May 2025

IW PERSPECTIVES

Emerging Cognitive Threats

By Rand Waltzman – Adjunct Senior
Information Scientist – RAND Corporation

Cognitive Warfare can be functionally defined as "the activities conducted in synchronization with other instruments of power, to affect attitudes and behaviors by influencing, protecting, and/or disrupting individual and group cognitions to gain an advantage." From the point of view of technique, we can say that "Cognitive Warfare is therefore the art of deceiving the brain or making it doubt what it thinks it knows." Deception is no less than an attack on cognitive abilities. In discussions of Cognitive Warfare, emotions are often described as limitations to cognitive abilities that must be overcome. However, research of the last several decades suggests that rather than being a limitation to cognitive abilities, emotions are integral to them. This in turn suggests that a major attack vector against cognitive abilities is through emotional manipulation. For example: "Neuroscientific and psychological research is now clear: No reasonable or rational decision-making is possible without emotions. Therefore, emotions are necessary for effective decision-making, including in war." So, to compromise an adversary's decision-making capability, one approach is to appropriately manipulate and compromise the emotions integral to decision-making.

Concepts of Cognitive Warfare draw heavily from the field of cognitive science. The neuroscientist Joseph LeDoux said of cognitive science that "a pure cognitive approach, one that omits consideration of emotions, motivations, and the like, paints an artificial, highly unrealistic view of real minds. Minds are not either cognitive or emotional, *they are both*, and more. Inclusion of work on emotion within the cognitive framework can help rescue this field from its sterile approach to the mind as an information-processing device that lacks goals, strivings, desires, fears, and hopes."

Much recent discussion of Cognitive Warfare places a great emphasis on Neuroscience and Technology and justifiably so. However, there is another set of technologies on the horizon that needs to be brought to the forefront of any discussion of Cognitive Warfare. Virtual Reality Environments (VREs), like Meta's Metaverse, of today and in the near future, do and will offer

The views expressed in these articles are those solely of the authors and do not reflect the policy or views of the Irregular Warfare Center, Department of Defense, or the U.S. Government.

tremendous capabilities for cognitive attacks based on emotional manipulation. We will have to learn to recognize and mitigate such attacks if our forces are to achieve and maintain cognitive superiority when operating in such environments. In addition to drawing on the already recognized areas of Human Factors and Medicine (HFM), Information Systems Technology (IST), System Analysis and Studies (SAS) and the Modelling and Simulation Group (NMSG) to understand the potential of VREs, we must look to the field of psychology.

In the last several decades, numerous experiments and psychotherapy development have clearly demonstrated the capabilities of even today's primitive VREs for purposes of emotional manipulation. VREs have been shown to be effective at inducing negative emotions such as fear and anxiety. For example, previous work has demonstrated the treatment of phobias using VREs to elicit fear in patients in a controlled way. While the ability to elicit fear could be used as part of therapy, it can also be used to make a target audience susceptible to harmful messaging like false rumors or conspiracy theories.

Imagine a scenario where a political candidate is giving a speech to millions of people in a VRE. While each viewer thinks they are seeing the same version of the candidate, they are each seeing a slightly different version. The difference is that for each and every viewer, the candidate's face has been subtly modified to resemble the viewer. This is done by blending features of each viewer's face into the candidate's face in such a way that the viewer is consciously unaware of any manipulation of the image. The result is that each viewer is more favorably disposed to the candidate than they might have been otherwise. It has long been known that mimicry can be exploited as a powerful tool for influence.

A series of studies from Stanford University has shown that slightly changing the features of a political figure to resemble each voter in turn would result in a significant influence on how people voted. The experiments took pictures of study participants and real candidates in a mockup of an election campaign. The pictures of each candidate were modified to resemble each participant in turn. They found that if 40% of the participant's features were blended into the candidate's face, the participants were entirely unaware the image had been manipulated. Even so, the blended picture significantly influenced the intended voting result in favor of the candidate. In one experiment, they demonstrated a 20% shift in voting between a control group that did not see the doctored images compared to a group that did. In a VRE, this type of mimicry effect can be produced at a massive scale.

