CNS

OCCASIONAL PAPER

#29 · JULY 2017

The Verification Clearinghouse: Debunking Websites and the Potential for Public Nonproliferation Monitoring

Bryan Lee and Kyle Pilutti

Middlebury Institute *of* International Studies at Monterey

James Martin Center for Nonproliferation Studies

The views, assessments, judgments, and conclusions in this report are the sole representations of the authors and do not necessarily represent either the official position or policy or bear the endorsement of the James Martin Center for Nonproliferation Studies, the Middlebury Institute of International Studies at Monterey, the President and Trustees of Middlebury College, or the US Department of State.

Acknowledgements:

The authors would like to thank Francisco Parada for his assistance in the research and analytical efforts for the initial phases of this paper.

James Martin Center For Nonproliferation Studies
www.nonproliferation.org

The James Martin Center for Nonproliferation Studies (CNS) strives to combat the spread of weapons of mass destruction by training the next generation of nonproliferation specialists and disseminating timely information and analysis. CNS at the Middlebury Institute of International Studies at Monterey is the largest nongovernmental organization in the United States devoted exclusively to research and training on nonproliferation issues.

Middlebury Institute of International Studies at Monterey
www.miis.edu

The Middlebury Institute of International Studies at Monterey, a graduate school of Middlebury College, provides international professional education in areas of critical importance to a rapidly changing global community, including international policy and management, translation and interpretation, language teaching, sustainable development, and nonproliferation. We prepare students from all over the world to make a meaningful impact in their chosen fields through degree programs characterized by immersive and collaborative learning and opportunities to acquire and apply practical professional skills. Our students are emerging leaders capable of bridging cultural, organizational, and language divides to produce sustainable, equitable solutions to a variety of global challenges.

James Martin Center for Nonproliferation Studies
Middlebury Institute of International Studies at Monterey
460 Pierce St., Monterey, CA 93940, U.S.A.
Tel: +1 (831) 647-4154
Fax: +1 (831) 647-3519

CONTENTS

EXECUTIVE SUMMARY

The rapid pace at which information spreads online has been a boon for information gathering but poses risks with respect to the truth of the information. This risk increases with high-stakes information, such as treaty compliance judgements, and it is one of the main reasons why publicly available information is usually excluded from verification efforts.

Nevertheless, just as military and intelligence experts have developed techniques to counter deception, techniques to debunk online deception have grown in power and sophistication. Most of these techniques can be applied by a dedicated amateur and allow the user to identify manipulations of images, video, and audio. Coupled with the rise of open source analytical tools, there is now the potential to construct a "verification clearinghouse," an online website that would allow both experts and members of the public to monitor and evaluate arms control treaties. There are a number of existing online debunking sites available to the public. The majority of them focus on smaller niche topics, such as sites dedicated to political fact checking or Internet hoaxes. In other cases, sites are maintained by enthusiastic experts who work to debunk cases that are of particular interest or importance to themselves or their fields.

These sites are important building blocks for future verification clearinghouse efforts. In our analysis, we characterize the existing attempts to address online deception, identifying the factors that make them particularly advantageous, and group them into three different types as shown below.

Types of Debunking Websites

Academic	Bulk	Explanative: Global, Regional, Topical
Debunking sites that are focused primarily on isolated cases and which offer a high degree of analytical rigor.	Debunking sites that focus on addressing as many classes of possible deception as possible.	Debunking sites that focus on providing detailed cases that focus on the explanation of why or why not a case is deceptive. These sites can be separated into those focused on regional, topical, or global issues.

Several technical and policy hurdles will need to be more fully resolved before a verification clearinghouse can be entirely implemented. Among the technical challenges are the difficulties of motivating participation, protecting against cheating, and the deception/counter-deception "arms race." Policy challenges principally revolve around the ethical, legal, and social implications or "ELSI" concerns. Both sets of challenges can be overcome, and the concept of a verification clearinghouse as one tool in a broader arms control and nonproliferation toolkit is worthy of exploration.

INTRODUCTION

The availability of online information has ushered in a new era of awareness about international events. People no longer have to rely on news agencies or government officials to hear about national security events occurring in foreign countries. They also no longer have to wait for information about these events, as social media applications and crowdsourced news sites provide unfiltered updates in real time. Much of the world, for example, learned of the 2015 Paris theater attack from eyewitness reports shared on Facebook and Twitter, not from the traditional news outlets or the Parisian authorities.[1]

Besides sharing news of world events, online information is being increasingly used to analyze them. For example, YouTube users posted videos of apparent chemical weapons attack victims in Aleppo with the explosive claim that the attacks were carried out by Syrian government forces. At the time, there were no reporters on the ground, and there were no credible government officials able to confirm or deny the claim. This did not deter concerned members of the academic and nongovernmental organization communities from conducting their own analysis.[2] Researchers at Human Rights Watch went through the video evidence to evaluate medical symptoms, as well as rocket debris, to reach the conclusion that chemical weapons were indeed used.[3]

Such methods of analysis might be expected from large, well-connected, and established human rights organizations or research universities, but much smaller organizations have done similar things. In early 2015, a researcher at the James Martin Center for Nonproliferation Studies came across photographs of North Korean president Kim Jong Un visiting a laboratory facility in his country. Aware of his recent statements regarding the Democratic People's Republic of Korea's nascent bioweapons production capability, she scrutinized laboratory equipment shown in the background of the photographs. Upon closer analysis, she determined that the equipment shown was dual-use, suitable both for the laboratory's stated purpose of producing pesticide and for the production of *Bacillus anthracis*.[4] Again, her findings led to international attention, but this time on the part of the North Korean authorities who claimed her analysis was without merit because she was a "trickster and riffraff."[5]

Competent analysis of open source information is not limited to teams and organizations. In 2014, following the downing of Malaysian airlines flight MH-17 over Ukraine, investigative journalist

1 Samuel Gibbs, "Facebook's Safety Check Leads Technology's Support of Paris," *The Guardian*, November 16, 2015, sec. Technology, <https://www.theguardian.com/technology/2015/nov/16/facebook-safety-check-technology-paris-terrorist-attacks>.

2 William J. Broad, "Rockets in Syrian Attack Carried Large Payload of Gas, Experts Say," *The New York Times*, September 4, 2013, sec. World / Middle East, <http://www.nytimes.com/2013/09/05/world/middleeast/rockets-in-syrian-attack-carried-large-payload-of-gas-experts-say.html>.

3 "Syria: Government Likely Culprit in Chemical Attack," Human Rights Watch, accessed November 19, 2016, <https://www.hrw.org/news/2013/09/10/syria-government-likely-culprit-chemical-attack>.

4 Lizzie Dearden, "North Korea 'Could Produce Military-Size Batches of Anthrax,'" *The Independent*, July 16, 2015, <http://www.independent.co.uk/news/world/asia/north-korea-could-produce-military-size-batches-of-anthrax-at-pesticide-factory-researcher-claims-10394624.html>.

5 M. Hanham, personal communication, July 13, 2015.

Elliot Higgins posted an entire series of photographic analyses on his website, Bellingcat.com. His analyses used a series of photographs gathered from witnesses on the ground and official sources. It painstakingly presented detailed evidence which demonstrated that the attack came from Russian-occupied territory in Ukraine, was carried out with a Russian missile, and was likely conducted by Russian separatists.[6] His analysis and general conclusions were virtually identical to the ones made independently by Dutch authorities in 2015 and published in their official investigation report.[7]

Many of these cases touching on international security receive prominent coverage in the expert community from analysts and journalists, but there is a dark side to the ubiquity of online information. For every carefully documented online case providing detailed open source evidence of war crimes or egregious state-sponsored behavior, there are dozens documenting, with apparent equal fidelity, the existence of mythical land creatures, alien abductions, and "deep-State" conspiracy theories.[8] While many of these efforts are easy for a skeptical observer to dismiss out of hand, many remain highly convincing. Even more troubling are state-sponsored efforts at online deception, which often combine sophisticated deception techniques with well-funded and coordinated propaganda campaigns.[9]

It does not take much imagination to envision a scenario where false or misleading online information can lead to tragic results. In the security arena, where tensions may already be high and actors have access to deadly force, the effects could be catastrophic.

Fortunately, rigorous verification of online information is possible. There are numerous, well-documented forensic techniques to identify false or misleading online information. Further, just as sophisticated analytical techniques are increasingly utilized outside of government intelligence agencies, the techniques of information forensics are also moving beyond the domain of law-enforcement and information technology experts. When these are combined with the powers of social media, interactive web technologies, and the crowd, or "collective intelligence," new opportunities arise for individuals and organizations who are dedicated to establishing the ground truth.

It comes as no surprise that fact-checking websites have grown almost as quickly as the sensationalist news stories, gossipy headlines, and Internet hoaxes that seem to dominate our inboxes and social media newsfeeds. Many of these websites employ similar techniques in their analytical efforts to those mentioned earlier. Sponsors of these sites range from news organizations, advocacy groups and political campaigns to skeptical individuals.

---

6 Bellingcat, "MH17 - The Open Source Investigation, Two Years Later," *Bellingcat*, July 15, 2016, <https://www.bellingcat.com/news/uk-and-europe/2016/07/15/mh17-the-open-source-investigation-two-years-later/>.

