Applying the Principles of Magic and the Concepts of Macrocognition to Counter-Deception in Cyber Operations

Simon HENDERSON[a], Robert HOFFMAN[b], Larry BUNCH[b], Jeff BRADSHAW[b]
a. Centre for Cyber Security & Information Systems, Cranfield University, Defence Academy of the UK
b. Institute for Human and Machine Cognition, Florida

ABSTRACT
Like magic tricks, most cyber attacks involve some form of deception. What are the key factors in cyber deception and how can we characterize and anticipate them? Concepts of macrocognition and the theory of magic are formative of a scheme to support cyberworkers as they try to make sense of complexity and dynamics, and act effectively in the face of uncertainty. This paper outlines a general theoretical foundation and multifactorial analytical scheme for the analysis of cyber attacks. In our primary case study we analyze the 4chan hack of the Time Magazine "Person of the Year" web poll. To demonstrate the extensibility of the scheme, we also deconstruct password cracking, footprinting, key logging and buffer flow cyber attacks.

BACKGROUND

Like magic tricks, cyber attacks involve deception, even when deception is not the sole purpose of the cyber attack. The primary purpose or intent of cyber attacks is of course to achieve some effect, and that intent can be enabled by deception. Most of the salient examples of cyber attacks involve deception that is brought about by various means. Deception is defined as a deliberate action to induce erroneous sensemaking and subsequent activity within a target audience to achieve and exploit an advantage (Henderson, 2011). The purpose of the deception is to bring about some influence on the defender, either to get the defender to do something or to keep the defender from doing something (e.g., noticing the attack).

Numerous treatises have been written on deception in military campaigns (see Cruickshank, 1979; Dewar, 1989; Gooch & Perlmutter, 1982; Holt, 2008; Howard, 1992; Latimer, 2001; Whaley, 2007). There are also numerous and respected discourses on the art and theory of deception that is employed in magic (Earl, 2012, 2013; Fitzkee, 1943, 1944, 1945; Higham, 2009, 2011; Jermay, 2003; Lamont and Wiseman, 1999; Ortiz 1994, 2006; Steinmeyer, 2004; Tamariz, 1987; Triplett, 1900). There are scientific studies of the rules of persuasion and influence (Cialdini, 2001), and there are useful analyses of con artistry (Cornelius, 2009; Lovell, 1996; Maurer, 2000; Robbins, 2008). The principles and analytical scheme presented here were originally conceived for the context of military deception planning (Henderson, 2007, 2011, 2012). Here, these ideas are synthesized into an operational procedure for defense against deception specifically in cyber defense.

A cyber attack can happen on a temporal scale that is so brief that it precludes human comprehension, analysis and intervention. Thus, one might assume that the only possible solution for cyberdefense operations is solely computational in nature. However, all forms of deception basically involve humans trying to mislead other humans. As the magician and theorist Daniel Fitzkee (1943) said,

> "Ultimately it is the spectator's mind which must be deceived, or there is no deception whatever. All of the apparatus we use, all of the secret gimmicks we employ, all of the sleights and stratagems we invoke—everything which identifies magic as mystery—the whole is designed to deceive the mind, and the mind alone, of the spectator."

Setting aside the ways in which cognitive work might play into cyber defense at the micro-temporal scale, cognitive work absolutely plays into cyber defense as macro scales where strategies and tactics are crucial.

DECEPTION CAN BE ANALYZED IN TERMS OF THE MAGICAL PLOYS

Discourses on magic have discussed dozens of magical "ploys." A ploy involves misleading the audience into thinking that their search for information is adequate and has been satisfied (Lamont and Wiseman, 1999). A clear example of a ploy is the "cover," when the magician waves one hand on a broad movement, thus distracting

attention from what his other hand is doing. Additional ploys are the "glance" (people will look where the magician is looking), the "missing page" (tell a story that has obvious gaps), the "surprise" (violation of expectations) the "exploit" (tapping biases expectations or prior beliefs), and the "reveal" (to make something obvious). Ploys can be understood in terms of which macrocognitive functions or processes they manipulate or disrupt (e.g., sensemaking, attention management, storytelling, etc.).

