

# Ambiguous Self-Induced Disinformation (ASID) Attacks: Weaponizing a Cognitive Deficiency

M Canham<sup>1,2</sup>, S Sütterlin<sup>3,4</sup>, TF Ask<sup>4,5</sup>, BJ Knox<sup>4,5,6</sup>, L Glenister<sup>2</sup>, and RG Lugo<sup>4,5</sup>

<sup>1</sup>*Chief Executive Officer, Beyond Layer Seven,  
United States*

*E-mail: mcanham@belay7.com*

<sup>2</sup>*School of Modeling Simulation and Training,  
University of Central Florida,  
Orlando, FL, USA*

*E-mail: mcanham@ist.ucf.edu*

<sup>3</sup>*Faculty of Computer Science,  
Albstadt-Sigmaringen University,  
Germany*

*E-mail: stefan.suetterlin@hs-albsig.de*

<sup>4</sup>*Faculty of Health, Welfare and Organisation, Østfold University College,  
Norway*

*E-mail: torvalda@stud.ntnu.no; benjamin.j.knox@ntnu.no; ricardo.g.lugo@ntnu.no,*

<sup>5</sup>*Department of Information Security and Communication Technology,  
Norwegian University of Science and Technology,  
Gjøvik, Norway*

<sup>6</sup>*Norwegian Armed Forces Cyber Defence,  
Norway*

*E-mail: bknox@mil.no*

**Abstract:** *Humans quickly and effortlessly impose context onto ambiguous stimuli, as demonstrated through psychological projective testing and ambiguous figures. This feature of human cognition may be weaponized as part of an information operation. Such Ambiguous Self-Induced Disinformation (ASID) attacks would employ the following elements: the introduction of a culturally consistent narrative, the presence of ambiguous stimuli, the motivation for hypervigilance, and a social network. ASID attacks represent a low-risk, low-investment tactic for adversaries with the potential for significant reward, making this an attractive option for information operations within the context of grey-zone conflicts.*

**Keywords:** *Disinformation, Cognitive Malware, Grey Zone Conflict, Ambiguous Self-Induced Disinformation Attacks*

## **Introduction**

*The Twilight Zone* episode “Monsters Are Due on Maple Street” depicts characters from a 1950s era American town who believe that they have witnessed a UFO. After several strange occurrences, the townspeople begin suspecting each other of being alien invaders disguised as humans. The episode concludes with the townspeople assaulting and murdering each other. The final scene reveals that aliens were behind the ambiguous occurrences that fed the townspeople’s paranoia. One alien says to the other “all we need to do is sit back, watch and let them destroy themselves”. This *Twilight Zone* episode provides an example of how ambiguity can be weaponized. Human cognition works to eliminate ambiguity by imposing meaning on noisy stimulus through context. This work argues that this cognitive proclivity may be leveraged to enhance a cyberattack, physical attack, or could be used as a form of disinformation attack by itself.

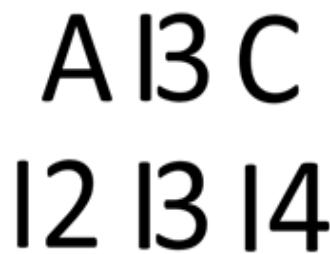
In the quest for information superiority and psychological dominance to weaken a rival country, entity, or alliance, factors that affect the flow and type of information play a significant role in the way decisions are made. Adversarial power in information warfare often relies upon deception, agitation, ambiguity, and propagation of narratives to influence people into making decisions adverse to their own interest, and that favor the deceiver or manipulator conducting the information or influence operation (Bagge 2019). Building on the fictional introductory example, this article introduces potential real-world examples where injecting ambiguous information into the information stream, combined with a culturally consistent narrative and hypervigilance, can weaponize a normal feature of human cognition. The resultant disrupting effects are self-induced, concealing the asymmetrical (military) means. This attack categorically falls within the concept of a Grey Zone conflict where approaches “encompass operations that fall short of warfare due to intensity, legality or ambiguity” (Fitton 2016, p.111). Exploiting this competition space between peace and war has renewed relevance, due to technological advances that can be utilized to attack human cognitive vulnerabilities.

The article proceeds by presenting a theoretical and historical grounding for Ambiguous Self-Induced Disinformation (ASID) attacks. The key elements of ASID attacks are explained before the authors present examples and forecast potential arenas where an ASID attack could be deployed. Finally, a future directions section highlights interdisciplinary approaches that can help develop a means by which to defend against such attacks.

## **Imposing Order on a Chaotic World**

The human mind is so strongly inclined toward imposing context onto ambiguous stimuli that a deliberate effort is required to ignore one’s own projections to perceive information as it truly exists. This cognitive imposition can be easily demonstrated by viewing ambiguous figures. In the figure below, the middle characters are perceived as either the capital letter ‘B’ or the number ‘13’, depending on the contextual information that is provided by the surrounding characters. In fact, the middle characters are identical and ambiguous. They are not in any way altered by the contextual characters; rather, humans cognitively impose the context on them through their background knowledge of the English alphabet and numerals. An observer lacking this knowledge will see the same characters in both contexts because he or she lacks the requisite knowledge to impose such

context. This occurs through a process referred to as *knowledge-driven processing*, also referred to as *top-down processing*. When information in our environment is interpreted, one's subjective experience is that it is being interpreted without bias and one is simply taking in what is there through a *data-driven process* (Cavanagh 1991; Goolkasian 1987; Theeuwes 2010; Kornmeier & Bach 2012; Ishizu 2013; Intaite *et al.* 2019). This data-driven processing, also referred to as *bottom-up processing*, is how incoming information is perceived when no context for what is being perceived is available. It is more common for one to process the stimuli encountered through the lens of one's prior knowledge, hence approaching it from a top-down or knowledge-driven perspective, even when one subjectively feels as though interpretation is without bias (Cavanagh 1991; Goolkasian 1987; Theeuwes 2010; Kornmeier & Bach 2012; Ishizu 2013; Intaite *et al.* 2019).



