# A Framework for Deception

**by Fred Cohen, Dave Lambert, Charles Preston, Nina Berry, Corbin Stewart, and Eric Thomas ***
July 13, 2001

- Fred Cohen: Fred Cohen & Associates, University of New Haven, Sandia National Laboratories
- Dave Lambert: SPAWAR Systems Center
- Charles Preston: Information Integrity, University of New Haven
- Nina Berry: Sandia National Laboratories
- Corbin Stewart: Sandia National Laboratories
- Eric Thomas: Sandia National Laboratories

This research was sponsored by the United States Department of Defense under MIPR1CDOEJG102 2112040 162-3825 P633D06 255X 633006.247.01.DD.00 JGBZZ.1 JOAN 1JG8CA

## Executive Summary

This paper overviews issues in the use of deception for information protection. Its objective is to create a framework for deception and an understanding of what is necessary for turning that framework into a practical capability for carrying out defensive deceptions for information protection.

### Overview of results:

We have undertaken an extensive review of literature to understand previous efforts in this area and to compile a collection of information in areas that appear to be relevant to the subject at hand. It has become clear through this investigation that there is a great deal of additional detailed literature that should be reviewed in order to create a comprehensive collection. However, it appears that the necessary aspects of the subject have been covered and that additional collection will likely be comprised primarily of detailing in areas that are now known to be relevant.

We have developed a framework for creating and analyzing deceptions involving individual people, individual computers, one person acting with one computer, networks of people, networks of computers, and organizations consisting of people and their associated computers. This framework has been used to model select deceptions and, to a limited extent, to assist in the development of new deceptions. This framework is described in the body of this report with additional details provided in the appendixes.

Based on these results; (1) we are now able to understand and analyze deceptions with considerably more clarity than we could previously, (2) we have command of a far greater collection of techniques available for use in defensive deception than was previously available and than others have published in the field, and (3) we now have a far clearer understanding of how and when to apply which sorts of techniques than was previously available. It appears that with additional effort over time we will be able to continue to develop greater and more comprehensive understanding of the subject and extend our understanding, capabilities, and techniques.

### Further Work:

It appears that a substantial follow-on effort is required in order to systematize the creation of defensive information protection deceptions. Such an effort would most likely require:

- The creation of a comprehensive collection of material on key subject areas related to deception. This has been started in this paper but there is clearly a great deal of effort left to be done.
- The creation of a database supporting the creation of analysis of defensive deceptions and a supporting software capability to allow that database to be used by experts in their creation and operation of deceptions.
- A team of experts working to create and maintain a capability for supporting deceptions and sets of supporting personnel used as required for the implementation of specific deceptions.

We strongly believe that this effort should continue over an extended period of time and with adequate funding, and that such effort will allow us to create and maintain a substantial lead over the threat types currently under investigation. The net effect will be an ongoing and increasing capability for the successful deception of increasingly skilled and hostile threats.

## Introduction and Overview

According to the American Heritage Dictionary of the English Language (1981):

*"deception" is defined as "the act of deceit"*
*"deceit" is defined as "deception".*

Since long before 800 B.C. when Sun Tzu wrote "The Art of War" [28] deception has been key to success in warfare. Similarly, information protection as a field of study has been around for at least 4,000 years [41] and

has been used as a vital element in warfare. But despite the criticality of deception and information protection in warfare and the historical use of these techniques, in the transition toward an integrated digitized battlefield and the transition toward digitally controlled critical infrastructures, the use of deception in information protection has not been widely undertaken. Little study has apparently been undertaken to systematically explore the use of deception for protection of systems dependent on digital information. This paper, and the effort of which it is a part, seeks to change that situation.

In October of 1983, [25] in explaining INFOWAR, Robert E. Huber explains by first quoting from Sun Tzu:

> "**Deception: The Key** The act of deception is an art supported by technology. When successful, it can have devastating impact on its intended victim. In Fact:
>
> > *"All warfare is based on deception. Hence, when able to attack, we must seem unable; when using our forces, we must seem inactive; when we are near, we must make the enemy believe we are far away; when far away, we must make him believe we are near. Hold out baits to entice the enemy. Feign disorder, and crush him. If he is secure at all points, be prepared for him. If he is in superior strength, evade him. If your opponent is of choleric temper, seek to irritate him. Pretend to be weak, that he may grow arrogant. If he is taking his ease, give him no rest. If his forces are united, separate them. Attack him where he is unprepared, appear where you are not expected."* [28]
>
> The ability to sense, monitor, and control own-force signatures is at the heart of planning and executing operational deception...
>
> The practitioner of deception utilizes the victim's intelligence sources, surveillance sensors and targeting assets as a principal means for conveying or transmitting a deceptive signature of desired impression. It is widely accepted that all deception takes place in the mind of the perceiver. Therefore it is *not* the act itself but the acceptance that counts!"

It seems to us at this time that there are only two ways of defeating an enemy:

> (1) One way is to have overwhelming force of some sort (i.e., an actual asymmetry that is, in time, fatal to the enemy). For example, you might be faster, smarter, better prepared, better supplied, better informed, first to strike, better positioned, and so forth.
>
> (2) The other way is to manipulate the enemy into reduced effectiveness (i.e., induced mis-perceptions that cause the enemy to misuse their capabilities). For example, the belief that you are stronger, closer, slower, better armed, in a different location, and so forth.

Having both an actual asymmetric advantage and effective deception increases your advantage. Having neither is usually fatal. Having more of one may help balance against having less of the other. Most military organizations seek to gain both advantages, but this is rarely achieved for long, because of the competitive nature of warfare.

---

### Overview of This Paper

The purpose of this paper is to explore the nature of deception in the context of information technology defenses. While it can be reasonably asserted that all information systems are in many ways quite similar, there are differences between systems used in warfare and systems used in other applications, if only because the consequences of failure are extreme and the resources available to attackers are so high. For this reason, military situations tend to be the most complex and risky for information protection and thus lead to a context requiring extremes in protective measures. When combined with the rich history of deception in warfare, this context provides fertile ground for exploring the underlying issues.

We begin by exploring the history of deception and deception techniques. Next we explore the nature of deception and provide a set of dimensions of the deception problem that are common to deceptions of the targets of interest. We then explore a model for deception of humans, a model for deception of computers, and a set of models of deceptions of systems of people and computers. Finally, we consider how we might design and analyze deceptions, discuss the need for experiments in this arena, summarize, draw conclusions, and describe further work.

---

## A Short History of Deception

### Deception in Nature

While Sun Tzu is the first known publication depicting deception in warfare as an art, long before Sun Tzu there were tribal rituals of war that were intended in much the same way. The beating of chests [44] is a classic example that we still see today, although in a slightly different form. Many animals display their apparent fitness to others as part of the mating ritual of for territorial assertions. [35] Mitchell and Thompson [35] look at human and nonhuman deception and provide interesting perspectives from many astute authors on many aspects of this subject. We see much the same behavior in today's international politics. Who could forget Kruschev banging his shoe on the table at the UN and declaring "We will bury you!" Of course it's not only the losers that 'beat their chests', but it is a more stark example if presented that way. Every nation declares its greatness, both to its own people and to the world at large. We may call it pride, but at some point it becomes bragging, and in conflict situations, it becomes a display. Like the ancient tribesmen, the goal is, in some sense, to avoid a fight. The hope is that, by making the competitor think that it is not worth taking us on, we will not have to waste our energy or our blood in fighting when we could be spending it in other ways. Similar noise-making tactics also work to keep animals from approaching an encampment. The ultimate expression of this is in the area of

nuclear deterrence. [45]

Animals also have genetic characteristics that have been categorized as deceptions. For example, certain animals are able to change colors to match the background or, as in the case of certain types of octopi, the ability to mimic other creatures. These are commonly lumped together, but in fact they are very different. The moth that looks like a flower may be able to 'hide' from birds but this is not an intentional act of deception. Survival of the fittest simply resulted in the death of most of the moths that could be detected by birds. The ones that happened to carry a genetic trait that made them look like a particular flower happened to get eaten less frequently. This is not a deception, it is a trait that survives. The same is true of the Orca whale which has colors that act as a dazzlement to break up its shape.

On the other hand, anyone who has seen an octopus change coloring and shape to appear as if it were a rock when a natural enemy comes by and then change again to mimic a food source while lying in wait for a food source could not honestly claim that this was an unconscious effort. This form of concealment (in the case of looking like a rock or foodstuff) or simulation (in the case of looking like an inedible or hostile creature) is highly selective, driven by circumstance, and most certainly driven by a thinking mind of some sort. It is a deception that uses a genetically endowed physical capability in an intentional and creative manner. It is more similar to a person putting on a disguise than it is to a moth's appearance.

---

## Historical Military Deception

The history of deception is a rich one. In addition to the many books on military history that speak to it, it is a basic element of strategy and tactics that has been taught since the time of Sun Tzu. But in many ways, it is like the history of biology before genetics. It consists mainly of a collection of examples loosely categorized into things that appear similar at the surface. Hiding behind a tree is thought to be similar to hiding in a crowd of people, so both are called concealment. On the surface they appear to be the same, but if we look at the mechanisms underlying them, they are quite different.