DECEPTION LEVERAGES MACROCOGNITION

Deception is achieved through the presentation of cue sequences that, via pattern recognition (framing), influence the process of mental modeling and thereby influence decision making. An attacker is able only to stimulate or direct a target's attention and problem detection. However, by manipulating attention in a structured way, the defender's mental modeling can be influenced. For example, repeated cue sequences can be used to condition a target's expectancies, pattern development, and direction of attention. Furthermore, if expectancies derive from the defender's own mental model—that is, the defender has no awareness that their mental model has itself been influenced—then the defender's expectancies will be vulnerable to deception. Another principle is that d*eception is more success*ful if it includes some form of emotional state induction, which can induce time pressure and interfere with reasoning.

These, and additional principles apply not just to individuals but to targets that operate on the basis of collective belief, that is, team and organizational sensemaking. Furthermore, the principles of deception necessarily invoke the fundamental macrocognitive processes and functions (see Klein, 2007; Klein, Moon and Hoffman, 2006; Klein, et al., 2003), especially sensemaking, mental modeling, and projection to the future. Cyber work brings additional macrocognitive functions into the mix, especially problem detection, flexecution, management of uncertainty, and management of attention.

From the cyber defender's perspective, the primary challenge for attention management is to answer the question *"Where do I look to find the data I need?"* From the cyber attacker's perspective, the defender's attempt to answer this question *c*an be influenced by directing the defender's attention, dividing the defender's attention, creating noise, exploiting the defender's inattention, concealing information, denying the defender the opportunity to find information, hiding the information, simulating the information in the "wrong" place, revealing false or bogus information, or substituting believable information for the genuine information. All of these are ploys utilized by magicians.

From the cyber defender's perspective, the primary challenge for sensemaking is to answer the question *"What counts as data?"* From the cyber attacker's perspective, the defender's sensemaking activity *c*an be influenced by making a moving target, revealing the data and in so doing lead the defender to believe that the data must be deceptive, inducing confirmation bias on the part of the defender (the defender seeks information that confirms their hypothesis), inducing disconfirmation bias (the defender does not seek information that would disconfirm an hypothesis), or swapping reality for an obvious and bogus deception (the "double-bluff").

Once the defender has an initial frame that determines what counts as data, the question for sensemaking is "*How do I understand these data?"* Will the initial frame be questioned and refined or questioned and rejected? From the cyber attacker's perspective, the defender's sensemaking activity *c*an be influenced by suggesting a pattern, supporting the verification of expectations, repeating pattern fragments to condition expectations, meeting the defender's expectations, dazzling (distracting) the defender, feeding the defender piecemeal information in order to stretch out the defender's sensemaking process, "accidentally" exposing the attacker's intent, such that the defender does not believe it, or fragmenting the pattern (make the defender invest effort to figure things out, thereby increasing the strength of their attachment to derived erroneous conclusions).

Once the defender has a frame that determines what counts as data, and the frame has been confirmed, or refined and improved, the question for is "*How do I act on these data?"* From the cyber attacker's perspective, the defender's flexecution activity *c*an be influenced by falsely confirming the attacker's intent and thereby causing the defender to engage in the wrong actions, falsely confirming that the defense has been effective, thereby causing the defender to cease an action, constraining the effectiveness of a defense, delaying the defender's actions, or channeling the defender's actions in certain directions or certain kinds of activity.

CASE STUDY

'Emily Williams' was a 2012 internet-based social engineering and technical attack conducted by two security researchers (Lakhani and Muniz, 2013) to gain access to a US Government VPN, take control of their email system, obtain access to confidential information, and obtain a physical laptop belonging to the organization. The attack was based on 'Robin Sage', another fictitious person created in 2009 as a demonstration in the ease of obtaining information from intelligence on US military personnel via social networks; the successful Robin Sage findings were presented at Black Hat 2010 (Ryan, 2010).

The researchers first created a false Facebook and LinkedIn profile for a character they named 'Emily Williams' ('Control Attention' via 'Planting', 'Show the False', 'Fragment Story Fragments'). An attractive waitress (exploiting Cialdini's notion of 'Liking') volunteered photos for the fictitious character. She actually worked at an establishment frequented by the target company's employees (the nearby Hooters) yet no employee recognised her in person at any time during the experiment.

Before targeting the government target's employees, Lakhani and Muniz built Williams's presence on social media, building hundreds of connections ('Show the false' via 'Inventing', 'Social Proof'), with only one man flagging her as suspicious. Another man asked how Emily might know him, and when the researchers answered with information they obtained from the man's social media profile ('Anticipate Suspicion Driven Searching', 'Show the False' via 'Mimicry), he said he did indeed remember the imaginary girl ('Memory is Attention in the Past').