**Figure 1:** Ambiguous figure

The presentation of ambiguous information that is also accompanied through a culturally consistent or plausible narrative (the latter serving as prior knowledge), can then lead to a form of self-induced disinformation effect. In **Figure 1**, above, it is difficult to ‘unsee’ the B or 13 even with the knowledge that this is a combination of capital letter I and the number 3 to create an ambiguous figure. This example involves low-level perception with known ground truth, but what happens when the truth is less concrete and the context more ambiguous?

For over 100 years, psychologists have relied upon projective testing to gain insights into the inner workings of the human mind. These tests are effective because patients project their own interpretation onto the ambiguous stimuli in the test. An example of projective testing can be found in the Thematic Apperception Test (Murray 1943). This test employs illustrations of characters engaged in ambiguous situations to which a patient or test subject is then asked what has led to the depicted event, what is happening in the moment of depiction, what the characters are experiencing, and what the outcome is. These tests demonstrate humans' ability to apply fabricated context to ambiguous stimuli, and this effect has been demonstrated through seeing figures in ambiguous images (Cavanagh 1991; Goolkasian 1987; Kornmeier & Bach 2012; Ishizu 2013; Intaite *et al.* 2019) or recognizing trends in stock market fluctuations (Cootner 1964; Fama 1970; Malkiel 2003).

This leads us to ask the question, what if this feature of human cognition, the capacity to project context onto ambiguous stimuli, could be weaponized as part of an information operation through the presentation of a culturally consistent narrative paired with ambiguous information?

### **Ambiguous Self-Induced Disinformation (ASID) Attacks**

Cyber operators and threat intelligence analysts are bombarded with information from sensors in machines, on networks, and from cyberspace in general. Sifting through too much information poses as much of a challenge as suffering from insufficient information. A major obstacle to suc-

cessfully detecting attacks is that much of the information to sift through is ambiguous; it might indicate an attack, or it might be innocuous, such as a user visiting an unauthorized website from a work computer. Previous research suggests that these ambiguous signals could be weaponized to suppress network defender responses to attacks (Sawyer *et al.* 2016). These *prevalence denial attacks*, also known as *grey signal attacks*, leverage the “artificial inflation of signal probability flooding a network with grey signals, purposely built to be flagged by algorithmic defense systems but easily identified as non-threats by human operators” (Sawyer *et al.* 2016, p 162). This signal flood raises the noise to signal ratio and thereby compromises operator accuracy by suppressing vigilance capability.

This paper argues that such effects can be enhanced with the introduction of a culturally consistent narrative that induces knowledge-driven cognitive processing within the target leading to self-induced disinformation influence. Such ASID attacks might be used by threat actors as a ‘meta-attack’ apart from, or in parallel with, other information operations. ASID attacks manipulate the target’s understanding of ambiguous stimuli by providing a context through which to interpret those stimuli. The principles underlying ASID attacks have been implicitly understood for some time but have not been explicitly recognized as a potent form of weaponized psychology. These attacks have been, and will be, highly effective because humans are prone toward imposing order on a chaotic environment. For example, imagine that an advanced persistent threat announces that he or she will begin attacking a specific target on a certain day, at a certain time. Armed with this information, network defenders will likely direct more acute focus on network traffic around the expected time of the attack. This increased focus will bias them toward interpreting ambiguous network activity as potentially malicious, even when no malicious activity is occurring. This enables a malicious actor to consume defender bandwidth and resources at almost no cost to the attacker. This tactic is not limited to a cyber defense context.

### **Mass Psychogenic Illness (MPI)—An Introduction**

Nearly every culture throughout history has experienced mass psychogenic illness (MPI) events. These MPI events follow a similar pattern with slight variations. An ambiguous event or environmental stimuli (sight, sound, smell) is encountered, followed shortly after by a reaction in an individual that triggers similar reactions in other individuals within a social network. An early example was the St. Vitus Dance outbreaks of the 14<sup>th</sup> and 15<sup>th</sup> centuries. These episodes involved contagious dancing that would sometimes last for weeks. Processions of dancers would often end at chapels dedicated to the saint (hence the name) with victims lying on the ground from exhaustion and frothing at the mouth (Baloh & Bartholomew 2020). MPI episodes such as this are typically divided into two categories: sensory and motor. The St. Vitus Dance epidemic was an example of a motor MPI. In the 20<sup>th</sup> and 21<sup>st</sup> centuries, sensory MPI events seem to be more common and are often brought about during high stress periods and in response to ambiguous stimuli such as sights, sounds, and smells (Baloh & Bartholomew 2020). For example, in 1989, shortly after Russian authorities deployed an irritant to disperse a crowd, approximately 400 Georgian schoolgirls began experiencing abdominal pain, skin irritation, and burning eyes after hearing rumors that they had also been exposed to the gas. Approximately one year later, several thousand residents of Kosovo experienced hyperventilation, burning sensations, headaches, dizziness, dry mouth, and nausea shortly after rumors circulated of Albania dispersing neurotoxins to poison the local populace; however, no attacks were ever documented or proven to have occurred (Baloh & Bartholomew 2020).

Humans tend to apply a *positive test strategy* (a confirmation bias) when evaluating whether a phenomenon or attribute is applicable to themselves (Cialdini 2016). In applying this strategy, people search their memory for positive instances, examples of something occurring or existing, rather than negative instances or the lack of something occurring or existing. If asked, “have you experienced any of these symptoms?” people will consider episodes from their past when they experienced these, not when they *did not* experience such symptoms. This means that the person asking the question may (intentionally or not) direct the person being asked through a psychological chute. Such *psychological chuting* causes people to focus on instances of a phenomenon’s occurrence and can alter their perception of likelihood and relevance of that phenomenon (Cialdini 2016). This suggests that the introduction of a narrative that accompanies a phenomenon is likely to cause some proportion of a given population to identify with that narrative and thus psychological chuting may very well be induced or facilitate MPI episodes.