At the heart of all deception, disinformation (a tool of deception) and influence operations is emotional manipulation. VREs, like Facebook's (now Meta's) Metaverse, will enable psychological and emotional manipulation and privacy violations of its users at a level unimaginable in today's communications media: social media, film, television and print. Malicious actors will take the arts of deception and influence to new heights—or depths, depending on one's point of view. As of this writing, we are not able to defend users against the threats posed either technically or legally.

According to the Encyclopedia of Multimedia, "Virtual Reality (VR) is the technology that provides almost real and/or believable experience in a synthetic or virtual way. The goal of Immersive VR is to completely immerse the user inside the computer-generated world, giving the impression to the user that he/she has 'stepped inside' the synthetic world." This definition seems straightforward and innocent enough. However, one can exploit the very same features that make VREs so attractive as communication environments to cause harm to their users. When it comes to emotional manipulation, two features of VREs are particularly important—presence and embodiment. The idea of presence is that participants feel they are communicating with one another directly without any type of computer interface. This is sometimes referred to as "the illusion of non-mediation." The sense of presence created by a VRE is an important element in eliciting strong emotional responses. The idea of embodiment is that the user has the feeling that his avatar or virtual body—his representation in the VRE—is his actual body.

So, how do presence and embodiment enable emotional manipulation? To understand this, we turn to the theory of embodied cognition. According to this theory, cognition depends on having a body that has certain perceptual and motor capabilities that are essential to reasoning, memory, language and emotion. Our bodily experiences and processes are the basis for understanding our own emotions as well as the emotions and intentions of others. Emotions are formed from multi-modal inputs. Body language is an important part of our communication. It includes a variety of nonverbal signals like eye gaze, gestures and facial expressions that we use to communicate intentions and emotions. Unlike verbal language, we often produce and perceive body language subconsciously, but it makes up a significant part of our communication. Our perceptual capabilities are not just for communicating with other people; they also allow us to notice a broad range of emotions and intents from our environments.

Combining the ideas of embodied cognition with the VRE features of presence and embodiment allows interaction between people and between people and environments that exploits the full range of human communication capabilities. Interaction at this level has not been possible in traditional social media environments (and is not yet available in current VREs although it will likely be available in the two to five-year future). That, of course, is both good news and terrible news. Good, because it allows for better communication. Terrible because it opens users not only to the full range of deceptive, dis-informative and influence techniques used in the physical world, but to what might be even more intense versions of them. So intense, in fact, that the users' virtual experiences seamlessly meld into real-world experiences, memories, and understandings. This changes the way they see and understand the world and affects the way they behave in it.

As immersive technologies become more sophisticated, they will allow the collection of a greater variety of biometric data. Current VR headsets, for example, are mostly limited to tracking body movement. Technology for tracking eye movements, however, is currently being adapted from domains like the aerospace industry. As sensors become smaller and headsets and other wearable devices become lighter and more manageable, virtual reality systems will be able to gather multiple streams of biometric data. VR devices that combine these streams of data will be able to measure and predict user behaviors and attitudes. There will be many who say they will use this capability to enhance the user experience. Rest assured that there will be others who, for example, use the technology to effectively manipulate a mob to burn the home of a perceived Frankenstein-like monster when there was no more than an innocent scientist within.

