Misperception, Self-Deception and Information Warfare

Lachlan Brumley, Carlo Kopp and Kevin Korb

Clayton School of Information Technology,
Monash University, Australia
E-mail: lbrumley@csse.monash.edu.au, carlo@csse.monash.edu.au, korb@csse.monash.edu.au

Abstract

Deception techniques are Information Warfare strategies commonly used by biological organisms and organisations to gain an advantage during competition. In this paper we examine two related techniques, misperception and self-deception, which we relate to the four canonical Information Warfare strategies and Boyd's OODA loop model.

Keywords

Information Warfare, Misperception, Self-Deception, Deception, OODA loop

INTRODUCTION

While the canonical Information Warfare (IW) strategies clearly explain the role of deception in Information Operations, the roles of misperception and self-deception are not so obvious. A common misperception about Information Warfare is that it is a new form of warfare, recently developed for the information age. This is not true, as Information Warfare techniques have been used by humans throughout their social and military history, and even longer by various biological entities in nature (Kopp and Mills, 2002).

Information Warfare

Information Warfare (IW) is defined by the United States Department of Defense as "any action to Deny, Exploit, Corrupt or Destroy the enemy's information and its functions; protecting ourselves against those actions and exploiting our own military information functions".

Previous works by both Borden (1999) and Kopp (2000) have categorised Information Warfare actions into four canonical strategies, which we will now briefly explain.

1. Degradation or Destruction (or Denial of Information)
   Aims to deny information to an opponent either by flooding the information channel with noise or by altering the object so that it more closely resembles the background noise of the information channel. This aims to reduce the amount of useful information that an opponent can obtain from the channel. Camouflage is one example of a denial strategy, as the user is hidden from the opponent and transmits no information about their presence, location or actions.

2. Corruption (or Deception and Mimicry)
   Aims to intentionally provide false information to an opponent. The user will attempt to mimic a signal that the opponent is familiar with, giving a false belief to the opponent.
   Examples are insects that disguise themselves as small twigs or leaves make it difficult for their predators to identify them and predators that disguise themselves as their prey to increase their chances of feeding. In the case of military actions, deception may consist of sending signals that convince an opponent of the false target of an invasion.

3. Denial [1] (or Disruption and Destruction)
   Aims to destroy or disrupt the opponent's information receiver, to inhibit their information gathering. An information receiver is any device that gathers information from the environment, such as eyes, ears, a

video camera or a radar receiver. Disrupting behaviour consists of any action that is intended to temporarily prevent the information receiver from gathering any information, for example radar jamming. Destructive behaviour consists of actions that destroy the information receiver, preventing its use altogether, such as capturing an opponent's spies or destroying their radar emplacements.

4. Denial [2] (or Subversion)

   Aims to modify the opponent to produce behaviour that is either self-destructive or to prevent behaviour that benefits the opponent. This is a much more complex behaviour that is not frequently used, probably due to its complexity. An example of Denial usage can be found in ant species, that enter the colony of others and alter the behaviour of the native ants to kill their own queen and serve the new queen.

5. Exploitation

   Aims to collect information from the opponent about their current state, location and actions. Exploitation is not actually an IW attack, as it does not affect the information channel or the receiver, however it is listed here for completeness.

OODA Loop

The Observation - Orientation - Decision - Action (OODA) loop model (Boyd 1986, Richards 2001) is one method of modelling an individual's event loop. This four step cycle models the information gathering, decision making and actions of an individual over time, with earlier behaviour providing feedback to current analysis and decision.

Military strategy is not the only realm where the concept of a event loop has been discovered and analysed. Neisser (1976) proposed a similar model for the psychology of perception, called the Perception Action Cycle. Norman (1990) has proposed a seven step action cycle for modelling human actions. The event loops of agents in Artificial Intelligence (Russell and Norvig 1995), Artificial Life and Robotics are also structured in a similar manner, with the agents proceeding through cycles of perceiving, deciding and acting. What is common to all of these behaviour models is a cycle of perception, decision, action and feedback, where current actions alter the state of the world and these changes provide feedback during future perception steps. We will focus on Boyd's model, as it intentionally separates the gathering of information and the assimilation of this information into the individual's perceived world model.

In the Observation phase information from the environment is acquired, which includes feedback from any previous actions taken. The new information is examined during the Orientation phase, where the individual uses it to update their model of the world. The individual's existing beliefs and knowledge are combined with the new information gathered, to produce a model of the world that reflects the individual's current perception of the state of the world. This model does not need to match reality, it is a product of the beliefs and processing abilities of the individual.

