be resolved without an admission or confession—even if conducted by the best and most equip police force in the country;[104] (2) criminal offenders ordinarily will not admit their guilt;[105] and (3) the nature of dealing with criminal offenders requires methods that are not considered "appropriate for the transaction of ordinary, everyday affairs by and between law-abiding citizens."[106]

In all, police interrogations serve an important function in our criminal legal system. While Inbau champions for "deal[ing] with criminal offenders on a somewhat lower moral plane than that upon which ethical, law-abiding citizens are expected,"[107] the importance of upholding constitutional protections against the "destructi[on] of human dignity" must be balanced with the legitimate need for law enforcement to "provide for the security of the individual and of his property."[108] As technology begins to rapidly advance, the law must shape and balance the contours of

---

102. *See supra* note 78 and accompanying text.

103. Fred E. Inbau, *Police Interrogation—A Practice Necessity*, 89 J. CRIM. L. & CRIMINOLOGY 1403, 1403–04 (1999).

104. *Id.* at 1405–06 (describing a case from his professional career with no trace of any evidence and concluding that "[w]ithout an opportunity for interrogation the police could not have solved this case.").

105. *Id.* at 1406 ("Self-condemnation and self-destruction not being normal behavior characteristics, human beings ordinarily do not utter unsolicited, spontaneous confessions. . . . [I]t is impractical to expect any but a very few confessions to result from a guilty conscience unprovoked by an interrogation.").

106. *Id.* at 1409–10 (describing a case about a woman who was murdered by her brother-in-law, "[t]he interrogation was "unethical[]" according to the standards usually set for professional, business, and social conduct. But the pertinent issue in this case was no ordinary, lawful, professional, business or social matter. It involved the taking of a human life by one who abided by no code of fair play toward his fellow human beings. The killer would not have been moved one bit toward a confession by subjecting him to a reading or lecture regarding the morality of his conduct. It would have been futile merely to give him a pencil and paper and trust that his conscience would impel him to confess. Something more was required—something which was in its essence an "unethical" practice on part of the interrogator.").

107. *Id.* at 1410.

108. Miranda v. Arizona, 384 U.S. 436, 457, 539 (1966).

practical and ethical law enforcement with the novel means for encroaching on individual liberty.

B. Deepfakes

Deepfakes involve the manipulation of video, photo, or audio files using machine learning.[109] Specifically, deepfakes are created through deep neural networks—a machine learning algorithm that learns from numerous test examples to produce an output.[110] Subpart 1 provides a general overview of how deepfakes work.[111] Subpart 2 analyzes whether deepfakes could alter a criminal suspect's memory.[112]

1. How Deepfakes Work

Generative neural networks, a type of deep neural network, are typically used to create deepfakes.[113] A neural network is created from numerous computer algorithms that together replicate how the human brain processes information.[114] The network is further comprised of numerous layers, where each layer includes a series of nodes; the nodes perform mathematical transformations to convert an input into a different output.[115] The more layers a network has, the "deeper" the network becomes; and the deeper a network becomes, the better the network is at creating seemingly real images.[116]

The compilations of algorithms which form the neural network are diverse in type. For example, some networks use two different algorithms—one algorithm that creates a deepfake and another algorithm that attempts to detect that deepfake—that work in tandem to strengthen the capabilities of the other.[117] This meaning, if the purpose of the

109. Karen Howard, *Deconstructing Deepfakes—How Do They Work and What Are the Risks?*, U.S. GOV'T ACCOUNTABILITY OFF. (Oct. 20, 2020), https://www.gao.gov/blog/deconstructing-deepfakes-how-do-they-work-and-what-are-risks [https://perma.cc/TY4J-FGNJ]; *Artificial Intelligence (AI) vs. Machine Learning*, COLUM. ENG'G, https://ai.engineering.columbia.edu/ai-vs-machine-learning [https://perma.cc/K56S-DTRN] (last visited Apr. 16, 2024) (describing machine learning as a pathway to and subcategory of AI).

110. *What the Heck is a Deepfake?*, *supra* note 36.

111. *See infra* Part II.B.1.

112. *See infra* Part II.B.2.

113. Doss, *supra* note 30.

114. *See id.*; *What the Heck is a Deepfake?*, *supra* note 36.

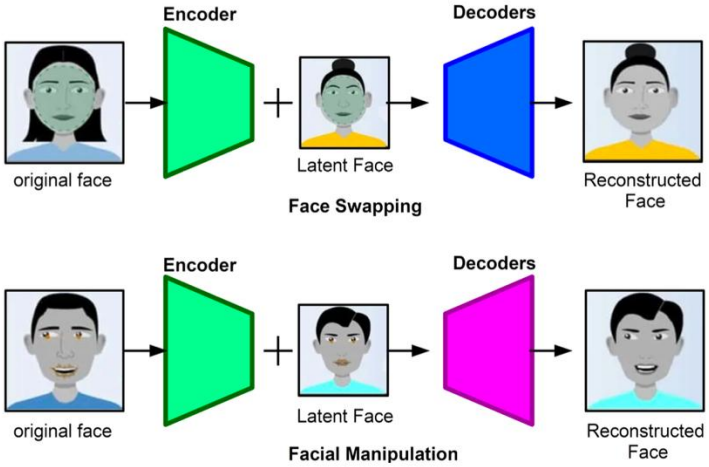
115. *What the Heck is a Deepfake?*, *supra* note 36.

116. *Id.*

117. *Id.*; *see also* Abdulqader M. Almars, *Deepfakes Detection Techniques Using Deep Learning: A Survey*, 9 J. OF COMPUT. & COMMC'N 20, 24–25 (2021) (describing encoder and decoder algorithms used for deepfake generation).

"creation algorithm" is to create as real a deepfake as possible, and the purpose of the "detection algorithm" is to accurately detect high quality deepfakes (that is, the "realest looking" deepfakes), the algorithms will continuously improve because each algorithm is using higher quality training data.[118]

The below image provides an exemplary illustration for the process of doctoring an image using deepfake technology[119]:



Beginning on the left, a photograph of an original face is extracted and processed by a deep neural network encoder.[120] An algorithm breaks down the first image into numerous discrete elements; using the example of facial manipulation, these discrete elements include representations of the nose shape, skin tone, eye color, hair color, wrinkles, and more.[121] Then, the discrete elements are input into an encoder to provide the neural network with a latent face—an "information rich" representation of the face.[122] A decoder then transforms the latent face back into a reconstructed face.[123] On the right, the final reconstructed image of a new face is created.[124]

---

118. *What the Heck is a Deepfake?*, *supra* note 36.

119. Robail Yasrab et al., *Fighting Deepfakes Using Body Language Analysis*, 3 FORECASTING 303, 306 (2021).

120. *Id.*

121. Jye Sawtell-Rickson, *What Is a Deepfake?*, BUILT IN (Apr. 4, 2024), https://builtin.com/machine-learning/deepfake [https://perma.cc/3D9W-8WQB].

122. *Id.*

123. *Id.*; Yasrab, *supra* note 119, at 305–06.

124. Yasrab, *supra* note 119, at 306.

A similar process applies to video footage and audio files.[125] By granularly dissecting each feature of a video frame or an audio sound byte, and iteratively superimposing and manipulating these features onto an existing video or audio file, one can create a whole new video or audio recording that seems real to the naked eye or the untrained ear.[126]

2. Can Deepfakes Alter One's Memory?

The "Misinformation Effect" is the theory that information obtained after an event has lapsed can distort an individual's memory.[127] False memories occur when an individual "inadvertently and unconsciously attribute[s] an internally generated mental experience to an incorrect source."[128] The phenomenon of "Imagination Inflation" occurs when an individual repeatedly imagines a fictional event and begins to generate fictional memories of that event.[129] These phenomena have been tested in psychological experiments to determine whether the use of a false, incriminating statement may lead an individual to falsely confess in Subpart a,[130] or whether the use of a false, incriminating video may lead an individual to falsely confess in Subpart b.[131]

a. The Use of a False, Incriminating Statement

Experiments demonstrate that a false, incriminating statement made by a third-party could lead an individual to accept guilt for a crime they did not commit.[132] A study conducted by Saul M. Kassin and Katherine L. Kiechel of Williams College (the "Kassin & Kiechel study") researched the frequency of when a false, incriminating statement would produce a false confession.[133] The study found that after a lead researcher confronted

125. *The Good and Bad Perspectives of Deepfakes*, KAMLESHGSINGH (Sept. 6, 2022), https://kamleshgsingh.com/2022/09/26/the-good-and-bad-perspectives-of-deepfakes [https://perma.cc/C8W3-SX9T].