> *"Historically, military deception has proven to be of considerable value in the attainment of national security objectives, and a fundamental consideration in the development and implementation of military strategy and tactics. Deception has been used to enhance, exaggerate, minimize, or distort capabilities and intentions; to mask deficiencies; and to otherwise cause desired appreciations where conventional military activities and security measures were unable to achieve the desired result. The development of a deception organization and the exploitation of deception opportunities are considered to be vital to national security. To develop deception capabilities, including procedures and techniques for deception staff components, it is essential that deception receive continuous command emphasis in military exercises, command post exercises, and in training operations."*
> --JCS Memorandum of Policy (MOP) 116 [10]

MOP 116 also points out that the most effective deceptions exploit beliefs of the target of the deception and, in particular, decision points in the enemy commander's operations plan. By altering the enemy commander's perception of the situation at key decision points, deception may turn entire campaigns.

There are many excellent collections of information on deceptions in war. One of the most comprehensive overviews comes from Whaley [11], which includes details of 67 military deception operations between 1914 and 1968. The appendix to Whaley is 628 pages long and the summary charts (in appendix B) are another 50 pages. Another 30 years have passed since this time, which means that it is likely that another 200 pages covering 20 or so deceptions should be added to update this study. Dunnigan and Nofi [8] review the history of deception in warfare with an eye toward categorizing its use. They identify the different modes of deception as concealment, camouflage, false and planted information, ruses, displays, demonstrations, feints, lies, and insight.

Dewar [16] reviews the history of deception in warfare and, in only 12 pages, gives one of the most cogent high-level descriptions of the basis, means, and methods of deception. In these 12 pages, he outlines (1) the weaknesses of the human mind (preconceptions, tendency to think we are right, coping with confusion by leaping to conclusions, information overload and resulting filtering, the tendency to notice exceptions and ignore commonplace things, and the tendency to be lulled by regularity), (2) the object of deception (getting the enemy to do or not do what you wish), (3) means of deception (affecting observables to a level of fidelity appropriate to the need, providing consistency, meeting enemy expectations, and not making it too easy), (4) principles of deception (careful centralized control and coordination, proper preparation and planning, plausibility, the use of multiple sources and modes, timing, and operations security), and (5) techniques of deception (encouraging belief in the most likely when a less likely is to be used, luring the enemy with an ideal opportunity , the repetitive process and its lulling effect, the double bluff which involves revealing the truth when it is expected to be a deception, the piece of bad luck which the enemy believes they are taking advantage of, the substitution of a real item for a detected deception item, and disguising as the enemy). He also (6) categorizes deceptions in terms of senses and (7) relates 'security' (in which you try to keep the enemy from finding anything out) to deception (in which you try to get the enemy to find out the thing you want them to find). Dewar includes pictures and examples in these 12 pages to boot.

In 1987, Knowledge Systems Corporation [26] created a useful set of diagrams for planning tactical deceptions. Among their results, they indicate that the assessment and planning process is manual, lacks automated applications programs, and lacks timely data required for combat support. This situation does not appear to have changed. They propose a planning process consisting of (1) reviewing force objectives, (2) evaluating your own and enemy capabilities and other situational factors, (3) developing a concept of operations and set of actions, (4) allocating resources, (5) coordinating and deconflicting the plan relative to other plans, (6) doing a risk and feasibility assessment, (7) reviewing adherence to force objectives, and (8) finalizing the plan. They detail steps to accomplish each of these tasks in useful process diagrams and provide forms for doing a more systematic analysis of deceptions than was previously available. Such a planning mechanism does not appear to exist today for deception in information operations.

These authors share one thing in common. They all carry out an exercise in building categories. Just as the long standing effort of biology to build up genus and species based on bodily traits (phenotypes), eventually fell to a mechanistic understanding of genetics as the underlying cause, the scientific study of deception will eventually yield a deeper understanding that will make the mechanisms clear and allow us to understand and create deceptions as an engineering discipline. That is not to say that we will necessarily achieve that goal in this short examination of the subject, but rather that in-depth study will ultimately yield such results.

There have been a few attempts in this direction. A RAND study included a 'straw man' graphic [17](H7076) that showed deception as being broken down into "Simulation" and "Dissimulation Camouflage".

> *"Whaley first distinguishes two categories of deception (which he defines as one's intentional distortion of another's perceived reality): 1) dissimulation (hiding the real) and 2) simulation (showing the false). Under dissimulation he includes: a) masking (hiding the real by making it invisible), b) repackaging (hiding the real by disguising), and c) dazzling (hiding the real by confusion). Under simulation he includes: a) mimicking (showing the false through imitation), b) inventing (showing the false by displaying a different reality), and c) decoying (showing the false by diverting attention). Since Whaley argues that "everything that exists can to some extent be both simulated and dissimulated, " whatever the actual empirical frequencies, at least in principle hoaxing should be possible for any substantive area."*[29]

The same slide reflects on Dewar's view [16] that security attempts to deny access and counterintelligence attempts while deception seeks to exploit intelligence. Unfortunately, the RAND depiction is not as cogent as Dewar in breaking down the 'subcategories' of simulation. The RAND slides do cover the notions of observables being "known and unknown", "controllable and uncontrollable", and "enemy observable and enemy non-observable". This characterization of part of the space is useful from a mechanistic viewpoint and a decision tree created from these parameters can be of some use. Interestingly, RAND also points out the relationship of selling, acting, magic, psychology, game theory, military operations, probability and statistics, logic, information and communications theories, and intelligence to deception. It indicates issues of observables, cultural bias, knowledge of enemy capabilities, analytical methods, and thought processes. It uses a reasonable model of human behavior, lists some well known deception techniques, and looks at some of the mathematics of perception management and reflexive control.

---

## Cognitive Deception Background

Many authors have examined facets of deception from both an experiencial and cognitive perspective.

Chuck Whitlock has built a large part of his career on identifying and demonstrating these sorts of deceptions. [12] His book includes detailed descriptions and examples of scores of common street deceptions. Fay Faron points out that most such confidence efforts are carried as as specific 'plays' and details the anatomy of a 'con' [30]. She provides 7 ingredients for a con (too good to be true, nothing to lose, out of their element, limited time offer, references, pack mentality, and no consequence to actions). The anatomy of the confidence game is said to involve (1) a motivation (e.g., greed), (2) the come-on (e.g., opportunity to get rich), (3) the shill (e.g., a supposedly independent third party), (4) the swap (e.g., take the victim's money while making them think they have it), (5) the stress (e.g., time pressure), and (6) the block (e.g., a reason the victim will not report the crime). She even includes a 10-step play that makes up the big con.

Bob Fellows [13] takes a detailed approach to how 'magic' and similar techniques exploit human fallibility and cognitive limits to deceive people. According to Bob Fellows [13] (p 14) the following characteristics improve the changes of being fooled: (1) under stress, (2) naivety, (3) in life transitions, (4) unfulfilled desire for spiritual meaning, (5) tend toward dependency, (6) attracted to trance-like states of mind, (7) unassertive, (8) unaware of how groups can manipulate people, (9) gullible, (10) have had a recent traumatic experience, (11) want simple answers to complex questions, (12) unaware of how the mind and body affect each other, (13) idealistic, (14) lack critical thinking skills, (15) disillusioned with the world or their culture, and (16) lack knowledge of deception methods. Fellows also identifies a set of methods used to manipulate people.

Thomas Gilovich [14] provides in-depth analysis of human reasoning fallibility by presenting evidence from psychological studies that demonstrate a number of human reasoning mechanisms resulting in erroneous conclusions. This includes the general notions that people (erroneously) (1) believe that effects should resemble their causes, (2) misperceive random events, (3) misinterpret incomplete or unrepresentative data, (4) form biased evaluations of ambiguous and inconsistent data, (5) have motivational determinants of belief, (6) bias second hand information, and (7) have exaggerated impressions of social support. Substantial further detailing shows specific common syndromes and circumstances associated with them.

Charles K. West [32] describes the steps in psychological and social distortion of information and provides detailed support for cognitive limits leading to deception. Distortion comes from the fact of an unlimited number of problems and events in reality, while human sensation can only sense certain types of events in limited ways: (1) A person can only perceive a limited number of those events at any moment, (2) A person's knowledge and emotions partially determine which of the events are noted and interpretations are made in terms of knowledge and emotion (3) Intentional bias occurs as a person consciously selects what will be communicated to others, and (4) the receiver of information provided by others will have the same set of interpretations and sensory limitations.