Throughout history, MPI episodes have included the following characteristics: 1. an ambiguous situation, event, or environmental stimuli, 2. a culturally consistent narrative, 3. stressful environment, 4. response from the affected group, 5. diffusion of condition across the social network.

1. *Ambiguous Situation, Event, Stimuli:* Across the globe, well over 7,000 cases of persons hear a chronic humming sound, which induces anxiety, irritation, and sometimes painful reactions in those who suffer from the malady (Baloh & Bartholomew 2020). Accused sources of the humming noise have included power lines, wind turbines, telecommunications towers, industry fans, radio waves, birds, insects, and other natural and artificial sources. Reported symptoms include: a vague vibrating sensation, headache, nose pressure, nosebleeds, ear pain, ear popping, chest tightness, insomnia, short-term memory loss, general unwellness, nausea, stabbing pains, and occasionally violent illness (Baloh & Bartholomew 2020). Researchers concluded that sufferers were misattributing common symptoms to the noises rather than being affected by them (Baloh & Bartholomew 2020). Subsequent studies found that, by informing subjects, prior to exposure, that sounds could either cause unpleasant symptoms or pleasant sensations (with therapeutic effects) it was possible to influence their interpretation of the subjective experience of those sounds (Crichton *et al.* 2014; Crichton *et al.* 2015; Baloh & Bartholomew 2020). These studies suggest that ambiguous stimuli (sound in this case) may be interpreted differently depending on the observer’s expectations and cultural context. The intensity of perceived physical symptoms can be enhanced or reduced by providing a coherent narrative modifying expectations. By providing a combination of ambiguous but not noxious thermal stimuli, positive, negative, or emotionally neutral narratives elicited corresponding degrees of physical pain, eliciting a labelling effect that has been termed ‘predictive coding’ (Bräscher *et al.* 2020a). In the context of perceived environmental intolerance, Bräscher and colleagues (2020b) elicited somatic symptoms after sham-exposure to what healthy participants believed to be electromagnetic fields caused by a (sham) Wi-Fi device. The same effect of so-called somatosensory illusions was elicited as predicted by the predictive processing theory of somatic symptom generation following narratively induced attribution patterns and symptom expectations (Wolters *et al.* 2021). One explanation as to why some individuals generate symptoms, whereas others do not, despite being exposed to the same ambiguous stimuli and explanatory narratives, seems to be, at least in part, explained by individual traits related to the cognitive processing of potentially threatening stimuli (Scheuren, Sütterlin & Anton 2014).
2. *Culturally Consistent Narrative:* The 1938 *War of the Worlds* radio broadcast became infamous for the public reaction that it provoked. The broadcast told the story of a Martian inva-

sion of Earth, portrayed as though it was being reported in real time through a news broadcast. This radio show played on cultural narratives that were popular at the time: that Mars might be inhabited and that there was a looming war in Europe. Large numbers of audience members who tuned into the program after the disclosure announcement, were unaware that this was in fact a radio play and not an actual news flash. A few people panicked to the point of confusing water towers for the three-legged Martian vehicles described in the program. The key point of this example is that the narrative of extraterrestrial visitation is culturally consistent for scientifically-oriented western cultures, while magically oriented cultures experience more instances of demonic possession. Collective narratives provide opportunities for knowledge-driven processing. They allow for interpretation (perceiving meaningful patterns instead of chaos) of the status quo, allow for projections into the future (what-ifs), and thus provide the basis for perceived control, resulting in stress alleviation (storytelling and narrative transportation). The stress alleviating effects of individual narratives are well established (Petko *et al.* 2015), and on a societal level, collective narratives can unfold a comparable effect of providing meaning to group members after the injection of ambiguous information into a social or socio-technical system. The collective sensemaking following such a grey signal injection provides an analogue to knowledge-based perceptual closure following the eradication of ambiguity. The collective narrative further maintains social bonds and group-level cooperation, which in turn contributes to reduced anxiety (see Bietti, Tilston & Bangerter 2019). Collective narratives have an inherent self-stabilizing effect by providing social proof to other individuals. The perceived state of the collective can also have measurable and potent biological effects on the individual. Perception of events related to the victory or defeat of a social group one considers oneself a member of (for example, as a spectating supporter of a football team or political party) increases and decreases testosterone levels, respectively (Bernhardt *et al.* 1998; Stanton *et al.* 2009), more than most own behaviors. This exemplifies how mere perceptions related to the current state of the collective may biologically manifest in the individual in powerful ways that could have downstream influences on sensations such as pain (Hau, Dominguez & Evrard 2004), emotional regulation and cognitions (Celec, Ostatníková & Hodosy 2015; Martin *et al.* 2009), and behaviors (for example, Dreher *et al.* 2016).

3. *Stressful Environment*: Throughout history, the majority, if not all, MPI events occur during times of stress for the affected group members. The 1990 MPI outbreak in Kosovo occurred at a time when rising ethnic tensions would eventually lead to the outbreak of civil war and Balkanization of the former Yugoslavia. The 1989 MPI episode in Georgia occurred amidst heightened tensions between Russia and Georgia during the Soviet Union disillusion. Even the earlier St. Vitus Dance MPI epidemics occurred on the footholds of the Black Death devastation of Europe (Baloh & Bartholomew 2020). The stressful tension of an era appears to provide the fertile ground from which an MPI might flourish. Both short-term and especially long-term stress can have several negative effects on cognitive information processing (Porcelli & Delgado 2017; Provenzano *et al.* 2019). This is mainly a consequence of the effect that acute and chronic stressors have on prefrontal cortical function (Maeng & Shors 2013; Snyder, Barry & Valentino 2015) which is partly resulting from stress-induced hormonal interactions between the hypothalamus and prefrontal cortices that promote executive dysfunction (Uribe-Mariño *et al.* 2016). Concomitantly, the salience and persistence of prior or current stressors could increase the perceived likelihood that future stressors will occur due to a readiness to interpret stimuli in a manner congruent with one's psychological state (Andersen 1990; Ander-