126. *Id.*

127. Gillian Murphy et al., *Face/Off: Changing the Face of Movies with Deepfakes*, PLOS ONE (July 2023) at 3–4, https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0287503 [https://perma.cc/P6H7-E46J].

128. *Id.* at 4.

129. *Id.*; Maryanne Garry & Devon L.L. Polaschek, *Imagination and Memory*, 9 CURRENT DIRECTIONS IN PSYCH. SCI. 6, 7–8 (Feb. 2000).

130. *See infra* Part II.B.2.a.

131. *See infra* Part II.B.2.b.

132. *See, e.g.*, Saul M. Kassin & Katherine L. Kiechel, *The Social Psychology of False Confessions: Compliance, Internalization, and Confabulation*, 7 AM. PSYCH. SOC'Y 125, 126–27 (1996).

133. *Id.* at 126.

a subject with an incriminating statement, that subject was likely to sign a confession.[134]

The Kassin & Kiechel study sought to emulate a criminal interrogation and confession—a topic which, in May 1996, had never been "systemically examined."[135] The Kassin & Kiechel study focused on three types of documented false confessions: (1) voluntary; (2) coerced-compliant; and (3) coerced-internalized.[136] A "voluntary" false confession is one where an individual confesses in the absence of external pressure.[137] A "coerced-compliant" false confession is one where an individual confesses for a particular reason, such as to end a contentious interrogation, securing a promise, or avoiding a threatened harm.[138] Lastly, a "coerced-internalized" false confession is one where an individual confesses because they actually became to believe that they committed the crime.[139]

The Kassin & Kiechel study was conducted as follows: (1) one subject and one confederate[140] filled out a brief survey concerning their typing experience and ability, spatial awareness, and reflex speed; (2) the subject and the confederate sat across from each other where the confederate would read aloud a list of letters that the subject typed onto a computer—the subject and the confederate would switch roles after some time elapsed; (3) before the experiment began, however, both the subject and confederate were specifically told not to press the "ALT" key because the ALT key would cause the computer system to crash; (4) after one minute of the subject on the computer, the computer would suddenly crash and the experimenter would accuse the subject of pressing the ALT key.[141] No subject ever hit the ALT key—all subjects remained in compliance during the entire experiment.[142] Additionally, the experiment included two other variables: first, the experimenters emulated a subject's "vulnerability" by varying the pace of the task—the faster the task, the more vulnerable the subject was to incriminating evidence—and second, to emulate the level

134. *Id.* at 126–27.
135. *Id.* at 125.
136. *Id.* at 125.
137. Kassin & Kiechel, *supra* note 132, at 125.
138. *Id.*
139. *Id.*
140. *Confederate*, https://www.alleydog.com/glossary/definition.php?term=Confederate [https://perma.cc/X3RC-2M5P] (last visited Apr. 16, 2024). A confederate in psychology is one who acts as a participant but is secretly on the research team. *Id.*
141. Kassin & Kiechel, *supra* note 132, at 126.
142. *Id.* at 126–27.

of "false incriminating evidence," the confederate would occasionally accuse the subject of pressing the ALT key.[143]

The Kassin & Kiechel study measured three forms of social influence: (1) compliance, (2) internalization, and (3) confabulation.[144] To measure compliance, the experimenter would ask the subject to sign a statement stating "I hit the 'ALT' key and caused the program to crash. Data were lost," with the consequence of receiving a phone call from the principal investigator.[145] To measure internalization, a second confederate was introduced.[146] After the experimenter accused the subject of pressing the ALT key, the experimenter would scold the subject and explain that the experiment would need to be repeated.[147] Once the experimenter left the room, the second confederate would subtly ask the subject "what happened" and recorded the subject's response verbatim.[148] Lastly, to measure confabulation, the experimenter would bring the subject back into the lab, read the list of typed letters, and asked the subject if they could identify where they hit the ALT key in the list of letters.[149] The purpose of this exercise was to determine whether the subject would recall a certain fake memory to help resolve liability.[150]

Of all the subjects, 69% signed the confession, 28% demonstrated internalization, and 9% confabulated details to corroborate the false, self-incriminating statements.[151] The table below provides a breakdown of the rates of confession, internalization, and confabulation among the participants[152]:

143. *Id.* at 127. In cases where the confederate did not turn on the subject, the confederate proclaimed they did not see anything. *Id.*

144. *Id.*

145. *Id.* at 126.

146. Kassin & Kiechel, *supra* note 132, at 126.

147. *Id.*

148. *Id.* ("The subject's reply was recorded verbatim and later coded for whether or not he or she had unambiguously internalized guilt for what happened [such as] 'I hit the wrong button and ruined the program'; 'I hit a button I wasn't supposed to.'"). Notably, the responses were conservatively scrutinized—a reply that began with "he said" or "I may have" or "I think" was not taken as internalization evidence. *Id.*

149. *Id.*

150. *Id.* at 126–27.

151. Kassin & Kiechel, *supra* note 132, at 127.

152. *Id.*

Form of influence	No witness		Witness	
	Slow pace	Fast pace	Slow pace	Fast pace
Compliance	35_a	65_b	89_{bc}	100_c
Internalization	0_a	12_{ab}	44_{bc}	65_c
Confabulation	0_a	0_a	6_a	35_b

The data presents two findings. First, when the confederate ("witness") would turn on the subject, the subject was more likely to comply, internalize, or confabulate.[153] Second, the more vulnerable ("fast pace") the subject, the more likely the subject was to comply, internalize, or confabulate.[154]

The principal drawback from the Kassin & Kiechel study is that "[the] procedure focused on an act of *negligence and low consequence* [which] may well explain why the compliance rate was high."[155] Nonetheless, the Kassin & Kiechel study's most important finding was that "many subjects privately internalized guilt for an outcome they did not produce, and that some even constructed memories to fit that false belief."[156] Thus, while compliance may have been impacted by the low stakes involved in the experiment, the resulting internalization by participants was "not seriously compromised by the laboratory paradigm that was used."[157]

b. The Use of a False, Incriminating Video

Another experiment demonstrates that the use of doctored videos may amplify the risk of generating false memories.[158] A study conducted by Robert Nash and Kimberly Wade of the University of Warwick (the "Nash & Wade study") researched the impact of doctored videos on false confessions and found that *all* individuals who viewed a video of themselves cheating at a task believed that they had actually cheated in

153. *Id.* at 126–27.
154. *Id.* at 127.
155. *Id.* (emphasis added).
156. Kassin & Kiechel, *supra* note 132, at 127.
157. *Id.*
158. *See, e.g.*, Robert Nash & Kimberly Wade, *Innocent But Proven Guilty: Eliciting Internalized False Confessions Using Doctored Video Evidence*, 23 APPLIED COGNITIVE PSYCH. 624, 633 (2009) (proclaiming itself as the "first study to demonstrate the dangers of modern digital manipulation technology when encouraging people to remember self-involving, recently occurring experiences . . . and on a broader level . . . show[ing] that seeing fake evidence is more convincing than being merely told of its existence" (internal citations omitted)).

that task.[159] The Nash & Wade study was conducted as follows: (1) subjects completed a computerized gambling task, then retired from the experiment for a break; (2) when the subjects returned from their break, the experimenters accused the participants of cheating on the task; (3) the experimenters told all participants that they were caught cheating on a video; and (4) the experimenters randomly selected half of the participants who were individually shown a fake video of them cheating during the experiment.[160]

The experimenters told the subjects that the experiment concerned a study on the impact of gambling with electronic and physical credit.[161] Unbeknownst to the subjects, the experimenters placed all the subjects in the "physical credit" group.[162] Additionally, the experimenters informed each subject that their participation would be video recorded.[163] The gambling task consisted of fifteen multiple choice questions administered on a computer, where each subject would digitally input an amount of money to gamble on each question.[164] If the subject answered the question correctly, a green check would appear on their computer screen with instructions to take a certain amount of "fake money" from the "bank."[165] If the subject answered the question incorrectly, a red cross appeared with instructions to return money to the bank.[166]