Al Seckel [15] provides about 100 excellent examples of various optical illusions, many of which work regardless of the knowledge of the observer, and some of which are defeated after the observer sees them only once. Donald D. Hoffman [36] expands this into a detailed examination of visual intelligence and how the brain processes visual information. It is particularly noteworthy that the visual cortex consumes a great deal of the total human brain space and that it has a great deal of effect on cognition. Some of the 'rules' that Hoffman describes with regard to how the visual cortex interprets information include: (1) Always interpret a straight line in an image as a straight line in 3D, (2) If the tips of two lines coincide in an image interpret them as coinciding

in 3D, (3) Always interpret co-linear lines in an image as co-linear in 3D, (4) Interpret elements near each other in an image as near each other in 3D, (5) Always interpret a curve that is smooth in an image as smooth in 3D, (6) Where possible, interpret a curve in an image as the rim of a surface in 3D, (7) Where possible, interpret a T-junction in an image as a point where the full rim conceals itself; the cap conceals the stem, (8) Interpret each convex point on a bound as a convex point on a rim, (9) Interpret each concave point on a bound as a concave point on a saddle point, (10) Construct surfaces in 3D that are as smooth as possible, (11) Construct subjective figures that occlude only if there are convex cusps, (12) If two visual structures have a non-accidental relation, group them and assign them to a common origin, (13) If three or more curves intersect at a common point in an image, interpret them as intersecting at a common point in space, (14) Divide shapes into parts along concave creases, (15) Divide shapes into parts at negative minima, along lines of curvature, of the principal curvatures, (16) Divide silhouettes into parts at concave cusps and negative minima of curvature, (17) The salience of a cusp boundary increases with increasing sharpness of the angle at the cusp, (18) The salience of a smooth boundary increases with the magnitude of (normalized) curvature at the boundary, (19) Choose figure and ground so that figure has the more salient part boundaries, (20) Choose figure and ground so that figure has the more salient parts, (21) Interpret gradual changes in hue, saturation, and brightness in an image as changes in illumination, (22) Interpret abrupt changes in hue, saturation, and brightness in an image as changes in surfaces, (23) Construct as few light sources as possible, (24) Put light sources overhead, (25) Filters don't invert lightness, (26) Filters decrease lightness differences, (27) Choose the fair pick that's most stable, (28) Interpret the highest luminance in the visual field as white, flourent, or self-luminous, (29) Create the simplest possible motions, (30) When making motion, construct as few objects as possible, and conserve them as much as possible, (31) Construct motion to be as uniform over space as possible, (32) Construct the smoothest velocity field, (33) If possible, and if other rules permit, interpret image motions as projections of rigid motions in three dimensions, (34) If possible, and if other rules permit, interpret image motions as projections of 3D motions that are rigid and planar, (35) Light sources move slowly.

It appears that the rules of visual intelligence are closely related to the results of other cognitive studies. It may not be a coincidence that the thought processes that occupy the same part of the brain as visual processing have similar susceptibilities to errors and that these follow the pattern of the assumption that small changes in observation point should not change the interpretation of the image. It is surprising when such a change reveals a different interpretation, and the brain appears to be designed to minimize such surprises while acting at great speed in its interpretation mechanisms. For example, rule 2 (If the tips of two lines coincide in an image interpret them as coinciding in 3D) is very nearly always true in the physical world because coincidence of line ends that are not in fact coincident in 3 dimensions requires that you be viewing the situation at precisely the right angle with respect to the two lines. Another way of putting this is that there is a single line in space that connects the two points so as to make them appear to be coincident if they are not in fact conincident. If the observer is not on that single line, the points will not appear coincident. Since people usually have two eyes and they cannot align on the same line in space with respect to anything they can observe, there is no real 3 dimensional situation in which this coincidence can actually occur, it can only be simulated by 3 dimensional objects that are far enough away to appear to be on the same line with respect to both eyes, and there are no commonly occuring natural phenomena that pose anything of immediate visual import or consequence at thast distance. Designing visual stimuli that violate these principles will confuse most human observers and effective visual simulations should take these rules into account.

Deutsch [47] provides a series of demonstrations of interpretation and misinterpretation of audio information. This includes: (1) the creation of words and phrases out of random sounds, (2) the susceptibility of interpretation to predisposition, (3) misinterpretation of sound based on relative pitch of pairs of tones, (4) misinterpretation of direction of sound source based on switching speakers, (5) creation of different words out of random sounds based on rapid changes in source direction, and (6) the change of word creation over time based on repeated identical audio stimulus.

First Karrass [33] then Cialdini [34] have provided excellent summaries of negotiation strategies and the use of influence to gain advantage. Both also explain how to defend against influence tactics. Karrass was one of the early experimenters in how people interact in negotiations and identified (1) credibility of the presenter, (2) message content and appeal, (3) situation setting and rewards, and (4) media choice for messages as critical components of persuasion. He also identifies goals, needs, and perceptions as three dimensions of persuasion and lists scores of tactics categorized into types including (1) timing, (2) inspection, (3) authority, (4) association, (5) amount, (6) brotherhood, and (7) detour. Karrass also provides a list of negotiating techniques including: (1) agendas, (2) questions, (3) statements, (4) concessions, (5) commitments, (6) moves, (7) threats, (8) promises, (9) recess, (10) delays, (11) deadlock, (12) focal points, (13) standards, (14) secrecy measures, (15) nonverbal communications, (16) media choices, (17) listening, (18) caucus, (19) formal and informal memorandum, (20) informal discussions, (21) trial balloons and leaks, (22) hostility releivers, (23) temporary intermediaries, (24) location of negotiation, and (25) technique of time.

Cialdini [34] provides a simple structure for influence and asserts that much of the effect of influence techniques is built-in and occurs below the conscious level for most people. His structure consists of reciprocation, contrast, authority, commitment and consistency, automaticity, social proof, liking, and scarcity. He cites a substantial series of psychological experiments that demonstrate quite clearly how people react to situations without a high level of reasoning and explains how this is both critical to being effective decision makers and results in exploitation through the use of compliance tactics. While Cialdini backs up this information with numerous studies, his work is largely based on and largely cites western culture. Some of these elements are apparently culturally driven and care must be taken to assure that they are used in context.

Robertson and Powers [31] have worked out a more detailed low-level theoretical model of cognition based on "Perceptual Control Theory" (PCT), but extensions to higher levels of cognition have been highly speculative to date. They define a set of levels of cognition in terms of their order in the control system, but beyond the lowest few levels they have inadequate basis for asserting that these are orders of complexity in the classic control theoretical sense. The levels they include are intensity, sensation, configuration, transition / motion, events, relationships, categories, sequences / routines, programs / branching pathways / logic, and system concept.

David Lambert [2] provides an extensive collection of examples of deceptions and deceptive techniques mapped into a cognitive model intended for modeling deception in military situations. These are categorized into cognitive levels in Lambert's cognitive model. The levels include sense, perceive feature, perceive form, associate, define problem / observe, define problem solving status (hypothesize), determine solution options, initiate actions / responses, direct, implement form, implement feature, and drive affectors. There are feedback and cross circuiting mechanisms to allow for reflexes, conditioned behavior, intuition, the driving of perception to higher and lower levels, and models of short and long term memory.

Charles Handy [37] discusses organizational structures and behaviors and the roles of power and influence within organizations. The National Research Council [38] discusses models of human and organizational behavior and how automation has been applied in this area. Handy models organizations in terms of their structure and the effects of power and influence. Influence mechanisms are described in terms of who can apply them in what circumstances. Power is derived from physicality, resources, position (which yields information, access, and right to organize), expertise, personal charisma, and emotion. These result in influence through overt (force, exchange, rules and procedures, and persuasion), covert (ecology and magnetism), and bridging (threat of force) influences. Depending on the organizational structure and the relative positions of the participants, different aspects of power come into play and different techniques can be applied. The NRC report includes scores of examples of modeling techniques and details of simulation implementations based on those models and their applicability to current and future needs. Greene [46] describes the 48 laws of power and, along the way, demonstrates 48 methods that exert compliance forces in an organization. These can be traced to cognitive influences and mapped out using models like Lambert's, Cialdini's, and the one we are considering for this effort.

Closely related to the subject of deception is the work done by the CIA on the MKULTRA project. [52] In June 1977, a set of MKULTRA documents were discovered, which had escaped destruction by the CIA. The Senate Select Committee on Intelligence held a hearing on August 3, 1977 to question CIA officials on the newly-discovered documents. The net effect of efforts to reveal information about this project was a set of released information on the use of sonic waves, electroshock, and other similar methods for altering peoples' perception. Included in this are such items as sound frequencies that make people fearful, sleepy, uncomfortable, and sexually aroused; results on hypnosis, truth drugs, psychic powers, and subliminal persuasion; LSD-related and other drug experiments on unwitting subjects; the CIA's "manual on trickery"; and so forth. One 1955 MKULTRA document gives an indication of the size and range of the effort; the memo refers to the study of an assortment of mind-altering substances which would: (1) "promote illogical thinking and impulsiveness to the point where the recipient would be discredited in public", (2) "increase the efficiency of mentation and perception", (3) "prevent or counteract the intoxicating effect of alcohol" (4) "promote the intoxicating effect of alcohol", (5) "produce the signs and symptoms of recognized diseases in a reversible way so that they may be used for malingering, etc." (6) "render the indication of hypnosis easier or otherwise enhance its usefulness" (7) "enhance the ability of individuals to withstand privation, torture and coercion during interrogation and so-called 'brainwashing', (8) "produce amnesia for events preceding and during their use", (9) "produce shock and confusion over extended periods of time and capable of surreptitious use", (10) "produce physical disablement such as paralysis of the legs, acute anemia, etc.", (11) "produce 'pure' euphoria with no subsequent let-down", (12) "alter personality structure in such a way that the tendency of the recipient to become dependent upon another person is enhanced", (13) "cause mental confusion of such a type that the individual under its influence will find it difficult to maintain a fabrication under questioning", (14) "lower the ambition and general working efficiency of men when administered in undetectable amounts", and (15) "promote weakness or distortion of the eyesight or hearing faculties, preferably without permanent effects".