7 "Investigation Crash MH17, 17 July 2014," Dutch Safety Board, July 17, 2014, <https://www.onderzoeksraad.nl/en/onderzoek/2049/investigation-crash-mh17-17-july-2014/publicatie/1658/dutch-safety-board-buk-surface-to-air-missile-system-caused-mh17-crash#fasen>.

8 *Top Paranormal Sites.com*, 2016, <http://www.topparanormalsites.com>/.

9 Tom Parfitt, "My Life as a pro-Putin Propagandist in Russia's Shadowy 'Troll Factory'; Journalist Poses as Housewife to Expose Firm That Pays Army of Bloggers to Spew Vitriol about the Kremlin's Critics," *The Sunday Telegraph (London)*, June 7, 2015.

What is surprising, however, is the relative lack of online fact-checking efforts in the national security arena. Such efforts are particularly important for arms control and nonproliferation policy because so much of the effort in this arena is rooted in transparency and trust. Moreover, publicly available technologies such as commercial satellite imagery are already being used by private organizations and citizens to monitor proliferation activities; however, there is wide variation in the expertise and quality of these efforts.[10]

It has been demonstrated that reliable techniques to verify online information are widely available and already utilized by private citizens. In addition, fact-checking websites are commonplace. Could the arms control community benefit from its own debunking effort? Might there be a possibility of a verification clearinghouse?

The remainder of this paper is an effort to answer that question. The beginning of this paper looks at the overall problem of deception, both online and off, as well as the subtopics of rumors and misinformation. Following this, we offer a detailed description of online tampering and tampering detection techniques. Next, we introduce the concept of a verification clearinghouse and demonstrate how online debunking websites function in general. Last, we explore the some of the technical and policy challenges involved in using verification clearinghouses as a tool for arms control and nonproliferation.

THE PROBLEM OF DECEPTION

Deception as a type of human interaction exists in some of the oldest writings on record, and many have concluded that it is an inherent behavior.[11] Despite its apparent evolutionary advantage, however, deception in the realm of online information is frequently harmful. In his study on the topic, Bowyer Bell defined deception as "the conscious, planned intrusion of an illusion seeking to alter a target's reality, replacing objective reality with perceived reality."[12] The methods to alter the perceived reality may differ, but the key point is that the deceiver always does so with explicit intent as part of a conscious process.[13]

10 Jeffrey Lewis, "Collected Thoughts on Phil Karber," Arms Control Wonk, December 7, 2011, <http://www.armscontrolwonk.com/archive/204799/collected-thoughts-on-phil-karber/>.

11 Charles F. Bond Jr and Michael Robinson, "The Evolution of Deception," *Journal of Nonverbal Behavior*, Vol. 12, No. 4 (1988): 295–307, doi:10.1007/BF00987597.

12 J. Bowyer Bell, "Toward a Theory of Deception," *International Journal of Intelligence and Counterintelligence*, Vol. 16, No. 2 (2003), pp. 244–279, doi:10.1080/08850600390198742.

13 Megan Wise and Dariela Rodriguez, "Detecting Deceptive Communication Through Computer-Mediated Technology: Applying Interpersonal Deception Theory to Texting Behavior," *Communication Research Reports*, Vol. 30, No. 4 (October 22, 2013), pp. 342–46, doi:10.1080/08824096.2013.823861; D. B Buller, J. B Stiff, and J. K Burgoon, "Behavioral Adaptation in Deceptive Transactions Fact or Fiction: Reply to Levine and McCornack," *Human Communication Research*, Vol. 22, No.4 (June, 1996), doi: 10.1111/j.1468-2958.1996.tb00381.x.

Despite ongoing research into online deception, a lack of standard methods and theory of deception pose a challenge.[14] Bell's deception cycle, however, is a useful point of departure.[15] Like all intentional acts, deception starts with planning, determining both the goal of the deception and the method. From there, the deceiver must construct a "ruse" and select an appropriate "channel." For example, in the online world, the ruse may be a fake persona and the channel could be an online dating site. A decision point follows, where the deceiver monitors to see if the illusion is accepted and what the response is. The deceiver then analyzes the response feedback and decides how to respond in turn.

Research supports the common-sense assertion that the Internet makes deception easier because any system designed to process information is also designed to manipulate it.[16] The Internet also makes countering deception difficult because of its support for anonymity. As Michael Tsikerdekis points out, even something as drastic as a complete change of gender can be done through simply typing in a new name.[17] Because of the range of possible motivations for online deception and the fundamental structure of the system, it remains unlikely that the risk of deception can ever be "engineered away."[18] Therefore, users of online systems should always be aware of the potential for information to be false or intentionally misleading.

Returning to Bell's framework, online deception in the planning stage begins with the goal and the method. In the online world, the goal is typically to deceive a particular individual. This may be termed personal deception and is the most likely to succeed, probably because individuals are less careful or deliberate in evaluating transactions than organizations.[19] This type of deception is frequently initiated over social media or direct e-mail connection, and the effects are usually limited to the target.

When the deceiver moves from targeting individuals to targeting groups, the deception methods necessarily become more complex. By spreading misinformation, one can manipulate the perception of a larger crowd. George Cybenko terms this "cognitive hacking" and identifies covert and overt methods for its perpetration.[20] Most forms of online deception depend on the spread of false

14 Avner Caspi and Paul Gorsky, "Online Deception: Prevalence, Motivation, and Emotion," *CyberPsychology & Behavior*, Vol. 9, No. 1 (February 23, 2006), pp. 54–59, doi:10.1089/cpb.2006.9.54; Jeffrey T. Hancock, "Digital Deception," 2007, <https://books.google.com/books?hl=en&lr=&id=zyhkAfDb_i4C&oi=fnd&pg=PA289&dq=related:aCCRSBQAwsYJ:scholar.google.com/&ots=oba21BgAon&sig=u4s2om51ZWQdr3Z_g-fwQYTJbXE>; Michelle Drouin et al., "Why Do People Lie online?' 'Because Everyone Lies on the Internet,'" *Computers in Human Behavior*, No. 64 (November 2016), pp. 134–142, doi:10.1016/j.chb.2016.06.052; Michail Tsikerdekis and Sherali Zeadally, "Online Deception in Social Media," *Communications of the ACM*, Vol. 57, No. 9 (September 2014), pp. 72–80, doi:10.1145/2629612.

15 Bell, "Toward a Theory of Deception."

16 Erik Gartzke and Jon Lindsay, "Weaving Tangled Webs: Offense, Defense, and Deception in Cyberspace," *Security Studies* Vol. 24, No. 2 (n.d.), pp. 316–48, doi:10.1080/09636412.2015.1038188.

17 Tsikerdekis and Zeadally, "Online Deception in Social Media," p. 3.

18 Gartzke and Lindsay, "Weaving Tangled Webs: Offense, Defense, and Deception in Cyberspace," p. 327.

19 Xinran Chen and Sei-Chen Joanna Sin, "'Misinformation? What of It?' Motivations and Individual Differences in Misinformation Sharing on Social Media," *Proceedings of the Association for Information Science and Technology*, Vol. 50, No.1 (2013), doi:10.1002/meet.14505001102, <http://onlinelibrary.wiley.com/doi/10.1002/meet.14505001102/abstract>.

20 George Cybenko, Annarita Giani, and Paul Thompson, "Cognitive Hacking: A Battle for the Mind," *Computer*, Vol. 35, No. 8 (August 1, 2002), pp. 50–56, doi:10.1109/MC.2002.1023788.

information in order to alter the target's perceived reality. As is the case with personal deception, this is commonly done via social media through the spreading of misinformation or rumors.

The definitive study of rumors was conducted in the mid-1940s in the context of World War II and remains widely uncontested. In this study, Gordon Allport and Leo Postman explored the social utility and rationale behind the spreading of rumors while examining both wartime and societal rumors. They defined a rumor as a "specific (or topical) proposition for belief, passed along from person to person, usually by word-of-mouth, *without secure standards of evidence being present*" (emphasis added). In addition, they identified "importance" and "ambiguity" as two key conditions that facilitate the creation and spread of rumors.[21] Finally, Allport and Postman believed that as rumors are transmitted, they become: *leveled, sharpened, and assimilated.* By this, they meant, rumors come to be "shorter and concise," with a "selective perception, retention, and reporting of a limited number of details," as they become more attractive and interesting to the listener.[22]

In combatting deception, it is helpful to remember the distinction between misinformation, or "information that has been shown to be inaccurate,"[23] and disinformation, which is "often the product of a carefully planned and technically sophisticated deceit."[24] Although dissemination of misinformation is commonly used in deception, its presence does not inherently imply the intent to deceive.

One of the most serious challenges of addressing online disinformation is identifying its source. Current approaches rely heavily on network analytic methods and are likely too sophisticated to be of much use to the casual user.[25] Moreover, clever deceivers do not necessarily originate or disseminate the misleading information.[26] With the accumulation of evidence of sophisticated state-directed online deception and hacking efforts,[27] source identification is likely to remain problematic.