After completing the fifteen questions, the participants left the study room for a break,[167] and the experimenters split the subjects into two groups: "See-Video Subjects" and "Told-Video Subjects."[168] Then, the experimenters altered a ten to twenty second segment of each of the See-Video" subjects' participation by digitally replacing the green check with a red cross.[169] The doctored video appeared as below[170]:

159. *Id.*at 633–34.
160. *Id.* at 626–28.
161. *Id.* at 626.
162. *Id.*
163. Nash & Wade, *supra* note 158, at 626.
164. *Id.* at 627.
165. *Id.* Each subject was given a pile of fake money to gamble with and another pile of fake money represented the bank. *Id.* at 626.
166. *Id.* at 627.
167. *Id.* at 626.
168. Nash & Wade, *supra* note 158, at 626.
169. *Id.* at 627–28.
170. *Id.*

The left image illustrates the video prior to any alteration, while the right image illustrates the video after alteration.[171] After the subjects returned from break, the experimenters individually informed each subject that they were caught on video stealing money from the bank.[172] This meaning, each subject was accused of improperly *taking* money from the bank when receiving a red cross, rather than *putting* money into the bank as instructed.[173] After accusing each subject of cheating, the experimenters played the doctored video for the "See-Video" subjects and did not show the "Told-Video" subjects anything.[174] During the experiment, no subject actually took money from the bank when they were not supposed to do so.[175]

The experimenters asked all subjects to sign a confession form to acknowledge that they took money from the bank when they should have returned it.[176] The experimenters explained that by signing the confession form, the subjects would not receive their compensation for participating in the study.[177] If any subject refused to sign the confession form, the subject could appeal the accusation to the "professor in charge" who would review the video; however, according to the experimenters, the professor in charge would likely withhold the subject's payment because "the video clearly showed that [the subject] took the money."[178]

Regardless of whether the subject elected to sign the confession, the experimenters asked each subject to write down their thoughts about the investigation.[179] The memorialization of the subjects' thoughts helped the

171. *Id.*
172. *Id.* at 625–28.
173. Nash & Wade, *supra* note 158, at 627.
174. *Id.* at 628.
175. *Id.* at 629.
176. *Id.* at 628.
177. *Id.*
178. Nash & Wade, *supra* note 158, at 628.
179. *Id.*

experimenters analyze the mental impressions of the exercise and was critical to determine whether the subjects "figured out the true nature of the experiment."[180] Subjects who signed the confession wrote their thoughts about the investigation with the confession, while subjects who refused to sign the confession wrote their thoughts in a separate waiting room while the experimenter spoke with the professor in charge.[181]

The experimenters planted a confederate in the waiting room who would initiate a conversation with the other subjects about the investigation.[182] The confederate recorded each subject's response.[183] Some subjects outright denied culpability, but others began to internalize and confabulate the events.[184] Based on their responses, each subject was further categorized into different subgroups: (1) "no internalization," "partial internalization," or "full internalization"; or (2) "no confabulation," "hypothesizing," or "full confabulation."[185] Subjects categorized as "partial internalization" made statements such as, "I think I messed up [the] experiment," whereas "full internalization" subjects made statements such as, "I took money when I was supposed to give it back."[186] Subjects categorized as "hypothesizing" made statements such as "I probably expected I was right, and didn't take any notice of the cross," whereas "full confabulation" subjects made statements such as "I was concentrating so hard on the money, I forgot to give rather than take when I was wrong."[187]

The Nash & Wade study found that 20% of all subjects partially internalized their actions, 63% of all subjects fully internalized the video, 7% of all subjects hypothesized about *why* they "cheated," and 3% of subjects fully confabulated details about the cheating scandal.[188] Additionally, the study found that the "See-Video" subjects were more likely to confess without resistance.[189] Moreover, the "See-Video" subjects filled in fake gaps in their memory to conform with the idea that they cheated.[190] Most strikingly, *every* participant—regardless of whether they saw the fake video—signed a "confession."[191] On the first confession

180. *Id.*
181. *Id.*
182. *Id.*
183. Nash & Wade, *supra* note 158, at 628–29.
184. *Id.* at 628.
185. *Id.* at 629.
186. *Id.*
187. *Id.*
188. Nash & Wade, *supra* note 158, at 629–30.
189. *Id.* at 629.
190. *Id.* at 630.
191. *Id.*

attempt, 87% of subjects signed the confession form, with the remaining 13% of subjects signing on the second confession attempt.[192]

In all, the study concluded that "a combination of social demand, phony evidence and false suggestion from a credible source can lead a substantial number of people to falsely confess and believe they committed an act they never did."[193] Notably, this study was conducted with technology from 2008—well before deepfake technology and other artificial intelligence allowed individuals to manipulate media in a remarkably indistinguishable manner.[194] Given that deepfakes alone can create a false memory rate of 49% in some studies,[195] the risk of generating false memories is amplified when using deepfake technology.[196]

c. The Cross-Over

The Kassin & Kiechel study demonstrates that an individual may recall false memories to explain a bad act, while the Nash & Wade study demonstrates that visual aids may intensify the false memories recalled.[197] The greatest limit to these studies is that individuals may be more reluctant to confess to a bad act if their life or liberty were at stake, rather than an insubstantial amount of money.[198] Additionally, both studies included subjects that were college students, which severely limits the diversity of the studies' population.[199]

---

192. *Id.* at 629.

193. Nash & Wade, *supra* note 158, at 629.

194. *See id.* at 624; *see supra* Part II.B.

195. Murphy, *supra* note 127, at 8, 14.

196. Nadine Liva & Dov Greenbaum, *Deep Fakes and Memory Malleability: False Memories in the Service of Fake News*, 11 AJOB NEUROSCIENCE 96, 100 (2020) ("Creating false memories to influence one's perception of reality is morally challenging. However, . . . it is not a new phenomena. What is new is the richness of the stimuli provided by technologies, such as deep fakes, and thus their increased potential to influence the human brain in terms of engaging neural networks and controlling mental, perceptual, emotional, and cognitive states."); Vaccari & Chadwick, *supra* note 21, at 1–2, 9; *c.f.* Murphy, *supra* note 127, at 14 ("Without a doubt, deepfakes have the potential to misinform and to cause real harm, but more empirical evidence is required before we can quantify these harms, weigh the benefits, and intervene where necessary.").

197. Kassin & Kiechel, *supra* note 132, at 127; Nash & Wade, *supra* note 158, at 625 ("[I]n both legal and everyday decision-making tasks people are more persuaded by visual than by verbal evidence.").

198. *But see* Nash & Wade, *supra* note 158, at 633 ("While our cheating event differs dramatically from the crimes that suspects are accused of, in particular because the act of taking money could have been considered unintentional, our findings are proof that many people will readily confess and develop erroneous beliefs if they are accused of an act and told about or confronted with false-video evidence.").

199. *Id.* at 626 ("Thirty students (13 males, 17 females; M = 21.20 years, SD = 2.48, range = 18–27) at Warwick University received £6 for participating.").

But regardless of these shortcomings, the underlying principle is concrete: individuals confronted with fake, tangible evidence against them fabricated memories to falsely rationalize non-existent actions.[200] Unlike a custodial interrogation, the subjects in the experiments were not in custody or subject to hours of interrogation; there would be no motive to "lie" just to end any interrogation proceeding in these studies because no interrogation proceeding ever existed.[201] The mere testimonial confrontation, followed by tangible evidence of the crime, was enough for a majority of subjects to confess without hours of interrogation.[202]

So, while the Kassin & Kiechel and Nash & Wade studies do not involve criminal suspects, the studies do involve the psychological effects of false evidence on accused individuals. Nonetheless, "[b]asic research shows that once people see an outcome as inevitable, cognitive and motivational forces conspire to promote their acceptance, compliance with, and even approval of the outcome."[203] If laboratory studies demonstrate that false evidence ploys over a decade ago can nearly double the rate of false confessions, deepfake technology is powerful enough to aggravate these already staggering statistics.[204] Deepfake distortion requires a different paradigm than other types of police distortion because deepfake technology and fake physical evidence in non-custodial settings are more likely to deceive individuals when compared to the existing hearsay distortion presented by police departments.[205] Once deepfake technology creeps into a custodial police setting, the rate of false and involuntary confessions is certain to increase.[206]

III. CONSTITUTIONALITY OF LYING DURING POLICE INTERROGATIONS

Three distinct constitutional principles limit the scope of police interrogations: (1) the *Miranda* safeguards associated with the Fifth Amendment Privilege Against Self-Incrimination; (2) the Right to Counsel afforded under the Sixth Amendment; and (3) the "voluntariness test" of the Due Process Clause.[207] A confession is inadmissible for any

200. Kassin & Kiechel, *supra* note 132, at 127; Nash & Wade, *supra* note 158, at 629.

201. *See, e.g.*, *False Confessions*, INNOCENCE PROJECT, https://innocenceproject.org/false-confessions [https://perma.cc/TY9Y-4VK8] (last visited Apr. 17, 2024).