sen & Schwartz 1992). On a neurocognitive level, negative affect decreases the efficiency of executive processing in favor of increased efficiency in mood-congruent salience processing of stimuli, meaning that the individual focuses on stimuli or an aspect of stimuli that corresponds with his or her existing negative mood (a form of negativity bias) (Provenzano *et al.* 2019). A stress-induced reduction in executive function that favors mood-congruent information processing, combined with negative expectations and a positive testing strategy, may shape sensory experiences in ways that facilitate individual-level psychogenic illness through predictive coding. It has been shown that expectations shape the encoding of perceptual stimuli at very early stages in perceptual processing (for example, Aitken, Turner & Kok *et al.* 2020; Panichello & Turk-Browne 2021), suggesting a mechanistic basis for how an interplay between biasing top-down mood-congruent processes and exposure to ambiguous stimuli can produce MPI events. Once an individual witnesses or starts to experience somatic symptoms, interpersonal stress-related processes such as co-rumination may promote vigilance for symptom onset in other individuals or even aid in the spreading, exacerbating, and maintenance of symptoms (Carlucci *et al.* 2018; Hankin, Stone & Wright 2010; Schwartz-Mette & Rose 2012). Co-rumination processes may additionally facilitate the continued hypothalamic-prefrontal cortex interactions that promote executive dysfunction and mood-congruent processing as well as a sustained failure to adapt to the MPI-associated stressors (Gianferante *et al.* 2014; Uribe-Mariño *et al.* 2016). Additional effects of executive dysfunction may include increased stress-induced immune responses that contribute to general feelings of being unwell (Ask, Lugo & Sütterlin 2018; Moriarity *et al.* 2019; Williams *et al.* 2019). These feelings may in turn be misattributed as symptoms (thus, proof) of illness. Together, the above neurocognitive and socio-cognitive factors demonstrate how stressful environments can interact with human cognition to produce MPI events.

4. *Response from the Affected Group:* A common element also present in most MPI episodes is the appearance of an Index Case, the first case which sets the expectations of those who follow. In the mid-1960s several African countries experienced MPI outbreaks of ‘Laughing Mania’. These outbreaks primarily afflicted young schoolgirls presenting uncontrollable laughing as well as episodes of screaming, twitching, trance states, and weeping. The pattern of contagion was for one or more of the afflicted to begin laughing uncontrollably and for this to spread to others. People in other villages became afflicted with the disorder after news of the illness spread. A common theme across different MPI outbreaks of the initial case set the trend of the symptomology for the subsequent cases. As mentioned in the Canadian study of sound sufferers, knowledge of the expression of symptoms may lead subsequent afflictees to attend more closely to otherwise normal sensations and thus attribute these to the ‘illness’. Curiously, this disorder did not seem to spread to adult populations, but remained primarily contained within school children, particularly schoolgirls, suggesting a social network component to the spread of MPI outbreaks.
5. *Diffusion across the Network:* In the early 1930s, several towns in rural Virginia reported a series of gas attacks committed by elusive culprits who always evaded capture and even left physical evidence of their presence. The index case was a family outside Roanoke County, Virginia, where it was reported on December 23, 1933, that someone had attempted to inject poisonous gas into his or her family’s residence. Police investigations failed to find a suspect or to identify any physical evidence to support the claim. After the story appeared in local newspapers, reports of similar incidents proliferated throughout the Virginia countryside. The

reports of these incidents skyrocketed after reporting but seem to have been localized to areas where the news was reported. In MPI episodes involving school children, it is common for children to become afflicted but not the adults. Similar trends were observed with the Laughing Mania outbreaks in Africa, suggesting once again that a social network component may contribute to the spread of MPI outbreaks.

Weaponizing the mechanics that drive MPI events could be categorized as a grey-zone tactic. Characterized by ambiguity, opacity of the actors involved, uncertainty concerning the relevant policy, and legal frameworks to address this attack vector are representative of the nature and form of this contemporary conflict space and reflect the challenges facing the security community attempting to understand, deter, and respond to it (Votel 2015). The significance and novelty of ASID attacks is the injection of ambiguous information as the primary, or peripheral, attack vector. The effects potentially transcend the information and psychological warfare domains to induce self-inflicted physical harm in targets.

### **Elements of an ASID Attack**

Examining the components of MPI outbreaks provides clues to the elements required to conduct a successful ASID attack. MPIs involve some ambiguous stimuli, either environmental or tied to an event. This ambiguous stimulus begs for resolution (Cialdini 2016). The next component, a stressful environment or circumstance, encourages hypervigilance and thus individuals begin attending closely to the ambiguous stimuli. Individuals within the social network then notice responses from affected individuals in the form of physical sensations. These sensations are often commonly experienced within a healthy population but are ignored or dismissed under ‘normal’ circumstances. Finally, the psychogenic condition becomes diffused across the social network. Drawing from MPI events as a guide, the authors assert that the following elements are necessary to successfully launch an ASID attack:

*Introduction of a Culturally Consistent Narrative:* Injecting a plausible narrative into the communication stream of the target population constitutes the required trigger to cause that population to engage in a positive test strategy to hyper-focus on ambiguous stimuli and interpret these through the lens of the weaponized narrative.

*The Presence of Ambiguous Stimuli:* The presence of ambiguous stimuli provides the circumstances that the attack(s) exploit to mire the target in the information noise of innocuous and irrelevant information. The search for positive instances will encourage the target to discover false positives that will consume time and resources.

*Motivation for Hypervigilance:* The target needs to have a reason to care about examining the ambiguous stimuli. Being exposed to a constant stream of ambiguous and (mostly) irrelevant information, humans lack the motivation for hypervigilance, unless a compelling reason exists. When an individual is suddenly compelled to direct intense focus on this ambiguous information stream, he or she will impose meaning onto that information. This hypervigilance can lead targets to interpret their experiences through cognitive modelling and experience formations. Such weaponization relies on cognitive processes that have been shown to be both implicit and self-perpetuating. Memory implanting has been vastly researched and demonstrates that individuals can create experiences or memories through suggestion (see Roediger & McDermott 1995; Frenda, Nichols & Loftus 2011). The previously described factors, alongside the accessibility of ambiguous infor-



mation, as well as actual reports of symptoms found in the media, would contribute to individuals incorrectly interpreting their experiences and creating false memories. These experiences are then shared, both verbally and non-verbally with others—causing a social contagion that would ‘infect’ others. Studies have shown that individuals can create experience and memories that lead to group experiences and collective memories, whereas individuals that experienced negative valence created more collective negative experiences, and this effect is stronger in smaller groups (Choi, Kensinger & Rajaram 2017).