A good summary of some of the pre-1990 results on psychological aspects of self-deception is provided in Heuer's CIA book on the psychology of intelligence analysis. [49] Heuer goes one step further in trying to start assessing ways to counter deception, and concludes that intelligence analysts can make improvements in their presentation and analysis process. Several other papers on deception detection have been written and substantially summarized in Vrij's book on the subject.[50]

---

## Computer Deception Background

In the early 1990s, the use of deception in defense of information systems came to the forefront with a paper about a deception 'Jail' created in 1991 by AT&T researchers in real-time to track an attacker and observe their actions. [39] An approach to using deceptions for defense by customizing every system to defeat automated attacks was published in 1992, [22] while in 1996, descriptions of Internet Lightning Rods were given [21] and an example of the use of perception management to counter perception management in the information infrastructure was given [23]. More thorough coverage of this history was covered in a 1999 paper on the subject. [6] Since that time, deception has increasingly been explored as a key technology area for innovation in information protection. Examples of deception-based information system defenses include concealed services, encryption, feeding false information, hard-to-guess passwords, isolated sub-file-system areas, low building profile, noise injection, path diversity, perception management, rerouting attacks, retaining confidentiality of security status information, spread spectrum, and traps. In addition, it appears that criminals seek certainty in their attacks on computer systems and increased uncertainty caused by deceptions may have a deterrent effect. [40]

The public release of DTK Deception ToolKit led to a series of follow-on studies, technologies, and increasing adoption of technical deceptions for defense of information systems. This includes the creation of a small but growing industry with several commercial deception products, the HoneyNet project, the RIDLR project at Naval Post Graduate School, NSA-sponsored studies at RAND, the D-Wall technology, [66] [7] and a number of studies and developments now underway.

- **Commercial Deception Products:** The dominant commercial deception products today are DTK and Recourse Technologies. While the market is very new it is developing at a substantial rate and new results

from deception projects are leading to an increased appreciation of the utility of deceptions for defense and a resulting increased market presence.

- **The HoneyNet Project:** The HoneyNet project is dedicated to learning and to the tools, tactics, and motives of the blackhat community and sharing the lessons learned. The primary tool used to gather this information is the Honeynet; a network of production systems designed to be compromised. This project has been joined by a substantial number of individual researchers and has had substantial success at providing information on widespread attacks, including the detection of large-scale denial of service worms prior to the use of the 'zombies' for attack. At least one Masters thesis is currently under way based on these results.

- **The RIDLR:** The RIDLR is a project launched from Naval Post Graduate School designed to test out the value of deception for detecting and defending against attacks on military information systems. RIDLR has been tested on several occasions at the Naval Post Graduate School and members of that team have participated in this project to some extent. There is an ongoing information exchange with that team as part of this project's effort.

- **RAND Studies:**

  In 1999, RAND completed an initial survey of deceptions in an attempt to understand the issues underlying deceptions for information protection. [18] This effort included a historical study of issues, limited tool development, and limited testing with reasonably skilled attackers. The objective was to scratch the surface of possibilities and assess the value of further explorations. It predominantly explored intelligence related efforts against systems and methods for concealment of content and creation of large volumes of false content. It sought to understand the space of friendly defensive deceptions and gain a handle on what was likely to be effective in the future.

  > This report indicates challenges for the defensive environment including: (1) adversary initiative, (2) response to demonstrated adversary capabilities or established friendly shortcomings, (3) many potential attackers and points of attack. (4) many motives and objectives, (5) anonymity of threats, (6) large amount of data that might be relevant to defense, (7) large noise content, (8) many possible targets, (9) availability requirements, and (10) legal constraints.
  >
  > Deception may: (1) condition the target to friendly behavior, (2) divert target attention from friendly assets, (3) draw target attention to a time or place, (4) hide presence or activity from a target, (5) advertise strength or weakness as their opposites, (6) confuse or overload adversary intelligence capabilities, or (7) disguise forces.
  >
  > The animal kingdom is studied briefly and characterized as ranging from concealment to simulation, at levels (1) static, (2) dynamic, (3) adaptive, and (4) premeditated.
  >
  > Political science and psychological deceptions are fused into maxims; (1) pre-existing notions given excessive weight, (2) desensitization degrades vigilance, (3) generalizations or exceptions based on limited data, (4) failure to fully examine the situation limits comprehension, (5) limited time and processing power limit comprehension, (6) failure to adequately corroborate, (7) over-valuing data based on rarity, (8) experience with source may color data inappropriately, (9) focusing on a single explanation when others are available, (10) failure to consider alternative courses of action, (11) failure to adequately evaluate options, (12) failure to reconsider previously discarded possibilities, (13) ambivalence by the victim to the deception, and (14) confounding effect of inconsistent data. This is very similar to the coverage of Gilovich [14] reviewed in detail elsewhere in this report.
  >
  > Confidence artists use a 3-step screening process; (1) low-investment deception to gage target reaction, (2) low-risk deception to determine target pliability, and (3) reveal a deception and gage reaction to determine willingness to break the rules.
  >
  > Military deception is characterized through Joint Pub 3-58 (Joint Doctrine for Military Deception) and Field Manual 90-02 [10] which are already covered in this overview.

  The report then goes on to review things that can be manipulated, actors, targets, contexts, and some of the then-current efforts to manipulate observables which they characterize as: (1) honeypots, (2) fishbowls, and (3) canaries. They characterize a space of (1) raw materials, (2) deception means, and (3) level of sophistication. They look at possible mission objectives of (1) shielding assets from attackers, (2) luring attention away from strategic assets, (3) the induction of noise or uncertainty, and (4) profiling identity, capabilities, and intent by creation of opportunity and observation of action. They hypothesize a deception toolkit (sic) consisting of user inputs to a rule-based system that automatically deploys deception capabilities into fielded units as needed and detail some potential rules for the operation of such a system in terms of deception means, material requirements, and sophistication. Consistency is identified as a problem, the potential for self-deception is high in such systems, and the problem of achieving adequate fidelity is reflected as it has been elsewhere.

  The follow-up RAND study [24] extends the previous results with a set of experiments in the effectiveness of deception against sample forces. They characterize deception as an element of "active network defense". Not surprisingly, they conclude that more elaborate deceptions are more effective, but they also find a high degree of effectiveness for select superficial deceptions against select superficial intelligence probes. They conclude, among other things, that deception can be effective in protection, counterintelligence, against cyber-reconnaissance, and to help to gather data about enemy reconnaissance. This is consistent with previous results that were more speculative. Counter deception issues are also discussed, including (1) structural, (2) strategic, (3) cognitive, (4) deceptive, and (5) overwhelming approaches.

- **Theoretical Work:** One historical and three current theoretical efforts have been undertaken in this area, and all are currently quite limited. Cohen looked at a mathematical structure of simple defensive network deceptions in 1999 [7] and concluded that as a counterintelligence tool, network-based deceptions could be of significant value, particularly if the quality of the deceptions could be made good enough. Cohen suggested the use of rerouting methods combined with live systems of the sorts being modeled as yielding the highest fidelity in a deception. He also expressed the limits of fidelity associated with system content, traffic patterns, and user behavior, all of which could be simulated with increasing accuracy for increasing cost. In this paper, networks of up to 64,000 IP addresses were emulated for high quality deceptions using a technology called D-WALL. [66]

  Dorothy Denning of Georgetown University is undertaking a small study of issues in deception. Matt Bishop of the University of California at Davis is undertaking a study funded by the Department of Energy on the mathematics of deception. Glen Sharlun of the Naval Post Graduate School is finishing a Master's thesis on the effect of deception as a deterrent and as a detection method in large-scale distributed denial of service attacks.

- **Custom Deceptions:** Custom deceptions have existed for a long time, but only recently have they gotten adequate attention to move toward high fidelity and large scales.

The reader is asked to review the previous citation [6] for more thorough coverage of computer-based defensive deceptions and to get a more complete understanding of the application of deceptions in this arena over the last 50 years.

Another major area of information protection through deception is in the area of steganography. The term steganography comes from the Greek 'steganos' (covered or secret) and 'graphy' (writing or drawing) and thus means, literally, covered writing. As commonly used today, steganography is closer to the art of information hiding, and is ancient form of deception used by everyone from ruling politicians to slaves. It has existed in one form or another for at least 2000 years, and probably a lot longer.

With the increasing use of information technology and increasing fears that information will be exposed to those it is not intended for, steganography has undergone a sort of emergence. Computer programs that automate the processes associated with digital steganography have become widespread in recent years. Steganographic content is now commonly hidden in graphic files, sound files, text files, covert channels, network packets, slack space, spread spectrum signals, and video conferencing systems. Thus steganography has become a major method for concealment in information technology and has broad applications for defense.

---

## The Nature of Deception

Even the definition of deception is illusive. As we saw from the circular dictionary definition presented earlier, there is no end to the discussion of what is and is not deception. This not withstanding, there is an end to this paper, so we will not be making as precise a definition as we might like to. Rather, we will simply assert that:

> Deception is a set of acts that seek to increase the chances that a set of targets will behave in a desired fashion when they would be less likely to behave in that fashion if they knew of those acts.