202. Nash & Wade, *supra* note 158, at 633.

203. Kassin, *supra* note 54, at 16–17.

204. *Id*. at 16–18.

205. *C.f. supra* note 47 and accompanying text.

206. Nash & Wade, *supra* note 158, at 633; Kassin, *supra* note 54, at 16–17; *see* Illinois v. Perkins, 496 U.S. 292, 296 (1990).

207. Stephen J. Schulhofer, *Confessions and the Court*, 79 MICH. L. REV. 865, 866–67 (1981).

purpose if it is involuntarily given, and is inadmissible for substantive purposes if the proper procedural safeguards described in *Miranda* are absent.[208] Additionally, a confession will be inadmissible if obtained in violation of the Sixth Amendment Right to Counsel, which attaches at the moment when formal judicial proceedings begin.[209] This Article assumes that the proper procedural safeguards of *Miranda* are satisfied and the Right to Counsel under the Sixth Amendment is inapplicable—thus, the focus remains on whether a confession is considered involuntary, which concerns an analysis of the "totality of the circumstances" to determine whether "overt physical coercion or patent psychological ploys" served to "overbear the will" of the suspect.[210] More specifically, this Article focuses on whether the presence of deepfakes alone are "patent psychological ploys" that should be considered involuntary *per se*.[211]

Generally, the police may lie to a suspect during custodial interrogations.[212] Coercive police activity, however, and the lies that spawn from such coercion, may sometimes render a confession involuntary and therefore inadmissible.[213] Subpart A details the constitutional history underlying specific tactics used by the police during interrogations.[214] Subpart B sets forth the current test under the law, as well as the constitutional protections formulated over time.[215] Subpart C argues that the use of deepfake technology alone should render a confession involuntary because of how powerful the deception produced by deepfakes may be.[216]

---

208. Miranda v. Arizona, 384 U.S. 436, 457 (1966) ("To be sure, the records do not evince overt physical coercion or patented psychological ploys. The fact remains that in none of these cases did the officers undertake to afford appropriate safeguards at the outset of the interrogation to insure that the statements were truly the product of free choice."). The *Miranda* Court recognized that the interrogation room atmosphere "carries its own badge of intimidation . . . not physical intimidation, but [the atmosphere] is equally destructive of human dignity." *Id.*; *see also* Harris v. New York, 401 U.S. 222, 226 (1971) ("The shield provided by Miranda cannot be perverted into a license to use perjury by way of a defense, free from the risk of confrontation with prior inconsistent utterances. We hold, therefore, that petitioner's credibility was appropriately impeached by use of his earlier conflicting statements.").

209. Massiah v. United States, 377 U.S. 201, 206–07 (1964).

210. *Miranda*, 384 U.S. at 457, 469; *see also* Rogers v. Richmond, 365 U.S. 534, 544 (1961).

211. *Miranda*, 384 U.S. at 457.

212. Frazier v. Cupp, 394 U.S. 731, 739 (1969).

213. Tinsley, *supra* note 46, §§ 10–26.

214. *See infra* Part III.A.

215. *See infra* Part III.B.

216. *See infra* Part III.C.

A. A Primer on the History of Police Interrogations and the Constitution

When determining whether a suspect's confession is admissible under the Due Process Clause, the Court in *Culombe v. Connecticut*[217] set forth the voluntariness test, which requires: (1) a confession that is "the product of an essentially free and unconstrained choice by its maker"; where (2) "his free will has [not] been overborne and his capacity for self-determination [is not] critically impaired."[218] A judge will determine voluntariness on a case-by-case basis, which typically involves an analysis of: (1) the totality of the circumstances leading up to and surrounding the confession; and (2) the psychological facts surrounding the suspect's mental state, which requires looking at how the suspect reacted to the "totality of the circumstances."[219] Then, the judge will apply the totality of the circumstances and psychological facts to the law and determine whether the suspect's reaction was legally significant.[220]

While *Culombe* concerns the Due Process Clause of the Fourteenth Amendment, the same voluntariness analysis applies to the Fifth Amendment because the Court in *Malloy v. Hogan*[221] officially incorporated the Fifth Amendment Privilege Against Self-Incrimination Clause under the Due Process Clause.[222] The *Malloy* Court also clarified that an inquiry of whether a confession was "free and voluntary" necessarily requires that the confession "not be extracted by any sort of threats or violence, nor obtained by any direct or implied promises, however slight, nor by the exertion of any improper influence."[223] The Court proclaimed that, "[g]overnments, state and federal, are thus constitutionally compelled to establish guilt by evidence independently and freely secured," and cannot prove a charge "against an accused out of his own mouth."[224]

The Court in *Massiah v. United States*[225] deviated from the rights afforded under the Fifth Amendment Privilege Against Self-Incrimination to hold that the Sixth Amendment Right to Counsel attaches to an indicted defendant under interrogation by the police in extrajudicial proceedings.[226]

---

217. 367 U.S. 568 (1961).
218. *Id.* at 602.
219. *Id.* 603–04.
220. *Id.* at 604.
221. 378 U.S. 1 (1964).
222. *Id.* at 6.
223. *Id.* at 7.
224. *Id.* at 8.
225. 377 U.S. 201 (1964).
226. U.S. CONST. amends. V–VI; *id.* at 206–07.

That same year, the Court clarified in *Escobedo v. State of Illinois*[227] that the Sixth Amendment Right to Counsel will attach when the investigation is no longer a "general inquiry" and the suspect is taken into custody to be questioned.[228] Hence, when the police denied Escobedo's request to speak with his attorney, and because he was not effectively warned of his right to remain silent, any statement elicited during that time must be suppressed.[229]

B. How the Court in Miranda *Reshaped the Bounds of Self-Incrimination*

The Court granted certiorari in *Miranda v. Arizona*[230] to clarify *Escobedo* and "explore some facets of the problems, thus exposed, of applying the privilege against self-incrimination to in-custody interrogation, and to give concrete constitutional guidelines for law enforcement agencies and courts to follow."[231] In *Miranda*, which consisted of several consolidated cases challenging the constitutionality of specific confessions, the Court attempted to set bright line rules for the police to follow: "the prosecution may not use statements, whether exculpatory or inculpatory, stemming from custodial interrogation of the defendant unless it demonstrates the use of procedural safeguards effective to secure the privilege against self-incrimination."[232] The Court defined "custodial interrogation" as one "initiated by law enforcement officers after a person has been taken into custody or otherwise deprived of his freedom of action in any significant way."[233] The Court outlined the "procedural safeguards" as follows: the suspect must be warned that (1) "he has a right to remain silent," (2) "any statement he does make may be used as evidence against him" in a court of law, and (3) "he is entitled to a lawyer and that if he cannot afford one, a lawyer will be provided for him prior to any interrogation."[234] Only after these rights are conveyed to and understood by the suspect may the suspect waive these rights, provided that "the waiver is made voluntarily, knowingly[,] and intelligently."[235] Hence, "[u]nless adequate protective devices are employed to dispel compulsion inherent in custodial surroundings, no

227. 378 U.S. 478 (1964).
228. *Id.* at 490–91.
229. *Id.* at 491–92.
230. 384 U.S. 436 (1966).
231. *Id.* at 441–42.
232. *Id.* at 444.
233. *Id.*
234. *Id.* at 444, 474.
235. *Id.* at 444.

statement obtained from the defendant can truly be the product of his free choice."[236]

The Court in *Illinois v. Perkins*[237] provides a stark exception to *Miranda* where the Court held that the use of an undercover agent to obtain a confession did not constitute a custodial interrogation necessitating the administration of *Miranda* warnings.[238] However, while "psychological ploys" that are designed to elicit incriminating responses typically constitute an interrogation sufficient to trigger *Miranda*,[239] Supreme Court precedent makes clear that coercion renders a sufficient warning null, and that "a finding of coercion need not depend upon actual violence by a government agent," because coercion can be both "mental as well as physical."[240]

Hence, the crux of any issue regarding the types of questions the police may ask concern the "voluntary" prong of *Miranda* and the jurisprudence detailing when a confession is made "voluntarily."[241] This is because a statement effectuating self-incrimination must be a voluntary waiver of the suspect's rights according to *Miranda*.[242] For example the Court in *Frazier v. Cupp*[243] held that the police falsely telling the defendant that his accomplice confessed alone will not render the defendant's confession

236. *Miranda*, 384 U.S. at 458.