21 Gordon W. Allport and Leo Postman, *The Psychology of Rumor* (New York: Henry Holt and Company, 1947), ix.
22 Ibid, 75, 86.
23 Chen and Joanna Sin, "'Misinformation? What of It?' Motivations and Individual Differences in Misinformation Sharing on Social Media," p. 1.
24 Don Fallis, "A Conceptual Analysis of Disinformation," Conference Paper, (February 28, 2009), p. 2, <https://www.ideals.illinois.edu/handle/2142/15205>.
25 Devavrat Shah and Tauhid Zaman, "Rumors in a Network: Who's the Culprit?," *IEEE Transactions on Information Theory*, Vol. 57, No. 8 (July 25, 2011), pp. 5163–5181, doi:10.1109/TIT.2011.2158885; Pedro C. Pinto, Patrick Thiran, and Martin Vetterli, "Locating the Source of Diffusion in Large-Scale Networks," *Physical Review Letters*, Vol. 109, No. 6 (August 10, 2012); Vahed Qazvinian et al., "Rumor Has It: Identifying Misinformation in Microblogs," *EMNLP '11 Proceedings of the Conference on Empirical Methods in Natural Language Processing*, (July 27, 2011), pp. 1589–99.
26 Fallis, "A Conceptual Analysis of Disinformation."
27 Ellen Nakashima, "U.S. Decides against Publicly Blaming China for Data Hack," *The Washington Post*, July 21, 2015; Julian Lindley-French, "NATO: Countering Strategic Maskirovka," Canadian Global Affairs Institute, May 2015, <http://www.cgai.ca/nato_countering_strategic_maskirovka>; Maria Snegovaya, "Putin's Information Warfare In Ukraine: Soviet Origins of Russia's Hybrid Warfare," September 2015, <http://understandingwar.org/sites/default/files/Russian%20Report%201%20Putin's%20Information%20Warfare%20in%20Ukraine-%20Soviet%20Origins%20of%20Russias%20Hybrid%20Warfare.pdf>.

Some researchers have argued for a self-correcting effect on social media because of the ability for rumors to be scrutinized by the online community.[28] Others have found that not to be the case and have shown that rumors and deliberately misleading information follow a typical threshold model pattern of information diffusion.[29] Critically, the pattern of information spread was not noticeably different between true and false information. In other words, on social media "rumors gradually acquire more credibility as more and more network neighbors acquire them."[30] This threshold is important for the spread of misinformation; once crossed, it is generally believed to be true by the community.

In summary, deception can be understood as an intentional effort to alter a target's perception of reality. All deception follows an iterative cycle of planning, targeting, and evaluating. In addition, it may be conducted online, offline, or using some combination of the two. Deception takes advantage of the power of rumor to spread both misinformation (wrong information) and disinformation (false information) and may be difficult to detect online because of ambiguities related to source, motive, and analytical method. Finally, individuals, organizations, and states practice deception, and all have been active in using the Internet to further their deceptive efforts.

DETECTING DECEPTION

While the universal nature of deception makes it hard to detect in general, there are numerous techniques available to detect a particular instance of deception. Manipulation of online information is described as tampering, and the field of information forensics has developed and refined several methods of tampering detection. As will be seen in the survey below, each has strengths and weaknesses. Taken as a whole, however, identifying whether a specific image or video has been altered is within the technical capability of virtually any determined analyst.

IMAGES

Discussions of image forensics usually make a distinction between image enhancement and image manipulation. Image enhancement is typically benign and includes such things as increasing contrast, changing brightness, or adjusting the hue or color. Still, in some circumstances, image enhancements can be used to deceive. A good example of unintentional deception was the presentation of labeled satellite imagery photos to the United Nations in the run up to the Second Gulf War (Figure 1).

---

28 Marcelo Mendoza, Barbara Poblete, and Carlos Castillo, "Twitter under Crisis: Can We Trust What We RT?," in *Proceedings of the First Workshop on Social Media Analytics* (ACM, 2010), pp. 71–79, <http://dl.acm.org/citation.cfm?id=1964869>.

29 Kate Starbird et al., "Rumor, False Flags, and Digital Vigilantes: Misinformation on Twitter after the 2013 Boston Marathon Bombing," iSchools, iConference 2014 Proceedings, January, 3, 2014, <https://www.ideals.illinois.edu/handle/2142/47257>; Jacob Ratkiewicz et al., "Detecting and Tracking Political Abuse in Social Media," in *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media* (AAAI Publications, 2011), <http://www.aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/view/2850>.

30 Jacob Ratkiewicz et al., "Detecting and Tracking Political Abuse in Social Media," 1.

The photographs themselves are blurry and indistinct, but with the addition of text labels, untrained observers were quick to accept the identification of possible weapons facilities at face value.[31]

Figure 1. Colin Powell UN Presentation 2003 (U.S. Department of State)
Source: http://opinionator.blogs.nytimes.com/2008/08/11/photography-as-a-weapon/

A much more common type of deception with images is image manipulation. This includes three types of actions: composition or splicing; retouching, healing, or cloning; and content embedding or steganography.[32] We will not be discussing steganography, or hiding additional information in images, because it requires special software to decode the hidden image, and we are examining image content as it is perceived by a typical website user.

Composition or splicing manipulations consist of merging one image with one or more parts of other images in order to create the illusion of one complete image.[33] These splicing manipulations are sometimes also known as *copy-move.* This type of manipulation is extremely common on the Internet and forms the basis of innumerable prank photographs and Internet memes. It is also surprisingly common on the national security stage as states try to embellish military or technical capabilities (Figure 2).

31 Errol Morris, "Photography as a Weapon - The New York Times," *Opinionator*, August 11, 2008, <http://opinionator.blogs.nytimes.com/2008/08/11/photography-as-a-weapon>/.
32 Anderson Rocha et al., "Vision of the Unseen: Current Trends and Challenges in Digital Image and Video Forensics," *ACM Computing Surveys (CSUR)* Vol. 43, No. 4 (2011), p. 26.
33 Ibid.

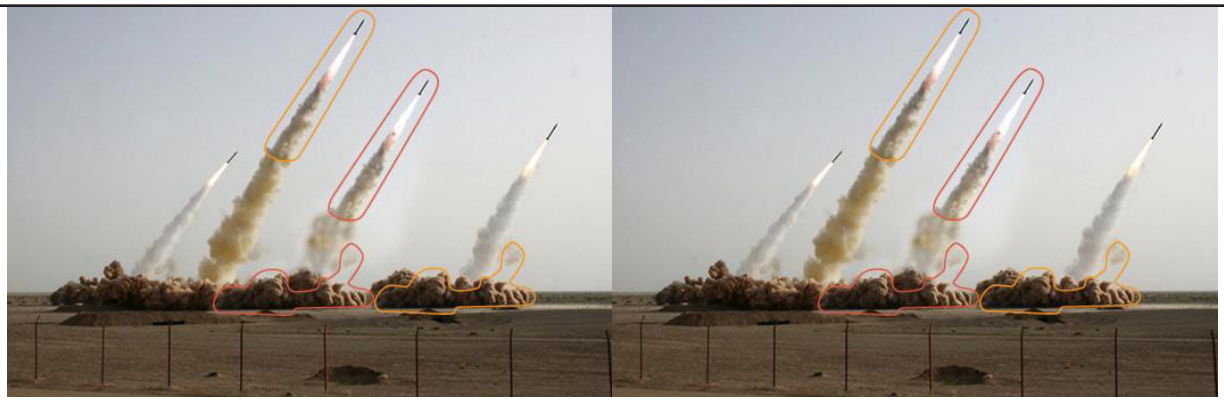

Figure 2. Duplicated missiles and dust clouds in Iranian test launch photograph identified in a New York Times analysis

Source: http://thelede.blogs.nytimes.com/2008/07/10/in-an-iranian-image-a-missile-too-many/

Splicing manipulations are used to add information to an image, but experience has long shown that deleting information can be just as powerful. This is done using retouching, healing, or cloning, but may be better known under the film-based term of *airbrushing*. In a cloning manipulation, an object or person is first cut out of the scene in question, then background imagery is duplicated and pasted into the empty space. Finally, the seams or overlaps are retouched or healed to hide evidence of the manipulation (Figure 3).

Figure 3. Chinese Standing Committee Member in original photo (left) and after cloning manipulation (right)

Source: http://english.chosun.com/site/data/html_dir/2016/01/11/20160111101791.html

Although the variety and types of manipulated images sometimes seems infinite, recent surveys show that image tampering detection techniques fall into five broad categories.[34] The first two of these are based on how the images are interpreted by the display software. These are pixel-based techniques and format-based techniques. The other three categories are related to how the images are captured and represented in the real world. These are camera-based techniques, light source techniques, and geometric measurement techniques.