Once a model has been constructed, the individual moves to the Decision phase where they consider potential actions and decide on a course of action. Finally in the Action phase, the individual performs their decided action, which causes a change in the state of world. The individual will now proceed through the cycle again to see how the world has changed since they last observed it.

DECEPTION

Deception is another label given to the canonical Information Warfare strategy Corruption. It is the act of transmitting false information to an opponent, with the intention that this information will give the opponent a false belief. The false belief is intended to cause the victim to suffer from a specialised misperception that will benefit the deceiver in a specialised way. The victim's future decisions will be based upon false information, allowing a deceiver to use deception to alter their opponent's behaviour to an advantage. One example of the use of strategic deception in a military campaign is the Allied invasion of France in 1944 and Haswell (1979) provides a detailed analysis of the deception strategies employed and the advantages provided by the various deceptions.

In nature deception is typically used either by predators mimicking another species considered to be harmless by their prey, or by prey mimicking a species that their predators will not predate upon. Several examples of mimicry used by insects has been covered by Kopp and Mills (2002). Mimicry benefits predators by allowing them to get much closer to their prey without alarming it, while prey benefit from escaping the attention of their predators. Both of these abilities are clearly advantageous and have been selected for by evolution.

When an individual's behaviour cycle is considered as an OODA loop, deception strategies can be considered attempts to alter the beliefs of the opponent by supplying deceptive information for the victim to collect during Observation. This information is stored as a belief and is consulted during the Orientation step when the deception victim attempts to update their model of the world. At this point, if the deception strategy has succeeded, then the victim's perceived state of the world will include the belief that the deceiver provided. The victim now goes on to Decide and then Act based upon their corrupted model of the world, which will leave them open to exploitation by the deceiver.

MISPERCEPTION

Misperception occurs when information is incorrectly interpreted by an individual, leading them to create a mental model of the world that does not reflect reality. They may not correctly perceive the intentions or actions of other individuals or perhaps even the existence or non-existence of other objects in the world. The problems caused by misperception become apparent when the individual makes decisions based upon their incorrect information, potentially leading them to disaster.

When considered in the context of the OODA loop model, a misperceiver collects information normally during the Observation step. This information is then passed on to the Orientation step to be processed and combined with the individual's existing beliefs to create a model of the world. However the individual incorrectly interprets the new information and creates a false belief which is added to their model of the world. A false belief may arise either due to the actions of an opponent or due to flaws in the individual's information collection apparatus. Whether the misperception is intentional or not is dependent on the source of the false belief that caused it. If the information that led to the false belief was placed by a Corruption attack, then the misperception would be intentional. If the false belief was caused due to an error in the Observation step then it may or may not be intentional. The error is caused by a flaw somewhere in the individual's information collection apparatus. If the flaw was caused by an opponent using a Denial (Disruption and Destruction) attack against the misperceiver, then the misperception would be intentional. However if the flaw was not caused by the opponent, then the misperception would be unintentional. These flaws could be physical shortcomings, such as poor eyesight or hearing, or information processing faults, such as biases, assumptions or existing beliefs that affect the interpretation and analysis of information.

An intentional misperception is caused by either directly using Corruption to cause the victim to adopt a false belief, or by indirectly using a Denial attack against the victim's information collecting apparatus. The former strategy gives the opponent control over the false belief, while the latter does not.

Unintentional misperception may be caused either by faulty information gathering devices or existing biases. Faults in information gathering devices may have been caused by the use of Denial (Disruption and Destruction) strategies by opponents. With a limited flow of information, an individual is more likely to misperceive their surroundings. Should they be wrong, they will have added false beliefs, which may later lead to misperceptions. The degradation of the information gathering devices does not have to be caused by an IW attack. They may also simply degrade over time by themselves - machinery breaks down, while people's eyesight and hearing degrade.

Existing biases are beliefs and information that an individual already has incorporated into their model of the world. When new information is processed it is done so in the context of the individual's existing beliefs. If the existing beliefs are false, then they will affect the processing of new information and may lead to misperception. In this case the misperception is not due to the intentional actions of an opponent but to existing beliefs that are not correct.