237. 496 U.S. 292 (1990).

238. *Id.* at 296. Notably, though, Justice Brennan's concurrence speaks to the matter here: "The deception and manipulation practiced on respondent raise a substantial claim that the confession was obtained in violation of the Due Process Clause." *Id.* at 301 (Brennan, J., concurring in judgment). Justice Brennan concurred only because the sole issue concerned applicability of *Miranda* but cautioned that certain interrogation techniques "are so offensive to a civilized system of justice that they must be condemned under the Due Process Clause of the Fourteenth Amendment." *Id.* (quoting Miller v. Fenton, 474 U.S. 104, 109–10 (1985)). In conclusion, Justice Brennan argues that the deliberate use of deception is incompatible "with a system that presumes innocence and assures that a conviction will not be secured by inquisitorial means," raising doubt that a suspect's will was not overborne. *Id.* at 303 (quoting *Miller*, 474 U.S. at 116).

239. *See, e.g.*, Arizona v. Mauro, 481 U.S. 520, 526–27 (1987); United States v. Lafferty, 503 F.3d 293, 305 (3d Cir. 2007) (describing psychological ploys to interrogate a codefendant in the same room as a defendant who had invoked *Miranda* rights); United States v. Orso, 266 F.3d 1030, 1033–34 (9th Cir. 2001) (describing psychological ploys to lie to a defendant about evidence during a ride to a formal interview), *overruled on other grounds by* Missouri v. Seibert, 542 U.S. 600 (2004), *and* United States v. Williams, 435 F.3d 1148 (9th Cir. 2006).

240. Arizona v. Fulminante, 499 U.S. 279, 287 (1991).

241. Frazier v. Cupp, 394 U.S. 731, 739 (1969).

242. *Id.* Notably, though, the Court in *Frazier* stated that "*Miranda* does not apply" to all post-*Escobedo* cases. *Id.* at 738–39.

243. 394 U.S. 731 (1969).

involuntary because lying alone is not enough to render an otherwise voluntary confession inadmissible.[244]

While the contours of what "lies" are permissible will depend on different circuit court applications, *Frazier* is a longstanding decision that is applicable in numerous contexts that insulates the practice of lying from a ruling of involuntariness.[245] This is especially true since the Court's ruling in *Colorado v. Connelly*[246] further requires "coercive police activity" as a "necessary predicate" prior to holding that a confession is involuntary.[247] For example, a false promise is not *per se* coercion if the promise is not specific.[248] The Sixth Circuit has held that promising leniency will render a confession coerced "if [the promise] was broken or illusory," where "fair-minded jurists could conclude" that the promise of leniency was genuine.[249] Other lies, such as fabricating a suspect's

244. *Id.* at 739.

245. *See, e.g.*, Oregon v. Elstad, 470 U.S. 298, 317 (1985); United States v. Jacques, 744 F.3d 804, 812 (1st Cir. 2014) ("Exaggerating the quality of their evidence, minimizing the gravity of Jacques's offense, and emphasizing the negative media attention that would attend Jacques's trial all fall safely within the realm of the permissible "chicanery" sanctioned by this and other courts"); Green v. Scully, 850 F.2d 894, 903 (2d Cir. 1988) (holding that a statement is not rendered involuntary by police misrepresenting the strength of the evidence against the suspect); United States v. Velasquez, 885 F.2d 1076, 1087–88 (3d Cir. 1989) (holding that false information directly related to a suspect's connection to the crime will not render a confession involuntary); United States v. Whitfield, 695 F.3d 288, 302–03 (4th Cir. 2012) (holding that the police's misrepresentation that the victim was alive and identified the defendant as the culprit did not render the defendant's statement involuntary); Ledbetter v. Edwards, 35 F.3d 1062, 1068 (6th Cir. 1994) (holding that a misrepresentation about fingerprint evidence will not render a confession involuntary); Robinson v. Skipper, 2020 WL 4728087, at *1–2 (6th Cir. July 13, 2020) (holding that falsely telling the defendant that two other suspects had implicated him in the victim's murder and that if he cooperated, the judge and jury would be more lenient did not make his confession involuntary); Johnson v. Pollard, 559 F.3d 746, 754–55 (7th Cir. 2009) (holding that a statement related to the strength of the State's case is not involuntary); Ortiz v. Uribe, 671 F.3d 863, 871 (9th Cir. 2011) (holding that lying about being a law enforcement officer will not make a confession involuntary); Lucero v. Kerby, 133 F.3d 1299, 1311 (10th Cir. 1998) (holding that misrepresentations about fingerprint evidence does not make a confession involuntary); Morgan v. Zant, 743 F.2d 775, 779–80 (11th Cir. 1984), *overruled on other grounds*, 784 F.2d 1479 (11th Cir. 1986) (holding that misrepresenting fingerprint evidence does not make a confession involuntary); United States v. Mohammed, 693 F.3d 192, 198 (D.C. Cir. 2012) (holding that a lie about testing positive test for drugs does not make a confession involuntary).

246. 479 U.S. 157 (1986).

247. *Id.* at 167.

248. *See* Arizona v. Fulminante, 499 U.S. 279, 285 (1991) (rejecting the argument that a confession could not be obtained by "any direct or implied promises," but finding a promise to protect the suspect from threatened violence by others rendered the confession involuntary).

249. *See* United States v. Binford, 818 F.3d 261, 271–72 (6th Cir. 2016) (explaining that, although broken or illusory promises may be coercive, "promises to recommend

connection to a crime, is less likely to lead to an involuntary confession.[250] As the Seventh Circuit has noted, "[o]f the numerous varieties of police trickery, . . . a lie that relates to a suspect's connection to the crime is the least likely to render a confession involuntary."[251]

Other lies which psychologically pressure a suspect may sufficiently "overbear a defendant's will" to make any subsequent statement inadmissible.[252] The First Circuit recognizes that psychological duress may suffice to overbear a suspect's will to force an involuntary confession.[253] For example, a statement to a suspect that non-cooperation will prolong separation from their family is a fact that might lead to an involuntary confession.[254] Other circuits, like the Eleventh Circuit, have found that deception which directly aims at the nature of a suspect's rights, or the consequences of waiving those rights, may lead to an involuntary statement.[255]

C. How Would Deepfakes Fit into the Current Standard?

Under the current "totality of the circumstances" test, deepfakes would be weighed in context with other police trickery. But deepfakes are more than a lie—deepfakes are a deceptive alteration of reality.[256] Consider the *Tankleff v. Senkowski*[257] case as an example.[258] The investigators staging a

---

leniency and speculation that cooperation will have a positive effect do not make subsequent statements involuntary" (quoting United States v. Delaney, 443 F. App'x 122, 129 (6th Cir. 2011)).

250. Holland v. McGinnis, 963 F.2d 1044, 1051 (7th Cir. 1992).

251. *Id.*

252. United States v. Jackson, 608 F.3d 100, 102–03 (1st Cir. 2010).

253. *Id.*

254. *See, e.g.*, Lynumn v. Illinois, 372 U.S. 528, 534 (1963) (finding the defendant's confession involuntary where "the petitioner's oral confession was made only after the police had told her that state financial aid for her infant children would be cut off, and her children taken from her, if she did not 'cooperate,'" among other factors); United States v. Tingle, 658 F.2d 1332, 1336 (9th Cir. 1981) (finding confession involuntary where agents made defendant "fear that, if she failed to cooperate, she would not see her young child for a long time"); Spano v. New York, 360 U.S. 315, 323 (1959) (lying to the defendant that his close friend would lose his job if defendant did not make a statement).