VIDEO

The enormous audience of video sharing sites such as YouTube and Vimeo mean that static images are no longer the only manipulation concern. Like still images, video can be manipulated in several ways, and the impact can be even more powerful. It was a falsely identified video, for example, that helped inflame the 2013 riots in the Indian city of Muzaffarnagar that killed sixty-three people and displaced forty thousand.[35]

Leaving aside intentional misattribution (see Figure 4), which seems to have been the case in India, video tampering typically is related to the broader form of video itself. At its most basic, video is a timed sequence of still images. Tampering, therefore, can be divided into three categories: spatial tampering, temporal tampering, and spatio-temporal tampering.[36]

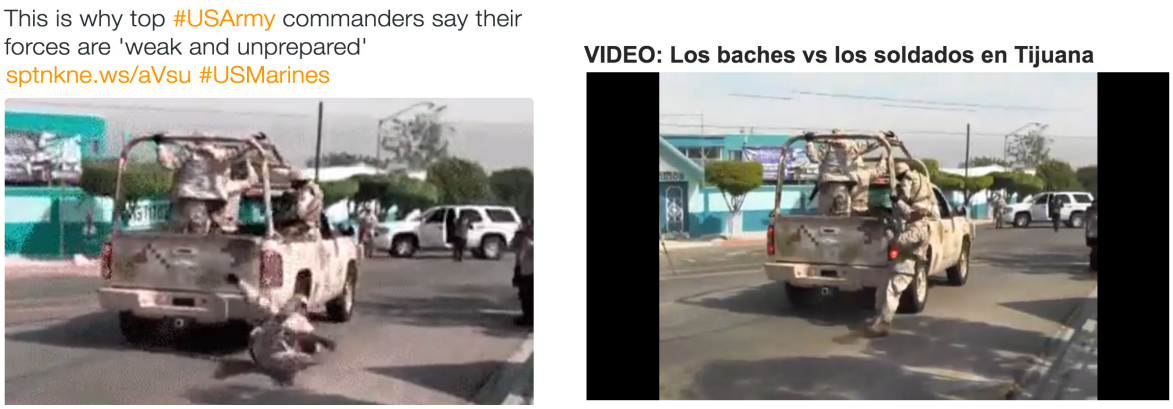

Figure 4. Tweet from Sputnik on left, original video from Mexican Military on right

Source: http://www.sandiegored.com/videos/2993/Los-baches-vs-los-soldados-en-Tijuana/

34 Hany Farid, "Image Forgery Detection," *IEEE Signal Processing Magazine*, Vol. 26, No. 2 (2009), pp. 16–25; Mohd Dilshad Ansari, S. P. Ghrera, and Vipin Tyagi, "Pixel-Based Image Forgery Detection: A Review," *IETE Journal of Education* Vol. 55, No. 1 (2014), pp. 40–46.

35 Manish Sahu, "Muzaffarnagar Riots: No End in Sight for Nine Pending Cases - One on 'Fake Video Clip,'" *Indian Express*, March 9, 2016, <http://indianexpress.com/article/india/india-news-india/muzaffarnagar-riots-no-end-in-sight-for-nine-pending-cases-one-on-fake-video-clip/>.

36 J. D. Gavade and S. R. Chougule, "Review of Techniques of Digital Video Forgery Detection," *Advances in Computer Science and Information Technology*, Vol. 2, No. 3 (March 2015), pp. 233–36.

Spatial tampering uses many of the same techniques seen earlier with image manipulation to copy, move, add, or delete objects in a video. Spatial tampering usually only requires commonplace photo editing software. Temporal tampering manipulates the sequence of frames to alter the apparent time sequence of the images. Typical attacks include frame addition, frame removal, and frame shuffling. Finally, as its name suggests, spatio-temporal tampering combines both types of manipulation.

The techniques for detection of video forgery are very similar to those to detect image manipulation. Camera-based techniques rely on detecting anomalies in sensors, lenses, or software processing to identify the origin of the source video.[37] Camera techniques in video suffer from the same deficiencies as those of playing images: namely, there is only a small database of reference baselines with which to compare readings.[38]

A more widely applicable detection technique is format-based. Just like digital images, videos are compressed as they are recorded using a particular compression scheme. In order to manipulate the video, it must first be decompressed for the manipulation then recompressed for final presentation. If an analysis of the video shows that it has been compressed multiple times, then it is an indicator that the video may be a forgery.[39]

The detection of pixel-based manipulations in video is not as advanced as those for images. Detection of copy-move attacks, i.e. copying some portion of the video and pasting it in place to duplicate or cover another image, is complicated by the fact that the video consists of multiple frames of information. An intra-frame attack can be analyzed like a still image, but an inter-frame attack requires different techniques. Various surveys have pointed out that this type of detection is still in its infancy and is hindered by a lack of standardized data sets and methods.[40]

In addition to the above methods for determining video authenticity, analysts can also examine the video's audio track to see if recording features such as reverberation match the depicted environment.[41] Detecting forged digital audio is complicated by the fact that digital recordings often do not contain a so-called "mains signal," which is the fluctuation in a recording signal that accompanies a microphone and is considered the standard means of determining whether a recording has been

37 Ibid.

38 Omar Ismael Al-Sanjary and Ghazali Sulong, "DETECTION OF VIDEO FORGERY: A REVIEW OF LITERATURE," *Journal of Theoretical & Applied Information Technology*, Vol. 74, No. 2 (April 20, 2015), <http://search.ebscohost.com/login.aspx?direct=true&profile=ehost&scope=site&authtype=crawler&jrnl=19928645&AN=102304166&h=qNN9VG9twAjwfZA6aYcrxYia5pgKrwE0sNs1obKa6rYjXnzgtVkvCnhx7Urp3Y%2B5wOHN4Sz85DbZQdZV31cU7Q%3D%3D&crl=c>.

39 K. Sitara and B. M. Mehtre, "Digital Video Tampering Detection: An Overview of Passive Techniques," *Digital Investigation*, No. 18 (September 2016), pp. 8–22, doi:10.1016/j.diin.2016.06.003.

40 Gavade and Chougule, "Review of Techniques of Digital Video Forgery Detection"; Sitara and Mehtre, "Digital Video Tampering Detection"; Al-Sanjary and Sulong, "DETECTION OF VIDEO FORGERY."

41 Simone Milani et al., "Audio Tampering Detection Using Multimodal Features," in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (IEEE, 2014), pp. 4563–4567, doi:10.1109/ICASSP.2014.6854466.

altered.[42] Nevertheless, research has shown that analyzing the audio compression data for signs of modification (double-compression) is an effective means to identify alterations.

There are many different tools and methods that can be used in the debunking of these different kinds of online deception. As the complex and technical nature of the many different types of debunking programs and tools are vast, we detailed a small sampling of some of the more common techniques we found in our research. The table can be found in Appendix 1 and additionally links them to the perceived skill level required to successfully use such a technique.

THE VERIFICATION CLEARINGHOUSE

In July 2014, Malaysian airlines flight MH-17 was en route from Amsterdam to Kuala Lumpur when it disappeared from the radar screen over the territory of Ukraine. Events could not have conspired better to create a more sensationalistic news story out of the tragedy. The disappearance came closely after the mysterious disappearance of another Malaysian Airlines flight, MH-370, over the Indian Ocean and took place during a period of increased terrorism warnings across Europe due to a shooting attack at the Jewish Museum of Belgium a little over a month before.[43] Moreover, many reporters recalled the accidental shoot-down of Siberian Airlines flight 1812 by Ukrainian air forces in 2001 and wondered if events had repeated themselves.[44]

Because of the surge of international interest in the incident, social media platforms were flooded with purported images of the crash site as well as claims and counterclaims of responsibility. Before long, international news media began to converge on the theory that the aircraft was shot down by a surface-to-air missile, likely launched from the disputed territory in Ukraine.[45] Russia blamed Ukraine, and Ukraine, of course, blamed Russia.

Accusations and counter-accusations flew as partisans and observers from all sides presented alternate analyses gleaned from an ever-growing pile of digital evidence. Who was right? Who was wrong? And, most importantly, how could anyone tell?

42 Rafal Korycki, "Authenticity Examination of Compressed Audio Recordings Using Detection of Multiple Compression and Encoders' Identification," *Forensic Science International*, No. 238 (May 2014), pp. 33–46, doi:10.1016/j.forsciint.2014.02.008.

43 Andrew Higgins, "3 Shot Dead at Brussels Jewish Museum," *The New York Times*, May 24, 2014, <https://www.nytimes.com/2014/05/25/world/europe/3-shot-dead-at-brussels-jewish-museum.html>.

44 Cogan Schneier, "History Shows Passenger Plane Shootdowns Often Mistakes," *USA Today*, July 17, 2014, <http://www.usatoday.com/story/news/world/2014/07/17/korean-airlines-flight-007-history/12786285/; "MH17 Crash: History of Passenger Planes Shot down," *BBC*, July 20, 2014, sec. World/Asia, <http://www.bbc.com/news/world-asia-28361223>.

45 Neil Macfarquhar, David M. Herszenhorn, and Rick Gladstone, "Kiev Suggests Plane May Have Been Hit by Missile," *International New York Times*, July 18, 2014, sec. NEWS, 3.

Arguing over the evidence is the linchpin of Western justice systems and is, of course, a key element in any treaty verification process. In a typical process, each side reviews evidence provided by a previously agreed-upon set of sensors and then presents a case based on that common evidence. The Bellingcat.com team investigation mentioned in the introduction was different, however, in that there was no agreed-upon set of sensors and the evidence presented came from multiple sources and was subject to tampering. Despite this, the Dutch-led Joint Investigation Team confirmed the accuracy of their conclusions more than two years after the crash, finding that a Russian missile shot down the plane and that the launching site was in rebel-controlled territory.[46]

Bellingcat offers a glimpse of the future of arms control verification. Through the use of crowd-sourced evidence and sophisticated analytical techniques, Bellingcat journalists were able to prove the veracity of the evidence and were also able to use that evidence to build a persuasive argument in support of their conclusions. Moreover, all of the evidence, methodologies, and conclusions were executed on the web in full view of the interested public. Regardless of the efforts to discredit the conclusions by bad faith actors and propagandists, the evidence and methods used were able to stand on their own.