We will generally limit our study of deceptions to targets consisting of people, animals, computers, and systems comprised of these things and their environments. While it could be argued that all deceptions of interest to warfare focus on gaining compliance of people, we have not adopted this position. Similarly, from a pragmatic viewpoint, we see no current need to try to deceive some other sort of being.

While our study will seek general understanding, our ultimate focus is on deception for information protection and is further focused on information technology and systems that depend on it. At the same time, in order for these deceptions to be effective, we have to, at least potentially, be successful at deception against computers used in attack, people who operate and program those computers, and ultimately, organizations that task those people and computers. Therefore, we must understand deception that targets people and organizations, not just computers.

### Limited Resources lead to Controlled Focus of Attention

There appear to be some features of deception that apply to all of the targets of interest. While the detailed mechanisms underlying these features may differ, commonalities are worthy of note. Perhaps the core issue that underlies the potential for success of deception as a whole is that all targets not only have limited overall resources, but they have limited abilities to process the available sensory data they are able to receive. This leads to the notion that, in addition to controlling the set of information available to the targets, deceptions may seek to control the focus of attention of the target.

In this sense, deceptions are designed to emphasize one thing over another. In particular, they are designed to emphasize the things you want the targets to observe over the things you do not want them to observe. While many who have studied deception in the military context have emphasized the desire for total control over enemy observables, this tends to be highly resource consumptive and very difficult to do. Indeed, there is not a single case in our review of military history where such a feat has been accomplished and we doubt whether such a feat will ever be accomplished.

> *Example: Perhaps the best example of having control over observables was in the Battle of Britain in World War II when the British turned all of the Nazi intelligence operatives in Britain into double agents and combined their reports with false fires to try to get the German Air Force to miss their factories. But even this incredible level of success in deception did not prevent the Germans from creating technologies such as radio beam guidance systems that resulted in accurate targeting for periods of time.*

It is generally more desirable from an assurance standpoint to gain control over more target observables, assuming you have the resources to affect this control in a properly coordinated manner, but the reason for this may be a bit surprising. The only reason to control more observables is to increase the likelihood of attention being focused on observables you control. If you could completely control focus of attention, you would only need to control a very small number of observables to have complete effect. In addition, the cost of controlling observables tends to increase non-linearly with increased fidelity. As we try to reach perfection, the costs presumably become infinite. Therefore, there should be some cost benefit analysis undertaken in deception planning and some metrics are required in order to support such analysis.

## All Deception is a Composition of Concealments and Simulations

Reflections of world events appear to the target as observables. In order to affect a target, we can only create causes in the world that affect those observables. Thus all deceptions stem from the ability to influence target observables. At some level, all we can do is create world events whose reflection appear to the target as observables or prevent the reflections of world events from being observed by the target. As terminology, we will call induced reflections **'simulations'** and inhibition of reflections **'concealments'**. In general then, all deceptions are formed from combinations of concealments and simulations.

Put another way, deception consists of determining what we wish the target to observe and not observe and creating simulations to induce desired observations while using concealments to inhibit undesired observations. Using the notion of focus of attention, we can create simulations and concealments by inducing focus on desired observables while drawing focus away from undesired observables. Simulation and concealment are used to affect this focus and the focus then produces more effective simulation and concealment.

## Memory and Cognitive Structure Force Uncertainty, Predictability, and Novelty

All targets have limited memory state and are, in some ways, inflexible in their cognitive structure. While space limits memory capabilities of targets, in order to be able to make rapid and effective decisions, targets necessarily trade away some degree of flexibility. As a result, targets have some predictability. The problem at hand is figuring out how to reliably make target behavior (focus of attention, decision processes, and ultimately actions) comply with our desires. To a large extent, the purpose of this study is to find ways to increase the certainty of target compliance by creating improved deceptions.

There are some severe limits to our ability to observe target memory state and cognitive structure. Target memory state and detailed cognitive structure is almost never fully available to us. Even if it were available, we would be unable, at least at the present, to adequately process it to make detailed predictions of behavior because of the complexity of such computations and our own limits of memory and cognitive structure. This means that we are forced to make imperfect models and that we will have uncertain results for the foreseeable future.

While modeling of enough of the cognitive structures and memory state of targets to create effective deceptions may often be feasible, the more common methods used to create deceptions are the use of characteristics that have been determined through psychological studies of human behavior, animal behavior, analytical and experimental work done with computers, and psychological studies done on groups. The studies of groups containing humans and computers are very limited at and those that do exist ignore the emerging complex global network environment. Significant additional effort will be required in order to understand common modes of deception that function in the combined human-computer social environment.

A side effect of memory is the ability of targets to learn from previous deceptions. Effective deceptions must be novel or varied over time in cases where target memory affects the viability of the deception.

## Time, Timing, and Sequence are Critical

Several issues related to time come up in deceptions. In the simplest cases, a deception might come to mind just before it is to be performed, but for any complex deception, pre-planning is required, and that pre-planning takes time. In cases where special equipment or other capabilities must be researched and developed, the entire deception process can take months to years.

In order for deception to be effective in many real-time situations, it must be very rapidly deployed. In some cases, this may mean that it can be activated almost instantaneously. In other cases this may mean a time frame of seconds to days or even weeks or months. In strategic deceptions such as those in the Cold War, this may take place over periods of years.

In every case, there is some delay between the invocation of a deception and its effect on the target. At a minimum, we may have to contend with speed of light effects, but in most cases, cognition takes from milliseconds to seconds. In cases with higher momentum, such as organizations or large systems, it may take minutes to hours before deceptions begin to take effect. Some deceptive information is even planted in the hopes that it will be discovered and acted on in months to years.

Eventually, deceptions may be discovered. In most cases a critical item to success in the deception is that the time before discovery be long enough for some other desirable thing to take place. For one-shot deceptions intended to gain momentary compliance, discovery after a few seconds may be adequate, but other deceptions require longer periods over which they must be sustained. Sustaining a deception is generally related to preventing its discovery in that, once discovered, sustainment often has very different requirements.

Finally, nontrivial deceptions involve complex sequences of acts, often involving branches based on feedback attained from the target. In almost all cases, out of the infinite set of possible situations that may arise, some set of critical criteria are developed for the deception and used to control sequencing. This is necessary because

of the limits of the ability of deception planning to create sequencers for handling more complex decision processes, because of limits on available observables for feedback, and because of limited resources available for deception.

> *Example: In a commonly used magician's trick, the subject is given a secret that the magician cannot possibly know based on the circumstances. At some time in the process, the subject is told to reveal the secret to the whole audience. After the subject makes the secret known, the magician reveals that same secret from a hiding place. The trick comes from the sequence of events. As soon as the answer is revealed, the magician chooses where the revealed secret is hidden. What really happens is that the magician chooses the place based on what the secret is and reveals one of the many pre-planted secrets. If the sequence required the magician to reveal their hidden result first, this deception would not work.[13]*

### Observables Limit Deception

In order for a target to be deceived, their observations must be affected. Therefore, we are limited in our ability to deceive based on what they are able to observe. Targets may also have allies with different observables and, in order to be effective, our deceptions must take those observables into account. We are limited both by what can be observed and what cannot be observed. What cannot be observed we cannot use to induce simulation, while what can be observed creates limits on our ability to do concealment.

> *Example: Dogs are commonly used in patrol units because of the fact that they have different sensory and cognitive capabilities than people have. Thus when people try to conceal themselves from other people, the things they choose to do tend to fool other people but not animals like dogs which, for example, might smell them out even without seeing or hearing them.*

Our own observables also limit our ability to do deceptions because sequencing of deceptions depends on feedback from the target and because our observables in terms of accurate intelligence information drive our ability to understand the observables of the target and the effect of those observables on the target.

### Operational Security is a Requirement

Secrecy of some sort is fundamental to all deception, if only because the target would be less likely to behave in the desired fashion if they knew of the deception (by our definition above). This implies operational security of some sort.

One of the big questions to be addressed in some deceptions is who should be informed of the specific deceptions under way. Telling too many people increases the likelihood of the deception being leaked to the target. Telling too few people may cause the deception to fool your own side into blunders.

> *Example: In Operation Overlord during World War II, some of the allied deceptions were kept so secret that they fooled allied commanders into making mistakes. These sorts of errors can lead to fratricide.[16]*

Security is expensive and creates great difficulties, particularly in technology implementations. For example, if we create a device that is only effective if its existence is kept secret, we will not be able to apply it very widely, so the number of people that will be able to apply it will be very limited. If we create a device that has a set of operational modes that must be kept secret, the job is a bit easier. As we move toward a device that only needs to have it's current placement and current operating mode kept secret, we reach a situation where widespread distribution and effective use is feasible.

A vital issue in deception is the understanding of what must be kept secret and what may be revealed. If too much is revealed, the deception will not be as effective as it otherwise may have been. If too little is revealed, the deception will be less effective in the larger sense because fewer people will be able to apply it. History shows that device designs and implementations eventually leak out. That is why soundness for a cryptographic system is usually based on the assumption that only the keys are kept secret. The same principle would be well considered for use in many deception technologies.