255. *See, e.g.*, Hart v. Att'y Gen. of the State of Fla., 323 F.3d 884, 894–95 (11th Cir. 2003) (holding that a detective contradicted *Miranda* warnings by telling the suspect that having a lawyer present would be a "disadvantage" and that "honesty wouldn't hurt him"); United States v. Beale, 921 F.2d 1412, 1435 (11th Cir.1991) (telling an illiterate defendant that signing a waiver form "would not hurt him").

256. Nick Petrić Howe & Benjamin Thompson, *This Isn't the Nature Podcast — How Deepfakes Are Distorting Reality*, NATURE: PODCAST, at 1:47–2:00 (Sept. 27, 2023), https://www.nature.com/articles/d41586-023-03042-1 [https://perma.cc/DNL6-LJH4].

257. 135 F.3d 235 (2d Cir. 1998).

258. *See supra* note 43 and accompanying text.

fake conversation with Tankleff's father was enough for Tankleff to question reality.[259] Imagine now if the police videotaped Tankleff's father in the hospital and manipulated the video and audio to frame Tankleff's father as saying "my son tried to kill me."[260] If the hearsay lie alone was enough make Tankleff question reality, the Nash & Wade and Kassin & Kiechel studies demonstrate that the introduction of fake, demonstrative evidence will increase the likelihood of a suspect to question reality. This is especially true if the suspect has no recollection of the event at all.[261]

The phenomenon in the Nash & Wade and Kassin & Kiechel studies, as well as the *Tankleff* case, help to explain why deepfake technology may heighten the risks of a false confession in other circumstances. For example, the Nash & Wade study demonstrated that when a suspect is confronted with the idea that incriminating evidence will be presented to the professor in charge, the suspect may begin to fabricate memories in an attempt to explain what happened, or accept guilt.[262] This phenomenon occurred at a lower rate with the Tell-Video Subjects[263]—this is akin to a verbal lie by the police—and at a high rate with the See-Video Subjects[264]—this is akin to a deepfake manipulation.[265] From this, the following conclusion can be drawn: if a criminal suspect is presented with video evidence of them committing the crime, or someone close to the suspect implicating them of the crime, the Kassin & Kiechel and Nash & Wade studies find that the suspect will at least "*think* [they] messed up."[266] If a lie can become this powerful to allow a person to assume liability in a crime, that lie should be barred from police use because it soils the purpose

259. *Tankleff*, 135 F.3d at 241.

260. *See supra* note 43 and accompanying text.

261. *See, e.g.*, *Tankleff*, 135 F.3d at 240.

262. Nash & Wade, *supra* note 158, at 630.

263. *Id.* The breakdown for the Tell-Video Subjects is as follows: 100% of subjects signed a confession, 60% experienced full internalization, 7% experienced partial internalization, 33% experienced no internalization, 0% experienced full confabulation, 7% experienced hypothesizing, and 93% experienced no confabulation. *Id.*

264. *Id.* The breakdown for the See-Video Subjects is as follows: 100% of subjects signed a confession, 67% experienced full internalization, 33% experienced partial internalization, 0% experienced no internalization, 7% experienced full confabulation, 7% experienced hypothesizing, and 87% experienced no confabulation. *Id.* This study demonstrates that, when subjects are shown with potentially damaging evidence again them, they will experience at least some internalization. *See id.*

265. *See id.*

266. *Id.* at 629–30. This is from the fact that the See-Video subjects did not experience any internalization. *Id.* The hypothetical question above merely conforms to the line of questioning attributed to the weakest form of internalization, such as "I think I missed up [the] experiment." *Id.*

of the voluntariness test: to avoid procuring false confessions from a suspect's lips.[267]

IV. SOLUTION: INTRODUCING A *PER SE* INVOLUNTARINESS RULE FOR DEEPFAKES AND OTHER FABRICATIONS CREATED BY ARTIFICIAL INTELLIGENCE

The voluntariness test serves three goals: (1) to protect against untrustworthy confessions;[268] (2) to stymie offensive police tactics to secure a confession;[269] and (3) to exclude involuntary-in-fact confessions.[270] Outside of violence, the Supreme Court has been reluctant to issue a bright line rule for when a confession may become involuntary because the Court has yet to classify one standalone tactic as so coercive or repugnant to require the automatic exclusion of a confession.[271] Indeed, "a categorical rule is inconsistent with the multi-factor, holistic approach to assessing voluntariness that . . . the Supreme Court ha[s] endorsed."[272] Several pre-*Miranda* cases had fashioned holdings which created a strong presumption of involuntariness, but none decried a "bright line rule" for any particular police tactic.[273]

---

267. *See, e.g.*, Miranda v. Arizona, 384 U.S. 436, 455–56, 460 (1966) (detailing the constitutional concerns with false confessions based on interrogation proceedings).

268. *See, e.g.*, Ashcraft v. Tennessee, 322 U.S. 143, 154 (1944) (holding that confining a defendant for thirty-six hours without rest or sleep to produce a confession "is so inherently coercive that its very existence is irreconcilable with the possession of mental freedom by a lone suspect against whom its full coercive force is brought to bear"); Spano v. New York, 360 U.S. 315, 320–21 (1959) ("The abhorrence of society to the use of involuntary confessions does not turn alone on their inherent untrustworthiness. It also turns on the deep-rooted feeling that the police must obey the law while enforcing the law; that in the end life and liberty can be as much endangered from illegal methods used to convict those thought to be criminals as from the actual criminals themselves.").

269. *See, e.g.*, Brown v. Mississippi, 297 U.S. 278, 282 (1936) ("[D]efendants were made to strip and they were laid over chairs and their backs were cut to pieces with a leather strap with buckles on it, and they were likewise made by the said deputy definitely to understand that the whipping would be continued unless and until they confessed").

270. *See, e.g.*, Townsend v. Sain, 372 U.S. 293, 309 (1963) ("If the confession which petitioner made . . . was in fact involuntary, the conviction cannot stand.").

271. *See, e.g.*, Schneckloth v. Bustamonte, 412 U.S. 218, 226 (1973) (recognizing that cases concerning involuntary confessions do not "turn[] on the presence or absence of a single controlling criterion" because each factor "reflect[s] a careful scrutiny of all the surrounding circumstances.").

272. United States v. Gonzalez-Garcia, 708 F.3d 682, 688 (8th Cir. 2013).

273. *See, e.g.*, Malinski v. New York, 324 U.S. 401, 405–07 (1945) (concerning the police keeping a defendant naked for several hours, "[i]f the confession had been the product of persistent questioning while Malinski stood stripped and naked, we would have a clear case. But it was not."); Lynumn v. Illinois, 372 U.S. 528, 534 (1963) (concerning the government withholding financial aid for the defendant's child if the defendant failed

If the Court was ever presented with the issues discussed herein, this Article would champion for a ruling that recognizes confessions produced from deepfake technology as *per se* involuntary. However, a more proactive approach is merited to ensure that no suspect endures the coercive power of deepfake technology. Thus, this Article suggests two methods for crafting a *per se* rule of involuntariness outside of the Court's purview. Subpart A details a solution for the strongest constitutional protections through the federal and state legislatures.[274] Subpart B explains a similar solution through local state governments that circumvents the