We believe that the combination of publicly-available information, transparent and well-communicated analytical methodologies, and public scrutiny of the evidence and the conclusions may form the basis of a future arms control monitoring and verification process. We call such a process a "verification clearinghouse" and define it as a publicly-accessible website that accepts digital evidence, subjects it to evaluation, assesses the veracity of the evidence, and uses this accepted evidence to establish or refute a claim.

We admit verification clearinghouses are exotic beasts, rarely seen in captivity. For that reason, in the next section we will explore their habitat and try to catalog the different varieties most likely to be found in the wilds of the Internet.

IDENTIFYING AND EVALUATING DEBUNKING WEBSITES

In this section, we describe the process used to identify useful characteristics and construct a typology of online debunking websites. Our approach started with a survey and review of existing debunking websites identified through Internet searches and references in the literature.[47] We then reviewed each site and selected a representative sample based on topics, approaches (e.g. individual authors or group sites), and recent activity. This resulted in the final list of twenty-one selected sites shown in Appendix 2. Following this, we identified the key features and characteristics of each site and used this analysis to classify the sites into one of five debunking archetypes.

46 "Investigation Crash MH17, 17 July 2014," Dutch Safety Board.

47 Jonathan C. Smith, *Pseudoscience and Extraordinary Claims of the Paranormal: A Critical Thinker's Toolkit* (Hoboken, NJ: John Wiley & Sons, 2011), <https://books.google.com/books?hl=en&lr=&id=sJgONrua8IkC&oi=fnd&pg=PT11&dq=(Smith,+Pseudoscience+and+Extraordinary+Claims&ots=leJZyN7RCn&sig=U9_cSmdzDWBLzuY8tuMPZ43iG3M>.

WEBSITE SELECTION AND EVALUATION

Despite the rapid increase in attention and effort to identify online deception, websites that focus primarily on debunking online deception are relatively few in number and range drastically in focus and topic selection.[48] Due to the scarcity of existing online debunking sites, we did not restrict sites based on topics and also included some sites that were not primarily focused on debunking but included it as a sub-focus of their site. The intent was to identify as many novel sites as possible.

Drawing on existing debunking sites, we analyzed the quality and nature of each site, noting the characteristics that proved to be useful to the user. First, we focused on analyzing sites that were generally regarded as the top debunking sites and had produced the most reputable results as determined by coverage in the major media. Working backwards, we cataloged the characteristics and features of these sites while taking note of what was missing in other, less successful sites. The qualities identified ranged from user interface characteristics, quality control tools, motivational identifiers, and quality indicators. The full survey of features can be found in Appendix 3.

Once a preliminary survey of features had been compiled, we identified debunking sites and mapped them to the table based on which characteristics were present. This exercise both helped with drawing the initial conclusions on the archetypes of different sites and provided a model for useful characteristics in the construction of a potential verification clearinghouse.

KEY SUCCESS FACTORS

Our analysis of the characteristics related to existing debunking sites resulted in the identification of four main success principles: transparency, accuracy, thoroughness, and impartiality. Each of the principles incorporates a number of features that help to explain the effectiveness and intended use of a successful debunking site.

TRANSPARENCY

Transparency was identified as an important aspect of online debunking initiatives because of the importance of trust in verification. In many cases, the debunking sites we identified maintain a degree of anonymity that can degrade trust in the analysis. Even if a site has brief biographies or an "about" page, most do not have the kind of name credibility (and presumed accountability) of a well-known university or government agency. Therefore, sites that had both detailed accounts of who has authored each of the debunked cases and detailed information on the origins of the site itself tended to appear more credible and are likely more successful.

48 Jerry Zhang and Myung Ko, "Current State of the Digital Deception Studies in IS," 2013, <http://aisel.aisnet.org/amcis2013/HumanComputerInteraction/GeneralPresentations/7/>.

ACCURACY

Arguably the most important of the identified principles, accuracy relates to the veracity of the claims and publications put out by the debunking site. As simple as the concept appears to be, judging the accuracy of sites proved to be difficult. We based our assessment on two factors: The first was whether or not the claims put forth in the site's postings were typically contested by the public. Whether or not a site is known for the validity of its assertions was a useful point of departure to gauge its overall reputation. Second, we looked for whether or not the site had been referenced or cited by other publications or newspapers. The use of the conclusions drawn by a debunking site in outside material indicates that there are other organizations that trust the validity of their findings.

THOROUGHNESS

We based our assessment of site thoroughness on an examination of site methods. First, we reviewed cases to see if the method used to debunk a case appeared to be complete in its analysis. Next, we considered whether each step of the debunking process was described in enough detail so that readers could understand and follow the logic path. We made special note of websites that detailed their debunking process so that the average user would be able to recreate the method used.

IMPARTIALITY

It is possible for a partial source to generate an impartial and effective analysis, but the results will always require an extra level of scrutiny. In the case of some regional debunking sites, for example, matters such as events in Crimea are colored based on whether they are debunked by organizations proclaiming anti- or pro-Russian sentiments. A site's impartiality, therefore, plays a large role in user acceptance of the validity of the debunking process and any conclusions drawn.

SECONDARY SUCCESS FACTORS

Additional factors and tools were identified in our analysis of existing debunking sites that appeared to increase chances of success while remaining separate from the analytical process. Of these, the first and possibly most important was an education section. Some sites, such as Bellingcat.com, maintain a section to teach users how to better identify and analyze online deception. This improves the quality of submissions and increases the probability that cases that make it to the expert-level evaluation are indeed cases of online deception. As mentioned earlier, unknowing online users can share misinformation unintentionally. An education section could also have the positive impact of lessening the number of false images and deceptive efforts spread online.

A clean, ad-free, and well-organized site appeared to be correlated with more reputable and credible debunking results. Ideally, sites are entirely advertisement free. At a minimum, sponsorship and the site's advertising policy should be stated clearly and unequivocally on the main landing page.

Finally, a comment section that allows for two-way communication between the analyst and the public also seemed to be a factor in a site's success and overall credibility. Comments not only identified potential flaws in the analysis but also encouraged site users to participate more actively in the analytical process. In order to further encourage public cooperation and involvement, all possible efforts need to be made in order to demonstrate that each claim is taken seriously. A comment section was further indication that this is done in a fair and open manner.

A TYPOLOGY OF DEBUNKING SITES

DEBUNKING TYPE CLASSIFICATION

After identifying key characteristics and outlining common principles, we classified the sites into one of five types based on their intention, methodology, and product (Table 1). These five types are: academic, bulk, explanative—regional, explanative—global, and explanative—topical. Debunking efforts seemed to be focused broadly on two separate goals. One grouping approaches debunking using a public crowdsourcing model, while the other utilizes a citizen science or expert publication model. Both have advantages and disadvantages and are tailored for different aspects of the online deception debunking effort. Below we describe each type in more detail.

Table 1. Debunking Site Typology

Academic	Bulk	Explanative: Global, Regional, Topical
Debunking sites that are focused primarily on isolated cases and which offer a high degree of analytical rigor.	Debunking sites that focus on addressing as many classes of possible deception as possible.	Debunking sites that focus on providing detailed cases that focus on the explanation of why or why not a case is deceptive. These sites can be separated into those focused on regional, topical, or global issues.

ACADEMIC

The academic debunking site is the archetype of the expert publication approach. The purpose of the academic debunking site is to showcase and explain the debunking process. It shows very little interaction with the public in the debunking process and typically only entertains limited public discussion in the comment sections of the site. This type of site does not typically focus on one specific field of expertise (either subject matter or geographic region), emphasizing instead the debunking argument. The level of validity and thoroughness in an academic debunking site will be high. Analysis is often produced by recognized experts in the field and will typically attempt to identify deception that has yet to be identified by the greater public or governments.

BULK

The bulk debunking site focuses on debunking/verifying as many possible cases of online deception as possible. The methodology used in the identification and later analysis of each case varies from site to site, and occasionally from case to case. Usually, however, bulk debunking sites will use a type of crowdsourced submission system and a staff of authors that works to ascertain the validity of the submitted claim.

EXPLANATIVE – REGIONAL, GLOBAL, AND TOPICAL

Explanative sites work within the bounds of a certain subject. Some might be focused on election-related rumors (topical), while others are focused on events in Ukraine (regional), and the remainder have a global approach that borders on the bulk type site. The key part of an explanative site is that it works to answer issues raised by the public in relation to the subject area and provides a reasonably thorough explanation as to its findings. The emphasis on explanation is the basic difference between explanative—global sites and the bulk type. Whereas a bulk site has the main goal of addressing as many claims as possible, at times only providing a true or false answer, an explanative—global site always emphasizes the explanation of the answer.

APPLICATIONS

Crowd-based "societal verification" has been a dream of the arms control community since the 1950s.[49] A crowd-based monitoring website could be a first step towards such a system. However, moving beyond the game of claim and counterclaim to the point of actual verification will likely require merging public observation with expert analysis.

We can use our typology to assist both in the better utilization of pre-existing sites as well as the creation of new ones. The diagram at Figure 5 can be used to identify the relevant features of existing sites and highlight the type of site most appropriate for the given monitoring or verification task.