A further consideration is the deterrent effect of widely published use of deception. The fact that high quality deceptions are in widespread use potentially deters attackers or alters their behavior because they believe that they are unable to differentiate deceptions from non-deceptions or because they believe that this differentiation substantially increases their workload. This was one of the notions behind Deception ToolKit (DTK). [19] The suggestion was even made that if enough people use the DTK deception port, the use of the deception port alone might deter attacks.

### Cybernetics and System Resource Limitations

In the systems theory of Norbert Weiner (called Cybernetics) [42] many systems are described in terms of feedback. Feedback and control theory address the notions of systems with expectations and error signals. Our targets tend to take the difference between expected inputs and actual inputs and adjust outputs in an attempt to restore stability. This feedback mechanism both enables and limits deception.

Expectations play a key role in the susceptibility of the target to deception. If the deception presents observables that are very far outside of the normal range of expectations, it is likely to be hard for the target to ignore it. If the deception matches a known pattern, the target is likely to follow the expectations of that pattern unless there is a reason not to. If the goal is to draw attention to the deception, creating more difference is more likely to achieve this, but it will also make the target more likely to examine it more deeply and with more skepticism. If the object is to avoid something being noticed, creating less apparent deviation from expectation is more likely to achieve this.

Targets tend to have different sensitivities to different sorts and magnitudes of variations from expectations.

These result from a range of factors including, but not limited to, sensor limitations, focus of attention, cognitive structure, experience, training, reasoning ability, and pre-disposition. Many of these can be measured or influenced in order to trigger or avoid different levels of assessment by the target.

Most systems do not do deep logical thinking about all situations as they arise. Rather, they match known patterns as quickly as possible and only apply the precious deep processing resources to cases where pattern matching fails to reconcile the difference between expectation and interpretation. As a result, it is often easy to deceive a system by avoiding its logical reasoning in favor of pattern matching. Increased rush, stress, uncertainty, indifference, distraction, and fatigue all lead to less thoughtful and more automatic responses in humans. [34] Similarly, we can increase human reasoning by reduced rush, stress, certainty, caring, attention, and alertness.

> *Example: Someone who looks like a valet parking person and is standing outside of a pizza place will often get car keys from wealthy customers. If the customers really used reason, they would probably question the notion of a valet parking person at a pizza place, but their mind is on food and conversation and perhaps they just miss it. This particular experiment was one of many done with great success by Whitlock. [12]*

Similar mechanisms exist in computers where, for example, we can suppress high level cognitive functions by causing driver-level response to incoming information or force high level attention and thus overwhelm reasoning by inducing conditions that lead to increased processing regimens.

### The Recursive Nature of Deception

The interaction we have with targets in a deception is recursive in nature. To get a sense of this, consider that while we present observables to a target, the target is presenting observables to us. We can only judge the effect of our deception based on the observables we are presented with and our prior expectations influence how we interpret these observables. The target may also be trying to deceive us, in which case, they are presenting us with the observables they think we expect to see, but at the same time, we may be deceiving them by presenting the observables we expect them to expect us to present. This goes back and forth potentially without end. It is covered by the well known story:

> *The Russian and US ambassadors met at a dinner party and began discussing in their normal manner. When the subject came to the recent listening device, the Russian explains that they knew about it for some time. The American explains that they knew the Russians knew for quite a while. The Russian explains they they knew the Americans knew they knew. The American explains that they knew the Russians knew that the Americans knew they knew. The Russian states that they knew they knew they knew they knew they knew they knew. The American exclaims "I didn't know that!".*

To handle recursion, it is generally accepted that you must first characterize what happens at a single level, including the links to recursion, but without delving into the next level those links lead to. Once your model of one level is completed, you then apply recursion without altering the single level model. We anticipate that by following this methodology we will gain efficiency and avoid mistakes in understanding deceptions. At some level, for any real system, the recursion must end for there is ground truth. The question of where it ends deals with issues of confidence in measured observables and we will largely ignore this issues throughout the remainder of this paper.

### Large Systems are Affected by Small Changes

In many cases, a large system can be greatly affected by small changes. In the case of deception, it is normally easier to make small changes without the deception being discovered than to directly make the large changes that are desired. The indirect approach then tells us that we should try to make changes that cause the right effects and go about it in an unexpected and indirect manner.

As an example of this, in a complex system with many people, not all participants have to be affected in order to cause the system to behave differently than it might otherwise. One method for influencing an organizational decision is to categorize the members into four categories: zealots in favor, zealots opposed, neutral parties, and willing participants. The object of this influence tactic in this case is to get the right set of people into the right categories.

> *Example: Creating a small number of opposing zealots will stop an idea in an organization that fears controversy. Once the set of desired changes is understood, moves can be generated with the objective of causing these changes. For example, to get an opposing zealot to reduce their opposition, you might engage them in a different effort that consumes so much of their time that they can no longer fight as hard against the specific item you wish to get moved ahead.*

This notion of finding the right small changes and backtracking to methods to influence them seems to be a general principle of organizational deception, but there has only been limited work on characterizing these effects at the organizational level.

### Even Simple Deceptions are Often Quite Complex

In real attacks, things are not so simple as to involve only a single deception element against a nearly stateless system. Even relatively simple deceptions may work because of complex processes in the targets.

> *As a simple example, we analyzed a specific instance of audio surveillance, which is itself a subclass of attack mechanism called audio/video viewing. In this case, we are assuming that the attacker is exploiting a little known feature of cellular telephones that allows them to turn on and listen to conversations without alerting the targets. This is a deception because the attacker is attempting to conceal the listening activity so that the target will talk when they otherwise might not, and it is a form of concealment because it is intended to avoid detection by the*

*target. From the standpoint of the telephone, this is a deception in the form of simulation because it involves creating inputs that cause the telephone to act in a way it would not otherwise act (presuming that it could somehow understand the difference between owner intent and attacker intent - which it likely can not). Unfortunately, this has a side effect.*

*When the telephone is listening to a conversation and broadcasting it to the attacker it consumes battery power at a higher rate than when it is not broadcasting and it emits radio waves that it would otherwise not emit. The first objective of the attacker would be to have these go unnoticed by the target. This could be enhanced by selective use of the feature so as to limit the likelihood of detection, again a form of concealment.*

*But suppose the target notices these side effects. In other words, the inputs do get through to the target. For example, suppose the target notices that their new batteries don't last the advertised 8 hours, but rather last only a few hours, particularly on days when there are a lot of meetings. This might lead them to various thought processes. One very good possibility is that they decide the problem is a bad battery. In this case, the target's association function is being misdirected by their predisposition to believe that batteries go bad and a lack of understanding of the potential for abuse involved in cell phones and similar technologies. The attacker might enhance this by some form of additional information if the target started becoming suspicious, and the act of listening might provide additional information to help accomplish this goal. This would then be an act of simulation directed against the decision process of the target.*

*Even if the target becomes suspicious, they may not have the skills or knowledge required to be certain that they are being attacked in this way. If they come to the conclusion that they simply don't know how to figure it out, the deception is affecting their actions by not raising it to a level of priority that would force further investigation. This is a form of concealment causing them not to act.*

*Finally, even if they should figure out what is taking place, there is deception in the form of concealment in that the attacker may be hard to locate because they are hiding behind the technology of cellular communication.*

*But the story doesn't really end there. We can also look at the use of deception by the target as a method of defense. A wily cellular telephone user might intentionally assume they are being listened to some of the time and use deceptions to test out this proposition. The same response might be generated in cases where an initial detection has taken place. Before association to a bad battery is made, the target might decide to take some measurements of radio emissions. This would typically be done by a combination of concealment of the fact that the emissions were being measured and the inducement of listening by the creation of a deceptive circumstance (i.e., simulation) that is likely to cause listening to be used. The concealment in this case is used so that the target (who used to be the attacker) will not stop listening in, while the simulation is used to cause the target to act.*

The complete analysis of this exchange is left as an exercise to the reader.. good luck. To quote the immortal Bard:

***"Oh what a tangled web we weave when first we practice to deceive"***

## Simple Deceptions are Combined to Form Complex Deceptions

Large deceptions are commonly built up from smaller ones. For example, the commonly used 'big con' plan [30] goes something like this: find a victim, gain the victim's confidence, show the victim the money, tell the tale, deliver a sample return on investment, calculate the benefits, send the victim for more money, take them for all they have, kiss off the victim, keep the victim quiet. Of these, only the first does not require deceptions. What is particularly interesting about this very common deception sequence is that it is so complex and yet works so reliably. Those who have perfected its use have ways out at every stage to limit damage if needed and they have a wide number of variations for keeping the target (called victim here) engaged in the activity.

## Knowledge of the Target

The intelligence requirements for deception are particularly complex to understand because, presumably, the target has the potential for using deception to fool the attacker's intelligence efforts. In addition, seemingly minor items may have a large impact on our ability to understand and predict the behavior of a target. As was pointed out earlier, intelligence is key to success in deception. But doing a successful deception requires more than just intelligence on the target. To get to high levels of surety against capable targets, it is also important to anticipate and constrain their behavioral patterns.