---

to cooperate, "[i]t is thus abundantly clear that the petitioner's oral confession was made only after the police had told her that state financial aid for her infant children would be cut off, and her children taken from her, if she did not 'cooperate.' . . . We think it clear that a confession made under such circumstances must be deemed not voluntary, but coerced."); Rogers v. Richmond, 365 U.S. 534, 543–44 (1961) (concerning the police pretending to bring in the defendant's wife—who suffered from an illness—for questioning, "[c]oncerning the feigned phone call that petitioner's wife be brought in to headquarters . . . we cannot but conclude that the question whether Rogers' confessions were admissible into evidence was answered by reference to a legal standard which took into account the circumstance of probable truth or falsity. And this is not a permissible standard" (footnote omitted)); Hayes v. Washington, 373 U.S. 503, 513 (1963) (concerning the rejection of the defendant's requests to call his wife or attorney until he cooperated, "[t]he uncontroverted portions of the record thus disclose that the petitioner's written confession was obtained in an atmosphere of substantial coercion and inducement created by statements and actions of state authorities."); Ward v. Texas, 316 U.S. 547, 555 (1942) (concerning the removal of the defendant from the jail to a distance place in an effort to conceal his whereabouts from friends and family, "[t]his Court has set aside convictions based upon confessions extorted from ignorant persons who have been subjected to persistent and protracted questioning, or who have been threatened with mob violence, or who have been unlawfully held incommunicado without advice of friends or counsel, or who have been taken at night to lonely and isolated places for questioning. Any one of these grounds would be sufficient cause for reversal. All of them are to be found in this case" (footnote omitted)); Leyra v. Denno, 347 U.S. 556, 561 (concerning a state employed psychiatrist who was disguised as a general practitioner to provide the defendant with the medical relief he needed, "the undisputed facts in this case are irreconcilable with petitioner's mental freedom 'to confess to or deny a suspected participation in a crime', and the relation of the confessions made to the psychiatrist, the police captain and the state prosecutors, is 'so close that one must [say] the facts of one control the character of the other"); Spano v. New York, 360 U.S. 315, 323 (1959) (concerning a lie that a childhood friend—who was a police officer—would lose his job if the defendant failed to comply, "Bruno's was the one face visible to petitioner in which he could put some trust. There was a bond of friendship between them going back a decade into adolescence. It was with this material that the officers felt that they could overcome petitioner's will. They instructed Bruno falsely to state that petitioner's telephone call had gotten him into trouble, that his job was in jeopardy, and that loss of his job would be disastrous to his three children, his wife and his unborn child. And Bruno played this part of a worried father, harried by his superiors, in not one, but four different acts, the final one lasting an hour. . . . We conclude that petitioner's will was overborne by official pressure, fatigue and sympathy falsely aroused after considering all the facts in their post-indictment setting.").

274. *See infra* Part IV.A.

otherwise arduous legislative process.[275] Subpart C provides support for imposing an unorthodox *per se* rule.[276]

A. Legislative Amendments from Congress or the State Legislature

Under the principles of federalism, the states are entitled to enact their own criminal statutes.[277] However, the Supremacy Clause ensures that the word of the federal government will trump the letter of state law where a conflict exists.[278] This creates fifty-two criminal jurisdictions, comprised of the fifty states, the District of Columbia, and federal law.[279] Notably, though, the protection against self-incrimination under the Fifth Amendment (and Fourteenth Amendment as applicable to the states) serves as the floor—any state law or state constitutional principle may provide greater, but not fewer, rights than those afforded under the United States Constitution.[280]

For an illustration, this Article will analyze federal and New York State law. Besides the Fifth Amendment, the admissibility of confessions under federal law is governed by 18 U.S.C. § 3501, which provides numerous factors when determining whether an admission is considered "voluntary."[281] Besides the Fourteenth Amendment and Article 1, section

---

275. *See infra* Part IV.B.

276. *See infra* Part IV.C.

277. *See, e.g.*, U.S. CONST. amend. X ("The powers not delegated to the United States by the Constitution, nor prohibited by it to the States, are reserved to the States respectively, or to the people.").

278. *See, e.g.*, U.S. CONST. art. VI, cl. 2 ("This Constitution, and the Laws of the United States . . . shall be the supreme Law of the Land; and the Judges in every State shall be bound thereby, any Thing in the Constitution or Laws of any State to the Contrary notwithstanding.").

279. *See, e.g.*, Paul H. Robinson, *Murder Mitigation in the Fifty-Two American Jurisdictions: A Case Study in Doctrinal Interrelation Analysis*, 47 TEX. TECH. L. REV. 19, 20 (2014).

280. U.S. CONST. amend. V; U.S. CONST. amend. XIV; U.S. CONST. art. VI, cl. 2.

281. 18 U.S.C. § 3501. Titled "Admissibility of Confessions," this statute provides:

> (b) The trial judge in determining the issue of voluntariness shall take into consideration all the circumstances surrounding the giving of the confession, including (1) the time elapsing between arrest and arraignment of the defendant making the confession, if it was made after arrest and before arraignment, (2) whether such defendant knew the nature of the offense with which he was charged or of which he was suspected at the time of making the confession, (3) whether or not such defendant was advised or knew that he was not required to make any statement and that any such statement could be used against him, (4) whether or not such defendant had been advised prior to questioning of his right to the assistance of counsel; and (5) whether or not such defendant was without the assistance of counsel when questioned and when giving such confession.

6 of the New York State Constitution, the admissibility of a confession under New York State law is governed by New York Criminal Procedure Law section 60.45, which also details several factors to determine whether a confession is voluntary.[282] Section 60.45 has also been interpreted as a bright line rule against statements made in reliance on a promise that creates a "substantial risk that the defendant might falsely incriminate himself."[283]

To protect against the use of deepfakes in interrogations, 18 U.S.C. § 3501(b) should be amended to include the following underlined provision:

> The presence or absence of any of the above-mentioned factors to be taken into consideration by the judge need not be conclusive on the issue of voluntariness of the confession.

Id.

282. N.Y. CRIM. PROC. LAW § 60.45 (McKinney 2021). Titled "Rules of Evidence; Admissibility of Statements of Defendants," this statute provides:

> [1] Evidence of a written or oral confession, admission, or other statement made by a defendant with respect to his participation or lack of participation in the offense charged, may not be received in evidence against him in a criminal proceeding if such statement was involuntarily made.
>
> [2] A confession, admission or other statement is "involuntarily made" by a defendant when it is obtained from him:
>
> (a) By any person by the use or threatened use of physical force upon the defendant or another person, or by means of any other improper conduct or undue pressure which impaired the defendant's physical or mental condition to the extent of undermining his ability to make a choice whether or not to make a statement; or
>
> (b) By a public servant engaged in law enforcement activity or by a person then acting under his direction or in cooperation with him:
>
> (i) by means of any promise or statement of fact, which promise or statement creates a substantial risk that the defendant might falsely incriminate himself; or
>
> (ii) in violation of such rights as the defendant may derive from the constitution of this state or of the United States.

Id.

283. People v. Thomas, 8 N.E.3d 308, 315 (N.Y. 2014) ("It is true that our state statute . . . treats as "involuntarily made" a statement elicited "by means of any promise or statement of fact, which promise or statement creates a substantial risk that the defendant might falsely incriminate himself" (internal citations omitted)); *see also* People v. Brown, 474 N.Y.S.2d 927, 929–31 (N.Y. App. Div. 1984) (citing People v. Vail, 457 N.Y.S.2d 933, 933-34 (N.Y. App. Div. 1982)) (relying on U.S. Supreme Court precedent in Bram v. United States, 168 U.S. 532 (1897) to support the proposition that CRIM. PROC. LAW § 60.45 codified the *Bram* rule as a *per se* bar to promises used in a confession). *But see* Arizona v. Fulminante, 499 U.S. 279, 285 (1991) (distinguishing Bram v. United States, 168 U.S. 532 (1897) ("Although the Court noted in *Bram* that a confession cannot be obtained by "'any direct or implied promises, however slight, nor by the exertion of any improper influence,'" . . . it is clear that this passage from *Bram* . . . does not state the [current] standard for determining the voluntariness of a confession") (internal citations omitted)).

> The presence or absence of any of the above-mentioned factors to be taken into consideration by the judge need not be conclusive on the issue of voluntariness of the confession, but the use of artificial intelligence during any interrogation is strictly prohibited.

Additionally, New York Criminal Procedure Law section 60.45(2)(a), and other similar state statutes, should be amended to include the following underlined portion:

> [2] A confession, admission or other statement is "involuntarily made" by a defendant when it is obtained from him:
>
> (a) By any person by the use or threatened use of physical force upon the defendant or another person, by the use of artificial intelligence, or by means of any other improper conduct or undue pressure which impaired the defendant's physical or mental condition to the extent of undermining his ability to make a choice whether or not to make a statement; . . .