49 Seymour Melman, "How Can Inspection Be Made to Work?," *Bulletin of the Atomic Scientists* Vol. 14, No. 7 (1958), pp. 270–272, doi:10.1080/00963402.1958.11453870.

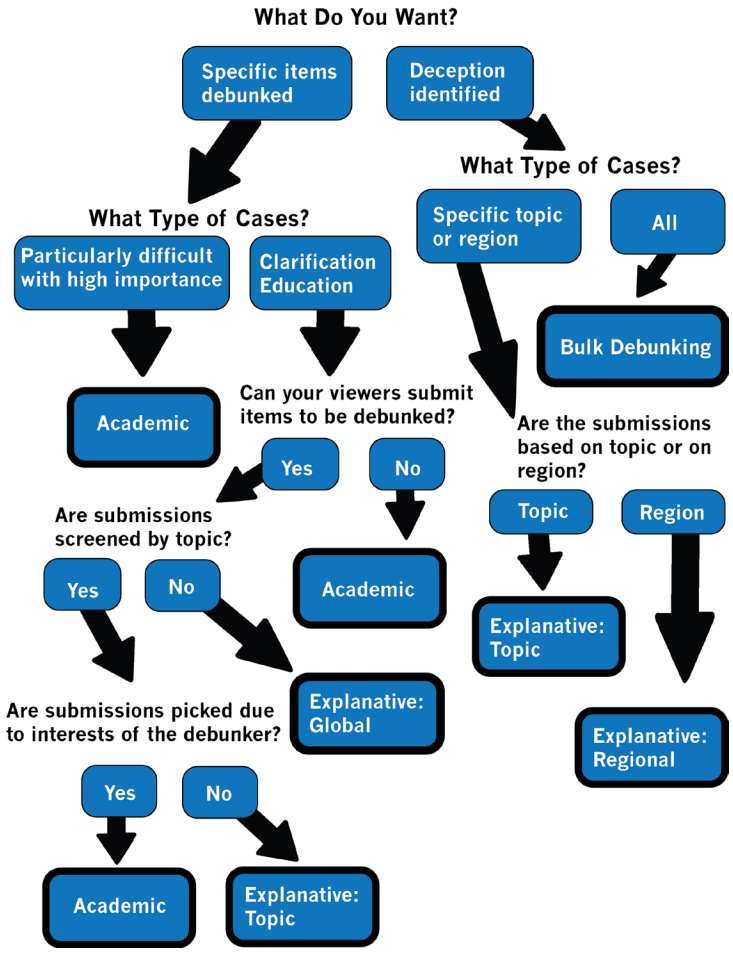

Figure 5. Site Type Flow Chart

The other potential use for the feature checklist in Appendix 3 and the typology table is for the creation of new debunking sites. This is resource-intensive but likely the most useful approach due to the scarcity of national security-focused online verification sites. Just as in evaluating existing sites, the checklist and table can be used to generate the structure and key features of the site. First, the user selects the objective and matches it to one of the identified types using Figure 5. Then, using the features list, the user can identify the critical features and tools that will be required.

Finally, the identified sites and types can be used as case studies when exploring design options. It is important to remember that not all the characteristics identified are necessary or work well with every other feature. For this reason, it is useful to look at existing sites to understand how the features and types function in the real world. For example, if the objective is to design a new explanative-regional type debunking website focused on conventional arms trading in Africa, then an analysis of existing explanative—regional sites is needed. Example explanative—regional sites such as Stopfake.org scored particularly high in the principles of verifiability and accuracy but got low scores in transparency and impartiality. Reviewing the examples identified will allow users to balance priorities against capabilities to provide the best chances at success.

DISCUSSION

We have defined a verification clearinghouse as a publicly-accessible website that accepts digital evidence, subjects it to evaluation, assesses the veracity of the evidence, and uses accepted evidence to establish or refute a claim. It is important to note that our definition is process-based and not restricted to a particular set of functions or software tools. Using this as a start, we attempted to identify websites that contained substantial portions of our necessary processes. As expected, we did not identify a single website that fulfilled all of our requirements, but we found a large number that shared elements of them. We grouped these into types to help clarify the features and processes that would likely be most effective in a treaty verification context.

To recall the challenge, treaty verification takes place in a context where deception is expected, that is, situations like the MH-17 example of accusation/counter-accusation will be the norm and not the exception. Moreover, in a public environment, many players would be involved in the deception effort, including self-motivated individuals, as well as state and non-state actors. The involvement of various actors would also imply various degrees of sophistication of any tampering efforts and a broad base from which to take advantage of the compounding effects of social media.

To overcome these challenges, a verification clearinghouse must have the proper provenance, people, and tools. Provenance means that ownership of the site must be clear and unequivocal in order to avoid charges of hidden manipulation. In a treaty context, joint operation between the affected parties or operation by an international compliance agency would be the obvious choice. For the people or experts tasked with the analysis, the key again is to avoid the impression of favoritism for one side or the other. Choosing a balanced group of experts from the treaty parties or using the staff of an international organization would be reasonable.

An additional consideration with respect to people is on the public side of the clearinghouse. Assuming that experts evaluate evidence received from the public, the clearinghouse must also have mechanisms to filter the nominations and prevent cheating.[50] This is not straightforward in a verification context because of the tension between security and anonymity. Ideally, a clearinghouse would accept nominations from any source, and the source would be assured of complete anonymity to avoid the possibility of retaliation. If the nomination is entirely anonymous, however, low quality or spurious nominations (either intentional or unintentional) may overwhelm the site.

A solution is to use a check and balance system containing three elements: content moderation, reputation management, and crowd voting.[51] Content moderation is the familiar system of flagging comments or submissions that seem spurious. This information can be deleted outright, or an inquiry can be sent requesting further information and clarification. Reputation management is a means whereby submitters are ranked based on factors such as quality of previous submissions,

50 Victor Naroditskiy et al., "Crowdsourcing Contest Dilemma," *Journal of The Royal Society Interface*, Vol. 11, No. 99 (August 20, 2014), doi:10.1098/rsif.2014.0532.

51 Vasco Furtado et al., "Collective Intelligence in Law Enforcement – The WikiCrimes System," *Information Sciences*, Vol. 180, No. 1 (January 2, 2010), pp. 4–17, doi:10.1016/j.ins.2009.08.004.

institutional affiliation, participation, or length of time of membership, and the submissions of those with high ranks are given more weight or priority. Finally, crowd voting lets the crowd itself determine the quality of the submission, with those that receive the most votes getting priority for analysis.

With respect to tools, any analytical software used for a verification clearinghouse should be restricted to open source.[52] This would allow for reproducibility of results and also reduce the potential for charges of manipulation. Our analysis has shown that the tools and techniques used to identify tampering are adequate to the task. They are particularly strong in detecting image manipulation, especially because the detection techniques are relatively easy to understand and reproduce. Detection of audio and video tampering is more problematic. The tools and methods are accurate and reliable, but the analysis can be time-consuming. Moreover, the determination of the answer relies on the calculation of an algorithm that is unlikely to be understood by a lay person. Nevertheless, the tools exist and, in most instances, are capable of determining whether digital information has been tampered with. Even in the cases where determining fraud is difficult, the capabilities available raise the cost of tampering and make committing fraud difficult and time-consuming.

It should also be pointed out that a verification clearinghouse does not have to fulfill a traditional treaty monitoring and verification role in order to be beneficial. Greater transparency in conflict areas or additional eyes on questionable activities in the research arena would serve to improve safety and security for all. A clearinghouse could also be useful as a platform for cooperative activities or other transparency and confidence-building measures.[53] Last, it could provide useful information as a type of supplemental monitoring system. That is, treaty parties would not be obligated to treat the information as official, but it would be publicly available. This would make it difficult to ignore or diminish obvious violations.

Given our understanding of the challenges and considerations, constructing a model verification clearinghouse should be possible. We have seen that it would be a mixed type, containing the crowd submission aspects of the bulk type, the expert-level evaluation of the academic type, and the targeted focus of the explanative type. Reviewing the dimensions inherent in each results in the following table.

52 Open Source Software Foundation, "Frequently Answered Questions, 'What Is "Open Source" Software?,'" Opensource.org, 2016, <https://opensource.org/faq#osd>.

53 Dan McMorrow, "Open and Crowd-Sourced Data for Treaty Verification" (DTIC Document, October 2014), <http://oai.dtic.mil/oai/oai?verb=getRecord&metadataPrefix=html&identifier=ADA614684>.

Table 2. Verification Clearinghouse Requirements

Dimension	Verification Clearinghouse Requirements	Associated Type
Audience	Public access	All
Data Type Analyzed	Image, video, audio, text	Academic
Source Citation	Sources must be identified	Academic, Explanative: Regional, Global, & Topical
Case Nomination	Crowd	Bulk
Debunking Method	Open source tools; expert-level; sophisticated; fully documented methods; repeatable technical analysis	Academic
Topic Area	Specialized (treaty verification)	Academic, Explanative: Topical
Provenance	Independent or sanctioned body	Not specific to one type
User Features	Comments; education section; reputation management; crowd voting	Academic
Authors	Disinterested experts; not anonymous	Not specific to one type

Conceptually, then, designing a verification clearinghouse is straightforward. Nevertheless, it is clear that several challenges to the concept remain. These can be divided into technical challenges related to the operation of the site and the conduct of the analysis and policy challenges related to the legal framework of the site and the acceptability of the findings.