In the case of computer hardware and software, in theory, we can predict precise behavior by having detailed design knowledge. Complexity may be driven up by the use of large and complicated mechanisms (e.g., try to figure out why and when Microsoft Windows will next crash) and it may be very hard to get details of specific mechanisms (e.g., what specific virus will show up next). While generic deceptions (e.g., false targets for viruses) may be effective at detecting a large class of attacks, there is always an attack that will, either by design or by accident, go unnoticed (e.g., not infect the false targets). The goal of deceptions in the presence of imperfect knowledge (i.e., all real-world deceptions) is to increase the odds. The question of what techniques increase or decrease odds in any particular situation drives us toward deceptions that tend to drive up the computational complexity of differentiation between deception and non-deception for large classes of situations. This is intended to exploit the limits of available computational power by the target. The same notions can be applied to human deception. We never have perfect knowledge of a human target, but in various aspects, we can count on certain limitations. For example, overloading a human target with information will tend to make concealment more effective.

*Example: One of the most effective uses of target knowledge in a large-scale deception was the deception attack against Hitler that supported the D-day invasions of World War II. Hitler was specifically targeted in such a manner that he would personally prevent the German military from responding to the Normandy invasion. He was induced not to act when he otherwise would have by a combination of deceptions that convinced him that the invasion would*

*be at Pas de Calais. They were so effective that they continued to work for as much as a week after troops were inland from Normandy. Hitler thought that Normandy was a feint to cover the real invasion and insisted on not moving troops to stop it.*

*The knowledge involved in this grand deception came largely from the abilities to read German encrypted Enigma communications and psychologically profile Hitler. The ability to read ciphers was, of course, facilitated by other deceptions such as over attribution of defensive success to radar. Code breaking had to be kept secret to in order to prevent the changing of code mechanisms, and in order for this to be effective, radar was used as the excuse for being able to anticipate and defend against German attacks. [41]*

**Knowledge for Concealment**

The specific knowledge required for effective concealment is details of detection and action thresholds for different parts of systems. For example, knowing the voltage used for changing a 0 to a 1 in a digital system leads to knowing how much additional signal can be added to a wire while still not being detected. Knowing the electromagnetic profile of target sensors leads to better understanding of the requirements for effective concealment from those sensors. Knowing how the target's doctrine dictates responses to the appearance of information on a command and control system leads to understanding how much of a profile can be presented before the next level of command will be notified. Concealment at any given level is attained by remaining below these thresholds.

**Knowledge for Simulation**

The specific knowledge required for effective simulation is a combination of thresholds of detection, capacity for response, and predictability of response. Clearly, simulation will not work if it is not detected and therefore detection thresholds must be surpassed. Response capacity and response predictability are typically for more complex issues.

Response capacity has to do with quantity of available resources and ability to use them effectively. For computers, we know pretty well the limits of computational and storage capacity as well as what sorts of computations can be done in how much time. While clever programmers do produce astonishing results, for those with adequate understanding of the nature of computation, these results lead clearly toward the nature of the breakthrough. We constantly face deceptions, perhaps self-deceptions, in the proposals we see for artificial intelligence in computer systems and can counter it based on the understanding of resource consumption issues. Similarly, humans have limited capacity for handling situations and we can predict these limits at some level generically and in specific through experiments on individuals. Practice may allow us to build certain capacities to an artificially high level. The use of automation to augment capacities is one of the hallmarks of human society today, but even with augmentation, there are always limits.

Response predictability may be greatly facilitated by the notions of cybernetic stability. As long as we don't exceed the capacity of the system to handle change, systems designed for stability will have predictable tendencies toward returning to equilibrium. One of the great advantages of term limits on politicians, particularly at the highest levels, is that each new leader has to be recalibrated by those wishing to target them. It tends to be easier to use simulation against targets that have been in place for a long time because their stability criteria can be better measured and tested through experiment.

## Legality

There are legal limitations on the use of deception for those who are engaged in legal activities, while those who are engaged in illegal activities, risk jail or, in some cases, death for their deceptions.

In the civilian environment, deceptions are acceptable as a general rule unless they involve a fraud, reckless endangerment, or libel of some sort. For example, you can legally lie to your wife (although I would advise against it), but if you use deception to get someone to give you money, in most cases it's called fraud and carries a possible prison sentence. You can legally create deceptions to defeat attacks against computer systems, but there are limits to what you can do without creating potential civil liability. For example, if you hide a virus in software and it is stolen and damages the person who stole it or an innocent bystander, you may be subject to civil suit. If someone is injured as a side effect, reckless endangerment may be involved.

Police and other governmental bodies have different restrictions. For example, police may be subject to administrative constraints on the use of deceptions, and in some cases, there may be a case for entrapment if deceptions are used to create crimes that otherwise would not have existed. For agencies like the CIA and NSA, deceptions may be legally limited to affect those outside the United States, while for other agencies, restrictions may require activities only within the United States. Similar legal restrictions exist in most nations for different actions by different agencies of their respective governments. International law is less clear on how governments may or may not deceive each other, but in general, governmental deception is allowed and is widely used.

Military environments also have legal restrictions, largely as a result of international treaties. In addition, there are codes of conduct for most militaries and these include requirements for certain limitations on deceptive behavior. For example, it is against the Geneva convention to use Red Cross or other similar markings in deceptions, to use the uniform of the enemy in combat (although use in select other circumstances may be acceptable), to falsely indicate a surrender as a feint, and to falsely claim there is an armistice in order to draw the enemy out. In general, there is the notion of good faith and certain situations where you are morally obligated to speak the truth. Deceptions are forbidden if they contravene any generally accepted rule or involve treachery or perfidy. It is especially forbidden to make improper use of a flag of truce, the national flag, the military insignia and uniform of the enemy, or the distinctive badges of the Geneva convention. [10] Those violating these conventions risk punishment ranging up to summary execution in the field.

Legalities are somewhat complex in all cases and legal council and review should be considered before any questionable action.

## Modeling Problems

From the field of game theory, many notions about strategic and tactical exchanges have been created. Unfortunately, game theory is not as helpful in these matters as it might be both because it requires that a model be made in order to perform analysis and because, for models as complex as the ones we are already using in deception analysis, the complexity of the resulting decision trees often become so large as to defy computational solution. Fortunately, there is at least one other way to try to meet this challenge. This solution lies in the area of "model-based situation anticipation and constraint". [5] In this case, we use large numbers of simulations to sparsely cover a very large space.

In each of these cases, the process of analysis begins with models. Better models generally result in better results but sensitivity analysis has shown that we do not need extremely accurate models to get usable statistical results and meaningful tactical insight. [5] This sort of modeling of deception and the scientific investigation that supports accurate modeling in this area has not yet begun in earnest, but it seems certain that it must.

One of the keys to understanding deception in a context is that the deceptions are oriented toward the overall systems that are our targets. In order for us to carry out meaningful analysis, we must have meaningful models. If we do not have these models, then we will likely create a set of deceptions that succeed against the wrong targets and fail against the desired targets, and in particular, we will most likely be deceiving ourselves.

The main problem we must first address is what to model. In our case, the interest lies in building more effective deceptions to protect systems against attacks.

These targets of such defensive deceptions vary widely and they may ultimately have to be modeled in detail independently of each other, but there are some common themes. In particular, we believe we will need to build cognitive models of computer systems, humans, and their interactions as components of target systems. Limited models of attack strengths and types associated with these types of targets exist [5] in a form amenable to simulation and analysis. These have not been integrated into a deception framework and development has not been taken to the level of specific target sets based on reasonable intelligence estimates.

There have been some attempts to model deceptions before invoking them in the past. One series of examples is the series of deceptions starting with the Deception ToolKit, [6] leading to the D-Wall, [7] and then to the other projects. In these cases, increasingly detailed models of targets of defensive deceptions were made and increasingly complex and effective deceptions were achieved.

## Unintended Consequences

Deceptions may have many consequences, and these may not all be intended when the deceptions are used. Planning to avoid unintended consequences and limit the effects of the deceptions to just the target raises complex issues.

> *Example: When deception was first implemented to limit the effectiveness of computer network scanning technology, one side effect was to deceive the tools used by the defenders to detect their own vulnerabilities. In order for the deceptions to work against attackers, they also had to work against the defenders who were using the same technology.*

In the case of these deception technologies, this is an intended consequence that causes defenders to become confused about their vulnerabilities. This then has to be mitigated by adjusting the results of the scanning mechanism based on knowledge of what is a known defensive deception. In general, these issues can be quite complex.

In this case, the particular problem is that the deception affected observables of cognitive systems other than the intended target. In addition the responses of the target may indirectly affect others. For example, if we force a target to spend their money on one thing, the finiteness of the resource means that they will not spend that money on something else. That something else, in a military situation, might include feeding their prisoners, who also happen to be our troops.

All deceptions have the potential for unintended consequences. From the deceiver's perspective this is then an operations security issue. If you don't tell your forces about a deception you risk it being treated as real, while telling your own forces risks revealing the deception, either through malice or the natural difference between their response to the normal situation and the known deception.

Another problem is the potential for misassociation and misattribution. For example, if you are trying to train a target to respond to a certain action on your part with a certain action or inaction on their part, the method being used for the training may be misassociated by the target so that the indicators they use are not the ones you thought they would use. In addition, as the target learns from experiencing deceptions, they may develop other behaviors that are against your desires.