SELF-DECEPTION

Self-deception is a special type of deception, where the deceiver and the victim are the same individual. While this may appear to be counterproductive at best, there are some explanations for how self-deception can be beneficial. Trivers (1976) proposes that self-deception can be beneficial if it is used to support lying. Trivers' hypothesis is that self-deception is used to convince a deceiver that they believe their own lie. When an individual communicates a lie to their intended victim, the victim will also receive a secondary message via the deceiver's body language that indicates the veracity of the message. If self-deception is used, the deceiver believes the lie and their body language will indicate that they are being truthful. However if self-deception is not used, then the deceiver's body language will indicate that they are lying and the victim may see through the deception. After the victim has been deceived, the deceiver restores the correct belief to their memory. The deceiver may then take advantage of the misperception that they gave to their opponent. This method uses self-deception as a support mechanism for deception.

It is argued by Ramachandran (1996) that self-deception cannot help deceivers in this way, as adopting the deceptive belief hides the goal of the deception from the deceiver. This leaves them unable to later benefit from the deception. Ramachandran's theory for the purpose of self-deception is that self-deception is used as a defense mechanism. The individual uses self-deception to create a coherent belief structure for themselves, which will impose stability on their behaviour. Individuals can therefore hide information from themselves that disagrees with their core beliefs. This theory also agrees with the belief held by psychologists and psychiatrists that self-deception is used to protect the user from harmful memories, by suppressing them. Organisations can also self-deceive, with one or more members hiding harmful information in order to maintain cohesion for the organisation.

Both of these theories require that the self-deceiver has some unconscious mechanism that controls the use of self-deception. It decides when self-deception will be used, implants the false belief and later restores the correct belief. We will refer to this mechanism as the self-deception controller, or simply controller. The controller is considered to be a black box within the self-deceiver, responsible for initiating and managing self-deceptive behaviour. This generalisation allows its use to be easily mapped to both individuals and organisations.

Self-deception aiding deception

Firstly we will consider the application of self-deception used to aid deception. When an individual lies, there will be other information that they are unconsciously transmitting. For a person, these could include their tone of voice and body language. All of these sources can indicate to the victim of the deception attempt that the deceiver may not be telling the truth. This is exactly what the deceiver does not want - it wastes any effort they have put into the deception and can cause the victim to distrust them.

We will consider the case of an organisation that wishes to convince others that they believe in X. However their lying can be detected by observers, who may notice that the organisation's behaviours do not match those of an organisation that believes in X. The observers can then conclude that the organisation is lying about their belief of X.

Self-deception allows the organisation to temporarily adopt the belief X (or at least the pretext that they do). This causes the organisation to integrate the false belief into their perception of the world during their Orientation step. To an observer the organisation now appears to state that they believe X and act in a manner that supports this statement. This combination of communication and behaviour is described by one of Haswell's (1985) principles of deception, multi-channel support. Multi-channel support is the use of multiple sources to transmit correlating false information to a victim during a deception attempt. The victim collects the verbal statement and the non-verbal behaviour that supports the veracity of the statement during its Observation step. During the Orientation step, both of these items will become integrated into the victim's perception of the world.

Ramachandran states that as long as the false belief is retained, the self-deceiver cannot benefit. Trivers' theory assumes that the self-deception controller thoroughly and permanently purges the false belief from the self-deceiver after the opponent has been deceived. This allows the self-deceptive organisation to purge their

belief of X and then take advantage of the deception that they used in some way.

In the context of Information Warfare, the lie is a Corruption strategy and self-deception is used to enhance the credibility of the lie. Self-deception and misperception both alter an individual's belief system, causing it to move further away from the actual true state of the world. This may be a risky thing to do, but the potential benefit of a more convincing deception is enough to tempt some.

There is also the potential for the self-deception to fail, which when the self-deception controller is unable to restore the correct belief. This in turn prevents their benefit from the deception that they have used and may even cause them to suffer from it. For example, if the false belief was intended to convince the victim to do something foolish, then it is possible that the self-deceiver will find themselves compelled to do this foolish thing. The self-deception is considered to fail as the self-deceiver cannot benefit and may end up worse off than if they had not used self-deception. This may be due to a fault in the self-deception controller or an error made by it, causing it to poorly manage or implement the use of self-deception.

Self-deception for suppressing information

Now we consider self-deception from Ramachandran's point of view, as a method of suppressing information that is harmful to the self. Why might information be harmful? It may reveal an organisation's weaknesses or it might conflict greatly with their perceived reality. In the latter case the self-deceiver is faced with the choice of simply hiding the offending information or incorporating it into their model, greatly damaging it. In the short term, the safer and easier path for them is the one of self-deception, however this may not be true in the long term. Andrews (2004) states that the use of self-deception in this way by various American intelligence agencies led to the destruction of the World Trade Center.