The chief technical challenge related to the verification clearinghouse that is dependent upon public nomination of cases is the difficulty of motivating participation. The crowdsourcing literature has identified numerous factors that influence a user's decision to participate in a crowdsourcing process, including altruism, monetary reward, reputational reward, and sense of community.[54] Unfortunately, no single factor is dominant, and the exact combination seems to change with each crowdsourcing effort. Research can inform the process design, but the final selection of motivating features can only be resolved through trial and error. One possible means of overcoming this challenge would be to partner with one of the major commercial crowdsourcing platforms such as Innocentive or Chaordix to take advantage of their experience and tested incentive methodologies.

A second key challenge related to operation of the site is the prevention of cheating and hacking. Cheating in this instance means an intentional effort to undermine the analytical process through some combination of multiple nominations, fraudulent voting, deletion of submitted work, or other

54 Haichao Zheng, Dahui Li, and Wenhua Hou, "Task Design, Motivation, and Participation in Crowdsourcing Contests," *International Journal of Electronic Commerce*, Vol. 15, No. 4 (Summer 2011), p. 57.

efforts intended to hinder the success of some group or individual. "Sock puppetry," or the creation of multiple accounts under fake names in order to skew voting results, is a common example of cheating found on crowdsourcing sites. To prevent this, careful consideration must be given to a stringent reputation management system, including dedicated site moderators. Strong moderation, together with a clear appeals process, can also help prevent flooding the system with spurious nominations that attempt to overwhelm the analysts. Hacking, of course, is an additional concern. The discussion of appropriate cyber security measures is beyond the scope of this study, but it is sufficient to say that the site design should build in the probability of large-scale, state-sponsored hacking attempts.

The final technical challenge is simply the difficulty inherent in the "arms race" between deceivers and investigators. As software becomes ever more sophisticated, tampering becomes easier to execute and harder to detect. This means that the accepted verification clearinghouse methodology must be flexible and open to a constant influx of new software tools and methods. It would not be acceptable, for example, to create a static list of approved software tools that could only be updated in line with treaty renegotiation windows. A procedure would have to be in place that would allow the selection implementation of new tools to happen much more rapidly in order to keep pace with the latest changes in the field of tampering and tampering detection.

A detailed study of the policy challenges to verification clearinghouses is beyond the intent of this paper, but two main concerns stand out. The first is whether state parties would be willing to cooperate in a verification forum that relied on publicly available information. In fact, they already do. State level cooperation that includes the use of publicly available information already takes place in the areas of wildlife management, fishing treaties, and environmental science.[55] Objections raised typically assert that the information involved is militarily sensitive and states would insist that it remain out of the public realm. This is perhaps the case, but it fails to explain numerous instances of cooperation between the public and international law enforcement efforts, which are equally sensitive within the realm of enforcement activity. This is especially the case recently, as law enforcement and counterterrorism efforts have increasingly merged. Given that there are numerous examples of this type of cooperation, and the fact that treaties can be framed in a way to clearly delineate all aspects of the process, the question of adopting a verification clearinghouse framework is a matter of will, not way.

A more pressing difficulty is the involvement of so-called Ethical, Legal, and Social Issues (ELSI) in societal verification efforts.[56] Foremost among these are the ethics of security. It is easy to imagine a situation where a terrorist organization or government retaliates against individuals who share information unfavorable to their position. Depending on the treaty and the monitoring environment, a verification clearinghouse might have to offer strict security controls along the lines of those already

55 "WHO WE ARE" Global Partnership for Sustainable Development Data, 2015, <http://www.data4sdgs.org/who-we-are/>.

56 "Redefining Societal Verification," Innovating Verification (Nuclear Threat Initiative, July 2014), <http://www.nti.org/media/pdfs/WG2_Redefining_Societal_Verification_FINAL.pdf?_=1405443059>; Kelsey Hartigan and Maynard Holliday, "A Review of Emerging Legal and Ethical Issues in Societal Verification," in *INMM Proceedings*, 2014.

implemented by leading media organizations to protect the identity of tipsters.[57] A related ethical concern is that of privacy. This is somewhat attenuated in the verification clearinghouse scenario because gathering individual data beyond some kind of basic identification (for reputation management) is not the purpose of the site. Moreover, users would be expected to acknowledge a terms of use statement that clarified how their information would be used in order to participate.

Legal frameworks for online data use and collection continue to evolve. However, current trends seem to support fewer restrictions on the use of publicly available data as long as it is not used for commercial gain.[58] This is increasingly being formalized in the European Union, where regulators and the courts are codifying many of the data practices used in online transactions.[59]

The social aspects of data collection are also evolving, and a verification clearinghouse should make a special effort to take cultural and political sensitivities into account. Participants cannot be expected to have common expectations of privacy and transparency in their submissions; a human rights NGO might have a different agenda in reporting than a partisan in a military conflict. In order to encourage participation by all sides, care must be given to a consistent application of rules and to the creation of an environment that recognizes cultural and political differences.[60]

CONCLUSION

The problem of deception remains the fundamental concern of arms control verification. This problem is exacerbated online by anonymity and the dynamics of rumors and misinformation. We hope to have demonstrated that the issue of false or misleading online information is not only manageable but is already being credibly addressed by a variety of organizations and individuals. Based on these existing examples, we believe a verification clearinghouse is possible and that elements of such a system already exist today in the real world.

Our review of existing debunking websites has identified several success factors common to all. The first is transparency, or clearly demonstrating the provenance of the site and the analysts associated with it. Next is accuracy, which is simply the demonstration of expertise, reliability, and credibility in technical analysis. Thoroughness was also important and means not only examining the question from all angles and attempting to answer reasonable objections but clearly showing the channels by which information is received and moved through the analytical process. Finally, and of particular

57 Andy Greenberg, "SecureDrop Project Will Pay To Install Media Outlets' WikiLeaks-Style Submission Systems," *Forbes.com*, October 15, 2013, <http://www.forbes.com/sites/andygreenberg/2013/10/15/securedrop-project-will-pay-to-install-media-outlets-wikileaks-style-submission-systems/#62035d5cf6b4>.
58 "Frequently Asked Questions - Creative Commons," Creative Commons, August 4, 2016, <https://creativecommons.org/faq/#what-is-creative-commons-and-what-do-you-do>.
59 Alan Travis and Charles Arthur, "EU Court Backs 'Right to Be Forgotten': Google Must Amend Results on Request," *The Guardian*, 13 2014, <https://www.theguardian.com/technology/2014/may/13/right-to-be-forgotten-eu-court-google-search-results>.
60 McMorrow, "Open and Crowd-Sourced Data for Treaty Verification."

importance in the verification context, is impartiality. Both the analysts involved and the owners of the forum must be (and must be seen by the public to be) fair and impartial observers.

There are two basic objections to the verification clearinghouse. The first is that it is not technically feasible because deception is so widespread and the risks of accepting false information are too high. The second is that it is not feasible from a policy perspective because states would never accept public information for monitoring purposes. On the second point, we can only point to current global cooperative efforts such as the UN-sponsored Global Partnership for Sustainable Development Data or the World Health Organization's Global Public Health Information Network as examples of the integration of public-sourced information into governmental monitoring systems.

On the technical front, freely available, analytical software and user groups have democratized tampering detection. Moreover, examples such as Bellingcat.com, 38North.com, and Armscontrolwonk.com clearly demonstrate that private-sector analysts are able to detect deception from sophisticated state actors such as Russia and North Korea. Indeed, the combination of new software tools, replicable research methods, and crowd scrutiny and discussion has made it possible to generate analytical judgments that are at the same level of quality as those that could be done previously only by national intelligence agencies.

Creating a workable verification clearinghouse seems eminently feasible. It would consist of a mix of elements from the types of debunking sites identified in this paper. This would likely be crowd-based case submission similar to the bulk type site, the expert level analysis and careful methodologies of the academic type, and the topical focus of the explanatory type.

National governments and international agencies already possess the technical expertise and crowd-sourcing experience required to set up and manage a verification clearinghouse. There are, of course, new policy challenges related to implementation of such a solution, but that is the case with every new technology: recall the difficulties faced by the Nixon and Carter administrations in convincing Congress to accept treaty monitoring by "national technical means."[61] In this instance, the benefits of increased transparency, low costs, and crowd-based accuracy all argue for such a solution. In addition, a verification clearinghouse could serve as a step towards engaging small or marginalized states in the cause of international nonproliferation security. By involving a state and its citizens, the clearinghouse could become a building block for future societal verification efforts.

61 Jeffrey Richelson, "National Security Archive Briefing Book No. 231 Declassifying the 'Fact Of' Satellite Reconnaissance," October 1, 2007, <http://webcache.googleusercontent.com/search?q=cache:vtUDSQIA7h0J:nsarchive.gwu.edu/NSAEBB/NSAEBB231/+&cd=2&hl=en&ct=clnk&gl=us>.