The addition of these succinct, yet broad clauses in each statute will provide for strong protections against deepfake technology and other advancements in artificial intelligence. A precautionary statutory bar will ensure that no criminal suspect is subject to the power of deepfake technology, but is not sweeping enough to swallow the voluntariness test. Additionally, this Article does not mean to suggest that any new advancement in technology should be granted an exception to the voluntariness test; here, however, the scientific evidence demonstrates the coercive power of deepfake technology to merit an exception.[284]

B. A Change in Local Government

Local government policy can expedite change and avoid the potential roadblocks frequently endured through typical statutory amendments. ABA Prosecution Standard 3-6.6 provides several principles for the presentation of evidence by the prosecutor.[285] These standards serve as

---

284. *See supra* Part II.B.2.

285. CRIM. JUSTICE STANDARDS FOR THE PROSECUTION FUNCTION Standard 3-6.6(d) (Am. Bar Ass'n 4th ed. 2017), https://www.americanbar.org/groups/criminal_justice/standards/ProsecutionFunctionFourthEdition [https://perma.cc/Y9AJ-BHXD]. Titled "Presentation of Evidence," this standard provides:

> (d) The prosecutor should not bring to the attention of the trier of fact matters that the prosecutor knows to be inadmissible, whether by offering or displaying

"best practices" for prosecutors, but will not "serve as the basis for the imposition of professional discipline.[286] Thus, while the ABA Standards are not binding, they provide guidance for the internal practices of district attorney's offices in the performance of certain functions.[287] To usher a change in interrogation practice, Prosecution Standard 3-6.6(d) should be amended to read as follows:

> The prosecutor should not bring to the attention of the trier of fact matters that the prosecutor knows to be inadmissible, whether by offering or displaying inadmissible evidence, asking legally objectionable questions, or making impermissible comments or arguments. Additionally, the prosecutor should not present evidence of any confession obtained through the use of artificial intelligence. If the prosecutor is uncertain about the admissibility of evidence, the prosecutor should seek and obtain resolution from the court before the hearing or trial if possible, and reasonably in advance of the time for proffering the evidence before a jury.[288]

Again, this succinct, yet broad clause in the ABA Standards can provide a workable guide for district attorneys when publishing policies for their office. Prosecution policy guidelines routinely detail pressing issues presented before a local office that the district attorney seeks to ameliorate.[289] As an exemplary provision for a newly elected or appointed district attorney, their office should adopt the following policy:

> A. PRESENTATION OF CONFESSION EVIDENCE
>
> 1. The Office will not present, as evidence of an element of a crime, a confession produced by means of artificial intelligence.

> inadmissible evidence, asking legally objectionable questions, or making impermissible comments or arguments. If the prosecutor is uncertain about the admissibility of evidence, the prosecutor should seek and obtain resolution from the court before the hearing or trial if possible, and reasonably in advance of the time for proffering the evidence before a jury.

Id.

286. *Id.* § 3-1.1(b).

287. *Id.*

288. *Id.* § 3-6.6(d).

289. Letter from Alvin L. Bragg, Jr., District Attorney, County of New York, to All Staff 1–2 (Jan. 3, 2022) ("Day One Letter Policies"), https://www.manhattanda.org/wp-content/uploads/2022/01/Day-One-Letter-Policies-1.03.2022.pdf [https://perma.cc/5T7R-U2G6].

> 2. The Office will not present, as impeachment evidence, a confession produced by means of artificial intelligence.
>
> 3. Where a confession has been procured by means of artificial intelligence or other inherently deceptive technological advances, the following rules apply:
>
> a) The Office will present such confession evidence to defense counsel;
>
> b) Safeguard any documentation of the confession; and
>
> c) Base the Office's case-in-chief solely on the understanding that such a confession is inadmissible under the laws of the United States Constitution.
>
> 4. If the Office is uncertain about the admissibility of a confession, the Office shall seek and obtain guidance from an independent group within the Office's Conviction Integrity Unit or, if such is unavailable, an independent group within the State's Committee on Professional Ethics.[290]

While blanket policies—typically, non-enforcement policies—have been scrutinized by scholars,[291] the above blanket policy does not completely bar the district attorney from bringing a charge, but merely pledges against the admission of particular evidence from the government's case-in-chief or use as impeachment evidence. The drawback to this solution though is that this policy is merely a policy—it would require self-enforcement by the district attorney's office, subject only to the will of the voter or distaste of the governor.[292] Regardless, even if this small change presents trouble in enforcement, any change in the current standard will lay the necessary groundwork for future incremental change by putting the public on notice that deepfake induced confessions will not be tolerated.

---

290. Memorandum from Alvin L. Bragg, Jr., District Attorney, County of New York, on Day One Polices & Procedures 1–5 (Jan. 3, 2022) ("Policy and Procedure Memorandum"), https://www.manhattanda.org/wp-content/uploads/2022/01/Day-One-Letter-Policies-1.03.2022.pdf [https://perma.cc/5T7R-U2G6].

291. *See, e.g.*, Zachary Price, *Blanket Nonenforcement Policies Are Unconstitutional in California*, SCOCA BLOG (Feb. 1, 2022), https://scocablog.com/616-2 [https://perma.cc/EW59-LMV3].

292. *See, e.g.*, N.Y. CONST. art. XIII, § 13 ("In each county a district attorney shall be chosen by the electors once in every three or four years as the legislature shall direct.").

C. Why an Involuntary Per Se Rule for Deepfake Produced Confessions is Necessary

The voluntariness test rests on a policy that confessions themselves may be false if given involuntarily, which supports exclusion of such evidence.[293] A *per se* rule barring deepfake confessions would avoid a rigorous, fact intensive analysis typically required by the court.[294] Additionally, a policy goal set by the district attorney's office would create local, self-enforcing rules that perform the function of the Fifth or Fourteenth Amendment exclusionary rule. A policy provision from a district attorney's office that excludes evidence from the government's case-in-chief or proffered for impeachment will prevent the jury from misusing a deepfake induced confession. While a prior statement is admissible for impeachment even if the statement was procured in violation of *Miranda*, a prior statement would be inadmissible if it was involuntarily produced.[295] Hence, the local bar on deepfake confessions for both substantive and impeachment purposes would solidify this goal. Moreover, the ethical rules, while unavailable to broker fines, would help to police internal policies and serve as a necessary benchmark.[296]

Working together, both the district attorney's office memorandum and the ABA Prosecution Standards would ensure that if the constitution or Congress could not deter the police from using deepfake technology, deterrence could at least spawn from the local government and serve as a "quasi-exclusionary rule."[297]

V. CONCLUSION

Advancements in technology beg the question: what amount of technology will be too much for the Due Process Clause to protect against? Deepfake technology is certainly short of telepathy,[298] but more powerful than a standard lie by the police. The voluntariness test should be continuously scrutinized as technology advances. The proposals above

---

293. Miranda v. Arizona, 384 U.S. 436, 455–56 (1966) (detailing the constitutional concerns with false confessions based on interrogation proceedings).

294. *See, e.g.*, *id*; 18 U.S.C. § 3501.

295. *See* Harris v. New York, 401 U.S. 222, 226 (1971).

296. *See* CRIM. JUSTICE STANDARDS FOR THE PROSECUTION FUNCTION Standard 3-6.6(d).

297. *See, e.g.*, Olmstead v. United States, 277 U.S. 438, 462 (1928), *rev'd on other grounds*, Katz v. United States, 389 U.S. 347 (1967) (ascribing the exclusionary rule applicable to the Fourth Amendment from *Weeks* to the Fifth Amendment).

298. I dedicate this footnote to Professor Fred Klein, Professor of Law at the Maurice A. Deane School of Law at Hofstra University. When asked, "what future technology do you believe would render a confession involuntary," his response was "some form of mindreading."

call for legislative assistance to ensure that the rights of individuals are not placed in jeopardy as technology advances. The strongest levels of protection will derive from the courts and federal or state legislatures, followed by local implementations of blanket exclusionary rules as part of a district attorney's policies.

To avoid having an individual subject to the coercive powers of deepfake technology, this Article champions for a proactive ban on using deepfake technology to induce confessions. This approach would inform ongoing discussions about the impact of technology on voluntariness, while a more reactive approach would leave individuals vulnerable to the next wave of potentially coercive technological advancements. Legal and ethical standards that squarely address the impact of deepfakes could pave the way before these technologies are widely misused. Proactivity is paramount to offer crucial safeguards for individual rights and to ensure the criminal legal system maintains its integrity in the digital age.