In this use of self-deception, the harmful information is collected and passed on to the self-deceiver's Orientation step. During the Orientation step, the individual fails to integrate the new information with their existing beliefs. This failure is due to the harmful nature of the information. The new information is discarded and a replacement belief that does not conflict with the existing beliefs is constructed. This maintains internal cohesion for the individual, at the expense of creating false beliefs.

Self-deception that is used in this way is a Degradation or Destruction strategy, with the self-deceiver hiding the harmful information from themselves - and potentially from others who may observe them. An example of this behaviour may be the classification of documents by a government of a nation that conflict with the image the government is attempting to project for the nation. The documents are classified and the false belief is propagated, thereby hiding the harmful information from the nation. In this example, the government plays the controller, while the nation assumes the role of the self-deceiver. Several instances of self-deception and perception management by nations and organisations are discussed by Van Evera (2002) and Kopp (2005).

Self-deception used in this way can fail if the self-deceiver is unable to retain the false belief, possibly because new information conflicts with the belief in such a way that it cannot be ignored. At some point the harmful true belief is reintroduced to the self-deceiver's beliefs and causes the self-deceiver to confront their problems. This belief reduces the cohesion in the self-deceiver and may force them to rethink their model of the world to accept the true belief.

CONCLUSIONS

Misperception and self-deception are related to deception and can both be explained in terms of the four canonical IW strategies. This relationship can also be identified by the similarities that all three share in the OODA loop. Misperception is the intended effect of a successful corruption attack, although intentional corruption actions are not the only source of misperception. Self-deception can be used either to aid a corruption attempt against an opponent or to hide information from oneself. Failure of a self-deception strategy can also be directly harmful, as the information that is discarded may not always be recovered, which makes self-deception a risky strategy to use by an attacker.

REFERENCES

Andrews, C. (2004) Belief Systems, Information Warfare and Counter Terrorism, Proceedings of the 5th Australian Information Warfare & Security Conference, Perth, WA, pp: 92-99.

Borden A. (1999) What is Information Warfare? *Aerospace Power Chronicles*, United States Air Force, Air University, Maxwell AFB, Contributor's Corner, URL: http://www.airpower.maxwell.af.mil/airchronicles/cc/borden.html [Online; accessed: 17-August-2005].

Boyd J.R. (1986) *Patterns of Conflict*, Briefing Notes, December, 1986 - unpublished, Defense and the National Interest, URL: http://www.d-n-i.net/boyd/pdf/poc.pdf [Online; accessed: 15-August-2005].

Haswell, J. (1979) *The Intelligence and Deception of the D-Day Landings*, Batsford, London.

Haswell, J. (1985) *The Tangled Web*, John Goodchild Publishers, Wendover.

Kopp, C. (2000) Information Warfare: A fundamental paradigm of infowar, Systems: Enterprise Computing Monthly pp. 46-55. URL: http://www.ausairpower.net/OSR-0200.html [Online; accessed: 17-August-2005].

Kopp, C. (2005) Classical Deception Techniques and Perception Management vs the Four strategies of Information Warfare. Draft paper.

Kopp C. and Mills B.I. (2002) Information Warfare and Evolution, *Proceedings of the 3rd Australian Information Warfare & Security Conference*, ECU, Perth. November, 2002. pp: 352-360.

Neisser, U. (1976) *Cognition and Reality*, W. H. Freeman, San Francisco.

Norman, D. A. (1990) *The Design of Everyday Things*, Doubleday, New York.

Ramachandran, V. S. (1996) The evolutionary biology of self-deception, laughter, dreaming and depression: Some clues from anosognosia, *Medical Hypotheses* **47**: 347-362.

Richards, C. W. (2001) Boyd's OODA loop, Slideshow. URL: http://www.d-n-i.net/fcs/ppt/boyds_ooda_loop.ppt [Online; accessed 15-August-2005].

Russell, S. J. and Norvig, P (1995) *Artificial Intelligence: A Modern Approach*, Prentice Hall, Englewood Cliffs, New Jersey.

Trivers, R. (1976) Preface in: *The Selfish Gene* (Dawkins, R., ed), Oxford University Press, pp. v-vii.

Van Evera, S. (2002) Why states believe foolish ideas: Non-self-evaluation by states and societies. URL: http://web.mit.edu/polisci/research/vanevera/why_states_believe_foolish_ideas.pdf [Online; accessed 15-August-2005].

COPYRIGHT