*Optional Social Network:* ASID attacks may infect an individual or an entire social network, but when targeted against a social network, the additional biases of *social proof* and *confirmation bias* are introduced and reinforced respectively. Most leading social media platforms employ algorithms that propagate content that increases engagement among its users, including content that induces strong negative emotions. These algorithms are not subject to any governmental or judicial regulations that inhibit their efficiency, thus offering a cost-effective and built-in ‘infrastructure’ for ASID attack propagation that requires minimal effort or understanding on the side of the adversary.

### **Potential Attack Examples**

Viewing the *War of the Worlds* broadcast response from the advantage of historical hindsight, the reactions of some of the audience may seem ridiculous. While some of the reported reactions may have been exaggerated, it is important to consider audience responses within the context of the era. This broadcast occurred on the eve of World War II breaking out and people sensed that war was on the horizon (Baloh & Bartholomew 2020). A case study of the event found that a significant portion of the listeners thought the ‘attacks’ were real, but that the reports were mistaking an attack by the German military as an alien invasion (Cantril 1947). Imagine that instead of being a publicity stunt by the radio network, this *War of the Worlds* broadcast was an information operation launched by the German military against the United States. This injection of information would superficially appear to be ambiguous or even benign to military officials; however, this ambiguity in fact would have played on the fears that many held about an impending war. The themes of the broadcast included invasion, poison gas dispersal, and air attacks while also including ambiguous information about the aliens themselves, such as the tripod vehicles that resembled water towers. Reports of residents shooting at these water towers were never confirmed and are suspected to have involved people from out of the area. If these reports are true, this would illustrate how real and serious physical harm could have resulted from the imposition of a harmful narrative onto ambiguous stimuli, in this case a three-legged water tower at night.

Considered within the historical context, it is easy to understand how an ASID attack might be deployed by an adversary to degrade or disrupt an organization’s ability to function. If the conditions are set (for example, the culturally consistent narrative, the presence of ambiguous stimuli, motivation for hypervigilance, and if necessary, the accelerator function of social networks) then the ambiguity and fear they spread act as a multiplier in the target group. Victims will complete the ‘play’ themselves, as the real-power mechanics driving MPI events are weaponized as a component of an information warfare campaign targeting human cognition.

*The Telegraphed Attack:* An advanced adversary who has established credibility through prior attacks and reputation may be capable of disrupting the operations of a target, or cause victims to use a significant number of resources, by merely threatening or telegraphing an attack. If their primary goal is the degradation of operational capability or distraction, then an ASID attack might be an

effective option. In this scenario, perpetrators of an advanced persistent threat might announce that they plan to launch an attack within a prescribed timeframe. If they are successful, information security staff from that organization will likely be more vigilant than their typical baseline during around the period of the announced attack. If this is a large organization, with virtual or distributed team and individual, then this attack might be enhanced by impersonating organizational members on social media. These impersonated individuals could post disinformation regarding false Indicators of Compromise (IoCs). To maximize effectiveness, these IoCs should describe ambiguous network traffic. This tactic could be highly effective because the injection of cognitive malware through social media channels would unlikely be anticipated and discovered in time to prevent the damage already accomplished by inducing hypervigilance within some operators.

*Social Media Attack:* Social media has provided a pathway for ASID attacks to propagate with little effort. Hot button topics are constantly enflamed on social media, which are exacerbated by misinformation and disinformation. An adversary could create culturally consistent content (stimuli) that is meant to mislead a hypervigilant population. The disinformation is amplified through social shares, which spread the content to an exponentially expanding network of social media users. This type of attack could be used to change opinion of voters in a contentious battleground state during an election period. A plausible scenario might involve a threat actor bringing attention to an ambiguous statement by a candidate about a hot button topic which causes a whirlwind of shares across the social media platform. By recruiting and inducing collective false memory within the targeted population, this ASID attack might cause a significant waste of time and resources to counter the effect and could lead to loss of confidence in election results. For instance, thousands of Russian-operated Twitter accounts posing as U.S. citizens have, since 2014, involved themselves in the public vaccine discourse while contributing with posts that are for, against, and neutral to vaccines (Bail *et al.* 2020; Broniatowski *et al.* 2018; Walter, Ophir & Jamieson 2020). In addition to spreading mis- and dis-information, many accounts will write posts that artificially conflate vaccine topics with topics that are inflammatory in the U.S., such as those related to skin color, socioeconomic disparities, veganism, gender inequality, and politics. This may be an attempt to make individuals process these topics in a conflated and mood- and narrative-congruent manner. If successful, ambiguous posts or news stories with incomplete information related to any of the topics may be processed according to a pattern of causal attribution that fits the narrative of specific groups. Some Russian-operated Twitter accounts will ask open-ended questions related to these topics (for example, “I believe in #vaccines, why don’t you? #VaccinateUS” (Broniatowski *et al.* 2018, pp. 1382). In absence of any supporting arguments, it is up to the reader to infer the viewpoints of the original poster according to his or her own narrative view, and it is also up to the reader to produce the arguments for why his or her narrative view of the topic is correct, thus convincing him or her that the attribution patterns are valid. In the scenario that a candidate for election expresses an ambiguous statement about a topic, that statement might be conflated with having specific opinions about other inflammatory topics.