Consider a few examples of the types of biometric data that an adversary could exploit. Galvanic skin response can discern how intensely a user is feeling an emotion. Facial muscles are known to reveal feelings like sadness, fear, surprise, anger, contempt, disgust and joy. Muscle tension can be measured using electromyography (EMG). One can use EMG to detect extremely rapid micro-expressions that would be almost impossible for someone to fake. The eye tracker can be used to see what the subject finds visually salient. Electroencephalography can measure how attentive or distracted a user is at any time. A system performing a combined analysis of these biometric data streams, and others will provide a comprehensive picture of the emotional state of a user in real-time. One can use that dynamic picture as a feedback loop to modify the stimulus to the user for maximum impact. Physiological characteristics related to walking and other movement are useful for gait tracking where one can identify a person by the particular way they walk. Other types of bodily gestures that a VRE could measure might be as revealing about the identity of a user as a fingerprint, retinal scan or a voice print.

The wide variety of biometric measurements will collectively enable the monitoring of how users act, interact and react at a level of detail previously not practical. Such monitoring will support the detailed profiling of their behaviors far beyond what has been possible with traditional social media. This, in turn, will support manipulation of users more effectively than ever before.

An important question is how users can be helped to make sure they do not misinterpret targeted, persuasive content as natural organic encounters. This problem has not even been solved in traditional social media environments where the levers of manipulation are not nearly as extensive as those that will be readily available in VREs. Consider, for example, a VRE where soldiers in an active combat area are able to gather, discuss events and generally commiserate with each other. The adversary could hack into such an environment and plant content, like two soldiers having a discussion that is designed to look authentic and whose purpose is to spread panic and demoralize the troops. This will be very difficult to detect with current technology, especially if their conversation makes use of a broad range of emotional communication channels like gestures, voice tone, pupil dilation, facial expressions, etc.

CONCLUSION

Cognitive Warfare is about degrading the adversary's cognitive capability. A key element of maintaining cognitive superiority in a Cognitive Warfare context is an effective cognitive defense. Emotions are integral to cognitive capability so a major attack vector against cognitive capability is emotional manipulation. Virtual Reality environments (VRE) will be a major battlefield of the future as current social media transitions into VREs. Because of their immersive characteristics, VREs offer opportunities for cognitive attacks based on emotional manipulation at a level far beyond what is possible in today's social media environments. Capabilities of VREs for emotional manipulation have already been demonstrated in the psychology and psychotherapy communities where current VREs are being used as part of a variety of psychotherapy treatments. In order to adequately defend our forces against such cognitive attacks and ensure cognitive superiority, we will have to learn to automatically recognize instances of these attacks. We should continuously develop techniques to inoculate our forces against them. We should also develop organizational structures and mechanisms to deliver awareness of and inoculation against cognitive attacks on an ongoing and continuous basis to personnel at all levels.

As a first step towards meeting the requirements of maintaining cognitive superiority in a VRE, we can develop a comprehensive taxonomy and catalog of emotional manipulation techniques in VREs. This may include how emotional manipulation techniques exist today and how they might evolve with VRE technology to serve as a foundation for the development of appropriate mitigation strategies to protect future forces. One approach to this would be to conduct a comprehensive survey and study of psychology and psychotherapy literature where there are reported findings of a wide variety of experiments and practical experiences using existing VREs for emotional manipulation. Compelling data and evidence will be in the form of reported proven VRE techniques for emotional manipulation that have been used as part of various psychotherapy treatments together with results of VRE experiments with emotional manipulation. The results of this analysis and study will be foundational to the development of mitigation and defensive strategies for maintaining cognitive superiority in Cognitive Warfare environments of the future. In parallel with these developments, we could begin to design the organizational and command structures that will be needed to implement the mitigation strategies for defense against cognitive attacks.

There are, without doubt, many characteristics of VREs that will greatly benefit a Military Metaverse. While we must prepare to take advantage of the benefits of this new technology, we must also prepare to face its potential dangers.

Dr. Rand Waltzman has created, planned and managed major research and development programs on the application of massive scale data analytics and machine learning in the areas of computer vision, insider threat detection and social media and strategic communications during two separate terms as a Program Manager at DARPA. He was most recently Senior Information Scientist and Deputy Chief Technology Officer at the RAND Corporation and is now retired - sort of.