APPENDICES

APPENDIX 1: COMMON DECEPTION AND DETECTION METHODS

Deception Tool	Detection Method	Difficulty Level
Copy Move: Such as when a section of the image is copied and pasted on the same image using such software applications as Photoshop. Or Composition or splicing manipulations: Merging one image with one or more parts of other images to create the illusion of one complete image (Rocha et al). (Common in memes and states trying to embellish military or technical capabilities)	This can be detected visually if the duplicated section is large enough or inconsistencies alert the viewer to an unnatural section of the image.	Novice
	Applications that analyze the pixels to see if any section had been edited. Sites such as: http://fotoforensics.com/	Enthusiast
Misattribution: When an image is labeled as evidence for something it is not. Similar to when an image or video is pulled from an old event as evidence for a recent occurrence.	This can be detected by a number of methods: A reverse Google image search can be used.	Novice
	For both photo and video deception, identification of context clues contained in the image can be used to determine the true origin.	Enthusiast
	A more advanced method could be geolocation identification, where the geographical features can be used to triangulate the location from satellite imagery.	Expert

Deception Tool	Detection Method	Difficulty Level
Intra-Frame Video Editing: When a video is decompressed and the image frames of the video are edited with such photo editing methods such as "copy move."	Just as this deception method needs a higher level of technological skill and access to the appropriate software, the detection requires similar know-how. The ability to decompress and evaluate the individual frames of a video is necessary for the detection of this deception method.	Enthusiast—Expert
False Geolocation Tag: When a false geolocation tag is used on social media applications such as Instagram or Twitter (similar to misattribution).	This can be identified visually if the region photographed is recognizably not where it is tagged to.	Novice—Enthusiast
	The geolocation can sometimes be found in the metadata of the actual photograph.	Enthusiast
	Other geolocation tags from the same user can be analyzed to see if their claim of being in a specific location is credible.	Enthusiast

APPENDIX 2: SELECTED DEBUNKING WEBSITES

Site Name	Website
Bellingcat	https://www.bellingcat.com
Snopes	http://www.snopes.com
Channel 4: Fact Check	https://blogs.channel4.com/factcheck/
Doubtful News	http://doubtfulnews.com/
Emergent	http://emergent.info
Metabunk	https://www.metabunk.org
The Skepdic's Dictionary	http://www.skepdic.com
France 24: Observers	http://observers.france24.com/en/
Stop Fake	http://www.stopfake.org/
Truth or Fiction	https://www.truthorfiction.com/
Fact Check	http://factcheck.org/
CSI: The Committee for Skeptical Inquiry	http://www.csicop.org/
Climate Feedback	http://climatefeedback.org/
Health News Review	http://www.healthnewsreview.org/
PESA Check	https://pesacheck.org/
ABC: Fact Check	http://www.abc.net.au/news/factcheck/
Full Fact	https://fullfact.org/ask/
Hoax of Fame	http://hoaxoffame.tumblr.com/
Claim Buster	http://idir-server2.uta.edu/claimbuster
Arms Control Wonk	http://www.armscontrolwonk.com/
Washington Post: The Intersect	https://www.washingtonpost.com/news/the-intersect/

Type Examples

Academic	Clearing-house	Explanative—Regional	Explanative—Global	Explanative—Topical
☐ Bellingcat ☐ Arms Con-trol Wonk	☐ Snopes	☐ Stop Fake ☐ ABC Fact Check	☐ France 24 Observers ☐ Doubtful News	☐ Climate Feedback ☐ Full Fact

APPENDIX 3: DEBUNKING SITE FEATURE IDENTIFICATION CHECKLIST

_____ Does this site give a clear explanation of how each case was debunked?

_____ Can this site's debunking process be replicated based on the information provided?

_____ Does this site show the ability to use photo analysis in debunking online deception?

_____ Does this site show the ability to use video analysis in debunking online deception?

_____ Does this site show the ability to use audio analysis in debunking online deception?

_____ Does this site use open source analysis tools to debunk the cases?

_____ Does this site analyze photo/video metadata?

_____ Does this site appear to debunk cases quickly?

_____ Does this site appear to use text analysis when debunking online deception?

_____ Does this site address issues from a diverse amount of locations in different regions?

_____ Does this site permit crowd submissions for cases?

_____ Does this site have an ability to comment on cases?

_____ Is this site free of advertisement?

_____ Does this site adequately link the reference material that is used to debunk the misinformation?

_____ Does this site give bios for the authors of debunked cases?

_____ Does this site have a portion dedicated to informing the public on how to better use open source tools to detect online deception?

_____ Is the site free of any kind of government associations?

_____ Are the authors of debunked cases from relevant backgrounds?

_____ Do the authors have bios posted to the site?

_____ Do the authors have CVs or information on educational backgrounds posted to site?

_____ Are the authors free from association with any kind of biased agency?

OCCASIONAL PAPERS AVAILABLE FROM CNS

online at http://nonproliferation.org/category/topics/cns_papers

***#33* Evaluating WMD Proliferation Risks at the Nexus of 3D Printing and Do-It-Yourself (DIY) Communities**
Robert Shaw, Ferenc Dalnoki-Veress, Shea Cotton, Joshua Pollack, Masako Toki, Ruby Russell, Olivia Vassalotti, Syed Gohar Altaf • 2017

***#32* Taiwan's Export Control System: Overview and Recommendations**
Melissa Hanham, Catherine Dill, Daniel Salisbury, P. Alex Kynerd, Raymond Wang • 2017

***#31* Revisiting Compliance in the Biological Weapons Convention**
James Revill • 2017

***#30* Crowdsourcing Systems and Potential Applications in Nonproliferation**
Bryan Lee • 2017

***#29* The Verification Clearinghouse: Debunking Websites and the Potential for Public Nonproliferation Monitoring**
Bryan Lee, Kyle Pilutti • 2017

***#28* Geo4nonpro.org: A Geospatial Crowd-Sourcing Platform for WMD Verification**
Melissa Hanham, Catherine Dill, Jeffrey Lewis, Bo Kim, Dave Schmerler, Joseph Rodgers • 2017

***#27* Searching for Illicit Dual Use Items in Online Marketplaces: A Semi-Automated Approach**
Bryan Lee, Margaret Arno, Daniel Salisbury • 2017

***#26* 2016 Symposium Findings on Export Control of Emerging Biotechnologies**
Steven Fairchild, Caroline R. M. Kennedy, Philippe Mauger, Todd J. Savage, Raymond A. Zilinskas • 2017

***#25* Outlawing State-Sponsored Nuclear Procurement Programs & Recovery of Misappropriated Nuclear Goods**
Leonard S. Spector • 2016

***#24* Strengthening the ROK-US Nuclear Partnership**
Miles A. Pomper, Toby Dalton, Scott Snyder, Ferenc Dalnoki-Veress • 2016

***#23* Replacing High-Risk Radiological Materials**
George M. Moore, Miles A. Pomper • 2015

***#22* A Blueprint to a Middle East WMD Free Zone**
Chen Kane, PhD • 2015

***#21* Biotechnology E-commerce: A Disruptive Challenge to Biological Arms Control**
Raymond A. Zilinskas, Philippe Mauger • 2015

***#20* Countering Nuclear Commodity Smuggling: A System of Systems**
Leonard S. Spector, Egle Murauskaite • 2014

***#19* Alternatives to High-Risk Radiological Sources**
Miles Pomper, Egle Murauskaite, Tom Coppen • 2014

***#18* Stories of the Soviet Anti-Plague System**
Casey W. Mahoney, James W. Toppin, Raymond A. Zilinskas, eds. • 2013

***#17* Ugly Truths: Saddam Hussein and Other Insiders on Iraq's Covert Bioweapons**
Amy E. Smithson, PhD • 2013

***#16* Rethinking Spent Fuel Management in South Korea**
Ferenc Dalnoki-Veress, Miles Pomper, Stephanie Lieggi, Charles McCombie, Neil Chapman • 2013

Older Papers

#15 Engaging China and Russia on Nuclear Disarmament • *2009*

#14 Nuclear Challenges and Policy Options for the Next US Administration • *2009*

#13 Trafficking Networks for Chemical Weapons Precursors: Lessons from the 1980s Iran-Iraq War • *2008*

#12 New Challenges in Missile Proliferation, Missile Defense, and Space Security • *2003*

#11 Commercial Radioactive Sources: Surveying the Security Risks • *2003*

#10 Future Security in Space: Commercial, Military, and Arms Control Trade-Offs • *2002*

#9 The 1971 Smallpox Epidemic in Aralsk, Kazakhstan, and the Soviet Biological Warfare Program • *2002*

#8 After 9/11: Preventing Mass-Destruction Terrorism and Weapons Proliferation • *2002*

#7 Missile Proliferation and Defences: Problems and Prospects • *2001*

#6 WMD Threats 2001: Critical Choices for the Bush Administration • *2001*

#5 International Perspectives on Ballistic Missile Proliferation & Defenses • *2001*

#4 Proliferation Challenges and Nonproliferation Opportunities for New Administrations • *2000*

#3 Nonproliferation Regimes at Risk

#2 A History of Ballistic Missile Development in the DPRK • *1999*

#1 Former Soviet Biological Weapons Facilities in Kazakhstan: Past, Present, and Future • *1999*

nonproliferation.org

Middlebury Institute *of*
International Studies at Monterey
James Martin Center for Nonproliferation Studies