*Hoax Security Warning:* It is common for security departments to send security warning emails to employees to caution against malicious email campaigns they have observed or other recent scams. A malicious threat actor might leverage these warnings to launch an ASID attack against the members of an organization by impersonating the security department in an email advising against a malware campaign being perpetrated against the organization. This email could contain descriptions of anomalous computer operations that in fact are normal but would appear as potentially malicious. Internet hoaxes such as these have already been observed in the wild with claims

that users should look for the presence of the ‘SULFNBK.exe’ within their Windows C: drive directory. The presence of this file would presumably indicate that their machine had been compromised, and that users should delete it from their system. In fact, this file was part of the normal operation of the machine and deleting it could potentially cause a system malfunction (AOL.com 2015). A recent study on college student responses to phishing emails found that, rather than reporting malicious emails to the university security department, some students posted screen shots of emails to subreddit forums (M. Constantino, personal communication, October 2, 2020). This presents another potential ASID attack vector, in which malicious actors could pose as students posting a legitimate email as a scam.

*False Bio-Attack:* A variation of the above scenario would be to impersonate someone from an organization with the claim that a biological pathogen was released within a facility and that several employees have become ill as a result. The symptoms should describe commonly experienced symptoms such as headaches, dizziness, fatigue, hyperventilation, dry mouth, and nausea. To increase effectiveness by inducing hypervigilance, the hoax would then suggest that anyone experiencing these symptoms should immediately quarantine themselves and contact his or her doctor, as the symptoms often rapidly increase in severity. This quarantine recommendation would serve two purposes. First, it would degrade operational capability, and second, it would provide an observable signal to other employees which might lead to additional ‘cases’, as has been observed in historical MPI events. The inducement of hypervigilance will statistically guarantee (in a large enough organization) that some individuals will experience these symptoms and associate them with the hoax attack. A variation of this hoax might describe exposure to electromagnetic radiation as the cause with similar symptoms and recommended course of action.

## **Future Directions**

Due to attack vectors relying on a combination of individual cognitive deficiencies, social processes, population-specific narratives, and uncertainties and ambiguities that are inherent to the cyber domain, defending against ASID attacks is a complex and interdisciplinary issue. Two major methodologies emerge for future research: development of a model for ASID attacks and designing training that focused on ASID resilience.

An ASID attack model could be leveraged to aid human analysts in recognizing an emerging or ongoing potential attack. This model may incorporate monitoring of current events, ongoing activities in certain regions, background information that could provide insight on how a community may react, and social media monitoring activities. This type of model may expose elements of an ASID attack that would typically be overlooked in the noise of today’s environment. The creation of a model to aid human analysts would promote rapid awareness of a potential or ongoing ASID attack and could also be used to train newer human analysts about what signals to focus on for better recognition.

Another direction would be to design ASID resilience training, which would at the very least require 1) awareness of which elements in an organization or a society that can be targeted by an ASID attack (for example, how shared narratives shape how a population makes causal attributions in a narrative-congruent manner), 2) knowledge of individual-level cognitive vulnerabilities that make them vulnerable targets and which skills will mitigate those vulnerabilities, and 3) organizational/societal-level implementation of both individualized and group-based training efforts. Developing efficient training programs for improving cognitive skills and increasing awareness

of and resilience to ASID attacks will require further research. Current research aiming to improve cognitive skills for situation awareness, communication, and decision-making during cyber threat situations indicate that cognitive abilities such as metacognition, self-regulation, and critical thinking improve human performance in such contexts (Jøsok *et al.* 2019; Knox *et al.* 2017, 2018; Massey *et al.* 2021). These abilities can be improved with training and may be potential topics for initial research efforts. Designing research projects that can model ASID attack scenarios with sufficient realism may require collaboration between neuroscientists, cognitive scientists, defense and intelligence community professionals, technologists, and social scientists. While collaborations allowing for correlational studies are becoming increasingly common in the literature, a recent review suggested that there was a general lack of experimental collaborations allowing for manipulation of key variables (Ask *et al.* 2021). Overcoming this hurdle may be one of several first important steps toward developing optimal defense strategies against ASID attacks.

## **Conclusion**

The long and well-documented history of Mass Psychogenic Illness (MPI) outbreaks suggests the potential effectiveness of injecting ambiguous information with a coherent narrative into the information stream as a means for causing disruption. This form of attack exploits normal cognitive functions to cause widespread disruption. Information warriors should be aware of the possibility of the employment of such tactics in grey-zone conflict scenarios as these attacks would reside below the threshold of overt military actions. For the actors and agencies who use time and energy to investigate, understand, resolve, explain, and potentially respond to ASID attacks, they must recognize the strategic validity of this approach and its capability to change how the victim, and they themselves, think and behave. Efforts should be made to better understand the key factors necessary to successfully launch such attacks and how to mitigate or disrupt their effects.

## **References**

- Aitken, F, Turner, G & Kok, P 2020, 'Prior expectations of motion direction modulate early sensory processing', *Journal of Neuroscience*, vol. 40, no. 33, pp. 6389-97.
- America Online (AOL.com) 2015, *Sulfnbk.exe Hoax Virus information*, viewed 23 September 2021, <<https://archive.is/GOnLX>>.
- Andersen, S & Schwartz, A 1992, 'Intolerance of ambiguity and depression: A cognitive vulnerability factor linked to hopelessness', *Social Cognition*, vol. 10, no. 3, pp. 271-98.
- Andersen, S 1990, 'The inevitability of future suffering: The role of depressive predictive certainty in depression', *Social Cognition*, vol. 8, no. 2, pp. 203-28.
- Ask, T, Lugo, R, & Sütterlin, S 2018, 'The neuro-immuno-senescence integrative model (NISIM) on the negative association between parasympathetic activity and cellular senescence', *Frontiers in Neuroscience*, vol. 12.
- , R, Knox, B & Sütterlin, S 2021, 'Human-human communication in cyber threat situations: A systematic review', *HCI International 2021 - Late Breaking Papers: Cognition, Inclusion, Learning, and Culture—HCII 2021*, eds. C Stephanidis, D Harris, W-C Li, D Schmorow, C Fidopiastis, M Antona, Q Gao, J Zhou, P Zaphiris, A Ioannou, R Sottolare, J Schwarz, M Rauterberg, LNCS, vol. 13096, pp. 21-43.