Once Williams had friends, the researchers updated her Facebook and LinkedIn profiles with just-hired status at the government target ('Mimicry', 'Generate Expectations'), and gave her an engineering title ('Authority'). The attractive, imaginary young woman connected with the target's employees via social media and connected with Human Resources, IT Support, Engineering and those in executive leadership roles (further 'Social Proof'). The congratulations for "her" new job subsequently rolled in.

As it was near the holidays, no one questioned when Williams posted seasonal cards to Facebook directed at specific targets among her co-workers - which they clicked, executed a Browser Exploitation Framework (BeEF) signed Java applet that opened a reverse shell back to Lakhani and Muniz via an SSL connection ('Liking', 'Reciprocity', 'Hide the Real' via 'Repackaging', 'Control Expectations', 'Simulation the Action').

Key logging was then used to gather passwords and insider information to gain access to the target agency ('Hide the Real' via 'Repackaging'). The researchers were able to figure out domain credentials to create an inside email address for Williams ('Show the False' via 'Invention'), VPN passwords to gain internal access and other methods to compromise the target.

The use of an inside email account subsequently enabled further social engineering ('Show the False' via 'Invention', 'Exploit Prior Beliefs'). Men working for the government agency were targeted to provide Williams, special treatment based on her attractive photograph ('Liking'). Some men offered to help Miss Williams at her new job by doing her a few favours; namely circumventing usual channels to get her a work laptop, and access to the organisation's network ('Pique Curiosity/Lure', 'Emotional Appeal').

We selected Operation Emily because it demonstrates how a single cyber attack episode can involve multiple strategies and multiple forms of deception. Thus, it serves as a useful case for the analysis based on the Theory of Magic. One should not assume the stereotype of the cyber attack as a single entity launching clever software and then just sitting back to see what happens. Cyber attacks are much more like army-on-army conflicts or like races. An attack can involve multiple and independent or even competitive hacker entities.

TOOL FOR ANALYSIS IN CYBER DEFENSE

The above ideas can be composed as a tool for cyber defense operations. The basic scheme is a matrix with columns such as:

1. Attacker's goal/intent,
2. Attacker's ploy,
3. Attacker's actions to implement the ploy (what changes or moves?),

4. Defender's indicators of an attack,
5. Defender's counter indicators (that there is no attack),
6. Defender's actions to prevent or mitigate the attack,
7. Defender's indicators that prevention has been achieved,
8. Attacker's indicators of attack progress,
9. Attacker's indicators that a defense has been engaged,
10. Attacker's bogus indicators of the success of a defense, and
11. Defender's bogus indicators of a defense success.

The rows of the matrix would be specific cyber attacks. This is illustrated in Table 2, below. This Table is for illustrative purposes; it includes only four types of cyber attack and only the first four columns in the complete matrix.

Table 1. A scheme for counter-strategies in cyber defense.

1. DECEIVER'S GOAL/INTENT	2. DECEIVER'S PLOY	3. WHAT CHANGES OR MOVES?	4. DEFENDER'S INDICATORS
ATTACK: FOOTPRINTING			
To acquire network, service and layout information; to map potential targets in the network.	Conceal (passive scanning); Camouflage (active scanning); Exploit inattention.	Connections are made from the attacker site.	Noise is created (Defender notices connection failures and hits into a darknet); Changes in the number of servers that are providing a service; Changes in the types of servers that are providing a service; Change in communications protocol that a server is using; Suspicious behavior on the part of a client (repeated failures).
ATTACK: PASSWORD CRACKING			
Access privileged information.	Camouflage; Making a moving target.	Information is transferred off the target's network.	Periodic password login failures over a number of different users.
ATTACK: KEY LOGGING			
Obtain personal information.	Camouflage (a process name that target does not recognize).	Accounts, passwords, financial information from the target computer back to the attacker	Virtually none; Target must know the attack is happening and look for it .
ATTACK: BUFFER OVERFLOW			
Take control of target's computer.	Exploit inattention.	Information packets are sent from the attacker's computer to the target's host computer; Overwrite of existing target information.	Bad packet.