Bagge, D 2019, *Unmasking Maskirovka: Russia's Cyber Influence Operations*, Defense Press, New York, NY, US.

Bail, C, Guay, B, Maloney, E, Combs, A, Hillygus, D, Merhout, D, Freelon, D & Volfovsky, A 2020, 'Assessing the Russian Internet Research Agency's impact on the political attitudes and behaviors of American Twitter users in late 2017', *Proceedings of the National Academy of Sciences*, vol. 117, no. 1, pp. 243-50.

Baloh, R & Bartholomew, R 2020, *Havana Syndrome: Mass psychogenic illness and the real story behind the embassy mystery and hysteria*, Springer Nature Switzerland AG, Cham, CH.

Bernhardt, P, Dabbs, J (Jr), Fielden, J & Lutter, C 1998, 'Testosterone changes during vicarious experiences of winning and losing among fans at sporting events', *Physiology & Behavior*, vol. 65, no. 1, pp. 59-62.

Bietti, L, Tilston, O & Bangerter, A 2019, 'Storytelling as adaptive collective sensemaking', *Topics in Cognitive Science*, vol. 11, no. 4, pp. 710-32.

Broniatowski, D, Jamison, A, Qi, S, AlKulaib, L, Chen, T, Benton, A, Quinn, S & Dredze, M 2018, 'Weaponized health communication: Twitter bots and Russian trolls amplify the vaccine debate', *American Journal of Public Health*, vol. 108, no. 10, pp. 1378-84.

Bräscher, A, Schulz, S, Van den Bergh, O & Witthöft, M 2020a, 'Prospective study of nocebo effects related to symptoms of idiopathic environmental intolerance attributed to electromagnetic fields', *Environmental Research*, vol. p. 190, 110019.

Bräscher, A, Sütterlin, S, Scheuren, R, Van den Bergh, O & Witthöft, M 2020b, 'Somatic symptom perception from a predictive processing perspective: An empirical test using the Thermal Grill Illusion', *Psychosomatic Medicine*, vol. 82, no. 7, pp. 708-14.

Cantril H, 1947, *The invasion from Mars: A study in the psychology of panic*, Princeton University Press, Princeton, NJ, US.

Carlucci, L, D'Ambrosio, I, Innamorati, M, Saggino, A & Balsamo, M 2018, 'Co-rumination, anxiety, and maladaptive cognitive schemas: When friendship can hurt', *Psychology Research and Behavior Management*, vol. 11, pp. 133-44.

Cavanagh, P 1991, 'What's up in top-down processing?', *Representations of Vision: Trends and Tacit Assumptions in Vision Research*, ed. A Gorea, Cambridge University Press, Cambridge, UK, pp. 295-304.

Celec, P, Ostatníková, D & Hodosy, J 2015, 'On the effects of testosterone on brain behavioral functions', *Frontiers in Neuroscience*, vol. 9.

Choi, H, Kensinger, E & Rajaram, S 2017, 'Mnemonic transmission, social contagion, and emergence of collective memory: Influence of emotional valence, group structure, and information distribution', *Journal of Experimental Psychology*, vol. 146, no. 9, pp. 1247-65.

Cialdini, R, 2016, *Pre-suasion: A revolutionary way to influence and persuade*, Simon & Schuster, New York, NY, US.

Constantino, M 2020, Personal communication with Matthew Canham, 2 October.

Cootner, P 1964, *The random character of stock market prices*. MIT Press, Cambridge, MA, US.

Crichton, F, Dodd, G, Schmid G, Gamble G, Cundy T & Petrie, K 2014, 'The power of positive and negative expectations to influence reported symptoms and mood during exposure to wind farm sound', *Health Psychology*, vol. 33, no. 12, pp. 1588-92.

—, Dodd, G, Schmid, G & Petrie, K 2015, 'Framing sound: sing expectations to reduce environmental noise annoyance', *Environmental Research*, vol. 142, pp. 609-14.

Dreher, J, Dunne, S, Pazderska, A, Frodl, T, Nolan, J & O'Doherty, J 2016, 'Testosterone causes both prosocial and antisocial status-enhancing behaviors in human males', *Proceedings of the National Academy of Sciences*, US, vol. 113, no. 41, pp. 11633-38.

Fama, E 1970, 'Efficient capital markets: A review of theory and empirical work', *Journal of Finance*, vol. 25, no. 2, pp. 383-417.

Fitton, O 2016, 'Cyber operations and gray zones: Challenges for NATO', *Connections*, vol. 15, no. 2, pp. 109-19.

Frenda, S, Nichols, R & Loftus, E 2011, 'Current issues and advances in misinformation research', *Current Directions in Psychological Science*, vol. 20, no. 1, pp. 20-3.

Gianferante, D, Thoma, M, Hanlin, L, Chen, X, Breines, J, Zoccola, P & Rohleder, N 2014, 'Post-stress rumination predicts HPA axis responses to repeated acute stress', *Psychoneuroendocrinology*, vol. 49, pp. 244-52.

Goolkasian, P 1987, 'Ambiguous figures: Role of context and critical features', *Journal of General Psychology*, vol. 114, no. 3, pp. 217-28.

Hankin, B, Stone, L & Wright, P 2010, 'Co-rumination, interpersonal stress generation, and internalizing symptoms: Accumulating effects and transactional influences in a multiwave study of adolescents', *Development and Psychopathology*, vol. 22, no. 1, pp. 217-35.

Hau, M, Dominguez O, Evrard H 2004, 'Testosterone reduces responsiveness to nociceptive stimuli in a wild bird', *Hormones and Behavior*, vol. 46, no. 2, pp. 165-70.

Intaite, M, Georgescu, A, Noreika, V, von Saldern, M, Vogeley, K & Falter-Wagner, C 2019, 'Adults with autism spectrum condition have atypical perception of ambiguous figures when bottom-up and top-down interactions are incongruous', *Autism: The International Journal of Research and Practice*, vol. 23, no. 5, pp. 1133-42.