GENERALIZED MODEL

This analysis suggests a general event model, depicted in Figure 1. This diagram covers only selected aspects of what we call the "Three Cycles." In Cycle One, an attack is launched, it is detected or not, and it succeeds or not. In Cycle Two, there is defense activity and deception activity, either of which might or might not be successful. In Cycle Three there is counter-deception and the use of bogus ploys and bogus indicators. We refer to these as Cycles because they involve closed loops, that is, they all have feedback implications (e.g., a defensive operation might be observable by the attacker and hence "give things away"). Obviously, Cycle Three is where things get highly complex and confusing (see Hoffman, et al., 2011).

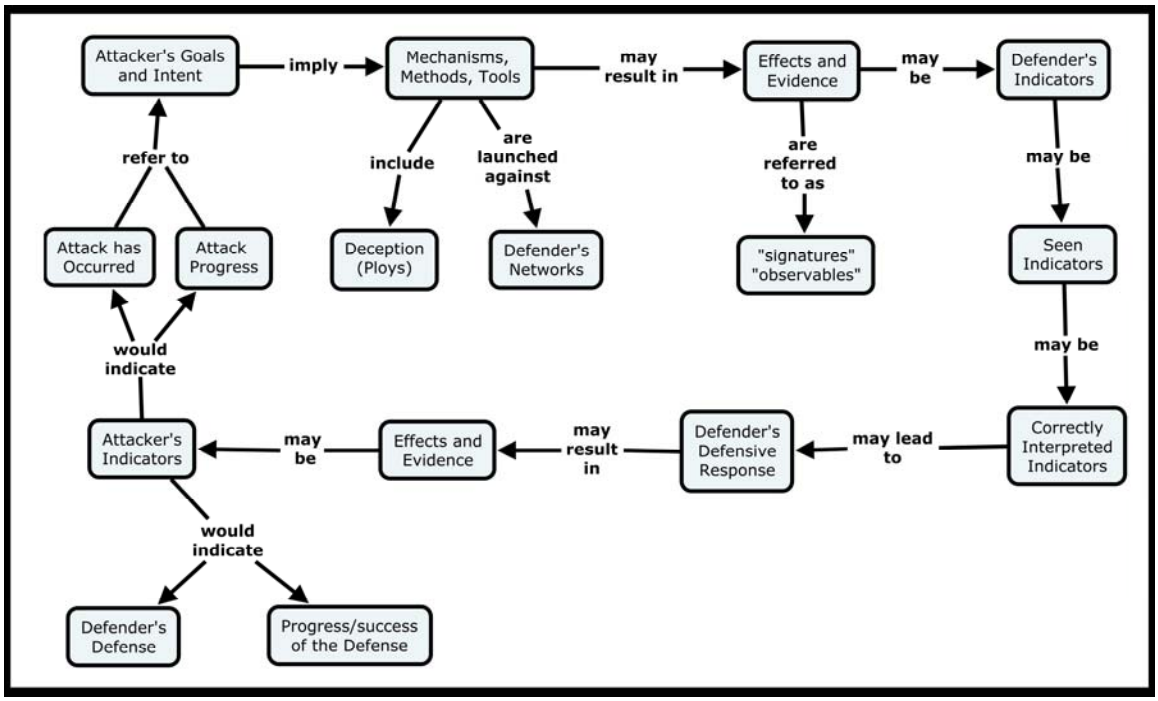

Figure 1. A process description of for cyber defense and offense based on the theory of magic and the concepts of macrocognition.

At a recent DoD meeting, a senior intelligence officer was asked:

> *Should intelligence operations always assume that the attacker is being deceptive, and that the attacker knows that the defender believes that the attacker is being deceptive, and that the defender will engage in counter-deception, and that the defender should engage in counter-deception? And can't this all not drive you nuts, but you have to think it through this deeply?*

The officer's answer was "Yes." Our current effort involves applying this analytical scheme to additional cyber attack and defense activities, and fleshing out the descriptive models of the "Three Cycles."

ACKOWLEDGEMENT

Preparation of a draft of this paper at IHMC was supported in part by a Subcontract to SoarTech, Inc., on an SBIR from the Office of Naval Research.

REFERENCES

Cialdini, R.B. (2001). The science of persuasion. *Scientific American, 284*, 76-81.
Cornelius, G. (2009). *The art of the con.* Alexandria, VA: American Correctional Association.
Cruickshank, C. G. (1979*). Deception in World War II*. New York: Oxford University Press.
Dewar, M. (1989). *The art of deception in warfare*. Newton Abbot: David & Charles.

Earl, B. (2012, 2013). *Less is more* (Vols 1, 2) Benjamin Earl
Fitzkee, D. (1943). *Showmanship for Magicians*. San Rafael, CA: San Rafael House
Fitzkee, D. (1944). *The trick brain.* San Rafael, CA: San Rafael House
Fitzkee, D. (1945). *Magic by misdirection.* San Rafael, CA: San Rafael House
Gooch, J., and Perlmutter, A. (Eds.). (1982). *Military deception and strategic surprise*. New York: Frank Cass and Company Limited.
Henderson, S. M. (2007). "Deception: A Guide to Exploiting the Psychological Basis of Deception in Military Planning." MIST/06/07/702/21/1.1. Farnborough: QinetiQ.
Henderson, S. M. (2011). Deceptive Thinking Workshop. Paper presented at the 1st MilDec Military Deception Symposium, 2nd-3rd November 2011, Defence Academy of the United Kingdom, Shrivenham.
Henderson, S. (2012). Military Deception: Learning From Other Domains. Presentation at the 2nd MILDEC Deception Symposium, Defence Academy of the United Kingdom, Shrivenham, 7-8 November 2012
Higham, J. (2009). *Secrets of improvisational magic*. London: Justin Higham.
Higham, J. (2011). *The Kosbe system: The mechanics of improvisation in card magic.* London: Justin Higham
Hoffman, R.R., Henderson, S., Moon, B., Moore, D.T., & Litman, J.A. (2011). Reasoning difficulty in analytical activity. *Theoretical Issues in Ergonomic Science, 12*, 225–240.
Holt, T. (2008). *The deceivers: Allied military deception in the Second World War* (Vols 1, 2). London: The Folio Society Ltd.
Howard, M. (1992). *Strategic deception in the Second World War*. London: Pimlico.
Jermay, L. (2003). *Building blocks.* Ashford: Alakazam Magic
Klein, G. (2007). Flexecution, Part 2: Understanding and supporting flexible execution: *IEEE Intelligent Systems, 22*, 108-112.
Klein, G., Moon, B. & Hoffman, R. R. (2006, November/December). *Making sense of sensemaking 2: A macrocognitive model.* IEEE Intelligent Systems, pp. 88-92.
Klein, G., Ross, K. G., Moon, B. M., Klein, D. E., Hoffman, R. R., & Hollnagel, E. (May/June, 2003). Macrocognition. *IEEE: Intelligent Systems*, pp. 81-85.
Lakhani, A., & Muniz, J. (2013). Social media deception. Paper presented at the RSAConference Europe 2013, October 29-31, 2013, Amsterdam.
Latimer, J. (2001). *Deception in war*. London: John Murray.
Lamont, P. & Wiseman, R. (1999). *Magic in Theory: An introduction to the theoretical and psychological elements of conjuring.* Hatfield, UK: University of Herfordshire Press.
Lovell, S. (1996). *How to cheat at everything: A con man reveals the secrets of the esoteric trade of cheating, scams, and hustles*. Philadelphia PA: Running Press Book Publishers
Maurer, D. W. (2000). *The big con: The story of the confidence man and the confidence trick.* London: Arrow Books
Ortiz, D. (1994). *Strong magic*. Washington, DC: Kaufman and Co.
Ortiz, D. (2006). *Designing miracles* (Vol 1). El Dorado Hills, CA: A1 MagicalMedia.
Robbins, T. (2008). The Modern Con Man: How to Get Something for Nothing. New York: Bloomsbury
Ryan, T. (2010). Getting in bed with Robin Sage. [Retrieved 12/01/2015, from http://www.omachonuogali.com/BlackHat-USA-2010-Ryan-Getting-In-Bed-With-Robin-Sage v1.0.pdf]
Steinmeyer, J. (2003). *Hiding the elephant: How magicians invented the impossible and learned to disappear*. New York: New York: Carroll & Graf Publishers.
Tamariz, J. (1987). *The magic way*. Madrid: Editorial Frakson.
Triplett, N. (1900). The psychology of conjuring deceptions. *The American Journal of Psychology, 11*, 439-510
Whaley, B. (2007). *Strategem: Deception and surprise in war*. Norwood, MA: Artech House.