Ishizu, T 2013, 'Disambiguation of ambiguous figures in the brain', *Frontiers in Human Neuroscience*, vol. 7.

Jøsok, Ø, Lugo, R, Knox, B, Sütterlin, S & Helkala, K 2019, 'Self-regulation and cognitive agility in cyber operations', *Frontiers in Psychology*, vol. 10, p. 875.

Knox, B, Jøsok, Ø, Helkala, K, Khooshabeh, P, Ødegaard, T, Lugo, R & Sütterlin, S 2018, 'Socio-technical communication: The hybrid space and the OLB model for science-based cyber education', *Military Psychology*, vol. 30, no. 4, pp. 350-9.

Knox, B, Lugo, R, Jøsok, Ø, Helkala, K & Sütterlin, S 2017, 'Towards a cognitive agility index: The role of metacognition in human computer interaction', *HCI International 2017 – Posters Extended Abstracts*, pp. 330-8.

Kornmeier, J & Bach, M 2012, 'Ambiguous figures—what happens in the brain when perception changes but not the stimulus', *Frontiers in Human Neuroscience*, vol. 6.

Maeng, L & Shors, T 2013, 'The stressed female brain: Neuronal activity in the prelimbic but not infralimbic region of the medial prefrontal cortex suppresses learning after acute stress', *Frontiers in Neural Circuits*, vol. 7.

Malkiel, B 2003, 'The efficient market hypothesis and its critics', *Journal of Economic Perspectives*, vol. 17, no. 1, pp. 59-82.

Martin, D, Burns, N & Wittert, G 2009, 'Free testosterone levels, attentional control, and processing speed performance in aging men', *Neuropsychology*, vol. 23, no. 2, pp. 158-67.

Massey, L, Smith, R, Whitaker, E & Wray R 2021, 'Designing learning experiences to encourage development of critical thinking skills. *Human Computer Interaction International 2021*, July 24-29, Virtual.

Moriarty, D, Ng, T, Titone, M, Chat, I, Nusslock, R, Miller, G & Alloy, L 2019, 'Reward responsiveness and ruminative styles interact to predict inflammation and mood symptomatology', *Behavior Therapy*, vol. 51, no. 5, pp. 829-42.

Panichello, M & Turk-Browne, N 2021, 'Behavioral and neural fusion of expectation with sensation', *Journal of Cognitive Neuroscience*, vol. 33, no. 5, pp. 814-25.

Petko, D, Egger, N, Schmitz, F, Totter, A, Hermann, T & Guttormsen, S 2015, 'Coping through blogging: A review of studies on the potential benefits of weblogs for stress reduction', *Cyberpsychology: Journal of Psychosocial Research on Cyberspace*, vol. 9, no. 2.

Porcelli, A & Delgado, M 2017, 'Stress and decision making: Effects on valuation, learning, and risk-taking', *Current Opinion in Behavioral Sciences*, vol. 14, pp. 33-9.

Provenzano, J, Verduyn, P, Daniels, N, Fossati, P & Kuppens, P 2019, 'Mood congruency effects are mediated by shifts in salience and central executive network efficiency', *Social Cognitive and Affective Neuroscience*, vol. 14, no. 9, pp. 987-95.

Roediger, H & McDermott, K 1995, 'Creating false memories: Remembering words not presented in lists', *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol. 21, no. 4, pp. 803-14.

Sawyer, B, Finomore, V, Funke, G, Warm, J, Matthews, G & Hancock 2016, 'Cyber vigilance: The human factor', *American Intelligence Journal*, vol. 32, no. 2, pp. 157-65.

Scheuren, R, Sütterlin, S & Anton, F 2014, 'Rumination and interoceptive accuracy predict the occurrence of the thermal grill illusion of pain', *BMC Psychology*, vol. 2, no. 1, pp. 1-15.

Schwartz-Mette, R & Rose, A 2012, 'Co-rumination mediates contagion of internalizing symptoms within youths' friendships', *Developmental Psychology*, vol. 48, no. 5, pp. 1355-65.

Snyder, K, Barry, M & Valentino, R 2015, 'Cognitive impact of social stress and coping strategy throughout development', *Psychopharmacology*, vol. 232, no. 1, pp. 185-95.

Stanton, S, Beehner, J, Saini, E, Kuhn C & Labar K 2009, 'Dominance, politics, and physiology: Voters' testosterone changes on the night of the 2008 United States presidential election', *PLOS One*, vol. 4, no. 10, p. 7543.

Theeuwes, J 2010, 'Top-down and bottom-up control of visual selection', *Acta Psychologica*, vol. 135, no. 2, pp. 77-99.

Uribe-Mariño, A, Gassen, N, Wiesbeck, M, Balsevich, G, Santarelli, S, Solfrank, B, Dournes, C, Fries, G, Masana, M, Labermeier, C & Wang, X 2016, 'Prefrontal cortex corticotropin-releasing factor receptor 1 conveys acute stress-induced executive dysfunction', *Biological Psychiatry*, vol. 80, no. 10, pp. 743-53.

Votel, J 2015, 'The Gray Zone', *White Paper*, USSOCOM, Tampa, FL, US.

Walter, D, Ophir, Y & Jamieson, K 2020, 'Russian Twitter accounts and the partisan polarization of vaccine discourse, 2015-2017', *American Journal of Public Health*, vol. 110, no. 5, pp. 718-24.

Williams, D, Koenig, J, Carnevali, L, Sgoifo, A, Jarczok, M, Sternberg, E & Thayer, J 2019, 'Heart rate variability and inflammation: A meta-analysis of human studies', *Brain, Behavior and Immunity*, vol. 80, pp. 219-26.

Wolters, C, Harzem, J, Witthöft, M, Gerlach, A & Pohl, A 2021, 'Somatosensory illusions elicited by sham electromagnetic field exposure: Experimental evidence for a predictive processing account of somatic symptom perception', *Psychosomatic Medicine*, vol. 83, no. 1, pp. 94-100.