## Counterdeception

Many studies appear in the psychological literature on counterdeception [50] but little work has been done on the cognitive issues surrounding computer-based deception of people and targeting computers for deception. No metrics relating to effectiveness of deception were shown in any study of computer-related deception we were able to find. The one exception is in the provisioning of computers for increased integrity, which is generally discussed in terms of (1) honesty and truthfulness, (2) freedom from unauthorized modification, and

(3) correspondence to reality. Of these, only freedom from unauthorized modification has been extensively studied for computer systems. There are studies that have shown that people tend to believe what computers indicate to them, but few of these are helpful in this context.

Pamela Kalbfleisch categorized counterdeception in face-to-face interviews according to the following schema. [53] (1) No nonsense, (2) Criticism, (3) Indifference, (4) Hammering, (5) Unkept secret, (6) Fait accompli, (7) Wages alone, (8) All alone, (9) Discomfort and relief, (10) Evidence bluff, (11) Imminent discovery, (12) Mum's the word (13) Encouragement, (14) Elaboration, (15) Diffusion of responsibility, (16) Just having fun, (17) Praise (18) Excuses, (19) It's not so bad, (20) Others have done worse, (21) Blaming (22) Buildup of lies, (23) No explanations allowed, (24) Repetition, (25) Compare and contrast, (26) Provocation, (27) Question inconsistencies as they appear, (28) Exaggeration, (29) Embedded discovery, (30) A chink in the defense, (31) Self-disclosure, (32) Point of deception cues, (33) You are important to me, (34) Empathy, (35) What will people think?, (36) Appeal to pride, (37) Direct approach, and (38) Silence. It is also noteworthy that most of these counterdeception techniques themselves depend on deception and stem, perhaps indirectly, from the negotiation tactics of Karrass. [33]

Extensive studies of the effectiveness of counter deception techniques have indicated that success rates with face-to-face techniques rarely exceed 60% accuracy and are only slightly better at identifying lies than truths. Even poorer performance result from attempts to counter deception by examining body language and facial expressions. As increasing levels of control are exerted over the subject, increasing care is taken in devising questions toward a specific goal, and increasing motivation for the subject to lie are used, the rate of deception detection can be increased with verbal techniques such as increased response time, decreased response time, too consistent or pat answers, lack of description, too ordered a presentation, and other similar indicators. The aide of a polygraph device can increase accuracy to about 80% detection of lies and more than 90% detection of truths for very well structured and specific sorts of questioning processes. [50]

The limits of the target in terms of detecting deception leads to limits on the need for high fidelity in deceptions. The lack of scientific studies of this issue inhibit current capabilities to make sound decisions without experimentation.

### Summary

The following table summarizes the dimensions and issues involved:

| | |
|---|---|
| Limited Resources lead to Controlled Focus of Attention | By pressuring or taking advantage of pre-existing circumstances focus of attention can be stressed. In addition, focus can be inhibited, enhanced, and through the combination of these, redirected. |
| All Deception is a Composition of Concealments and Simulations | Concealments inhibit observation while simulations enhance observation. When used in combination they provide the means for redirection. |
| Memory and Cognitive Structure Force Uncertainty, Predictability, and Novelty | The limits of cognition force the use of rules of thumb as shortcuts to avoid the paralysis of analysis. This provides the means for inducing desired behavior through the discovery and exploitation of these rules of thumb in a manner that restricts or avoids higher level cognition. |
| Time, timing, and sequence are critical | All deceptions have limits in planning time, time to perform, time till effect, time till discovery, sustainability, and sequences of acts. |
| Observables Limit Deception | Target, target allies, and deceiver observables limit deception and deception control. |
| Operational Security is a Requirement | Determining what needs to be kept secret involves a trade off that requires metrics in order to properly address. |
| Cybernetics and System Resource Limitations | Natural tendencies to retain stability lead to potentially exploitable movement or retention of stability states. |
| The Recursive Nature of Deception | Recursion between parties leads to uncertainty that cannot be perfectly resolved but that can be approached with an appropriate basis for association to ground truth. |
| Large Systems are Affected by Small Changes | For organizations and other complex systems, finding the key components to move and finding ways to move them forms a tactic for the selective use of deception to great effect. |
| Even Simple Deceptions are Often Quite Complex | The complexity of what underlies a deception makes detailed analysis quite a substantial task. |
| Simple Deceptions are Combined to Form Complex Deceptions | Big deceptions are formed from small sub-deceptions and yet they can be surprisingly effective. |
| Knowledge of the Target | Knowledge of the target is one of the key elements in effective deception. |
| Legality | There are legal restrictions on some sorts of deceptions and these must be considered in any implementation. |
| Modeling Problems | There are many problems associated with forging and using good models of deception. |
| Unintended Consequences | You may fool your own forces, create mis-associations, and create mis-attributions. Collateral deception has often been observed. |

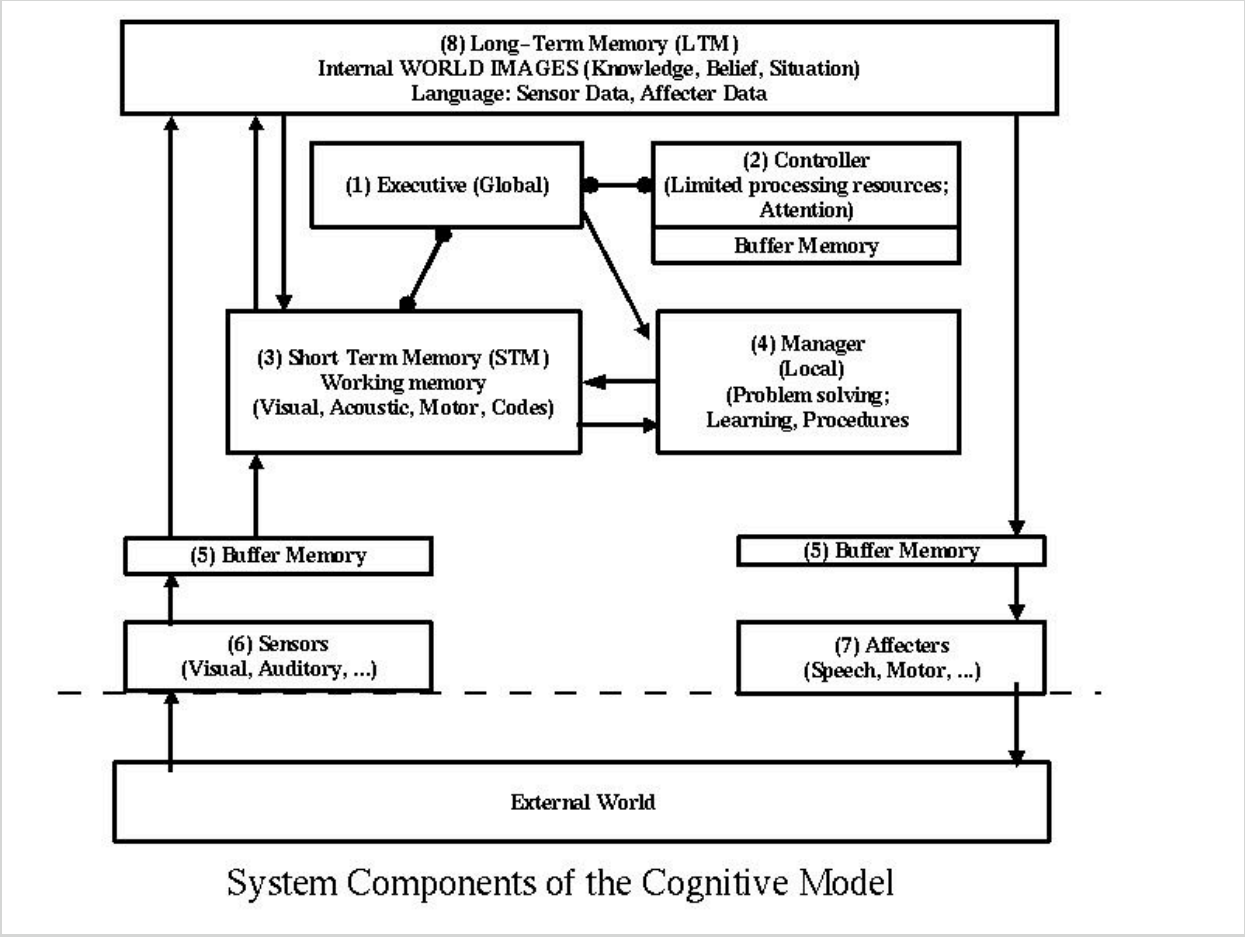
| | |
|---|---|
| Counterdeception | Target capabilities for counterdeception may result in deceptions being detected. |

## A Model for Human Deception

By looking extensively at the literature on human cognition and deception, a model was formed of human cognition with specific focus on its application to deception. This includes Lambert's data collection and mapping into his model of human deception.

### Lambert's Cognitive Model

We begin with Lambert's model of human cognition. [2] This model is linked to the history of psychological models of brain function and cognition and, as such, does not represent so much the physiology of the brain as the things it is generally believed to do and the manner in which it is generally believed to operate. There is no sense that this model will be found to match physiology in the long run, however, it is useful because it relates to a great deal of other experimental work that has been done on deception and the limits of human perception. It may also be related to perceptual control theory's notions of orders of control and, through that mechanistic view, to physiology. [31]



System Components of the Cognitive Model

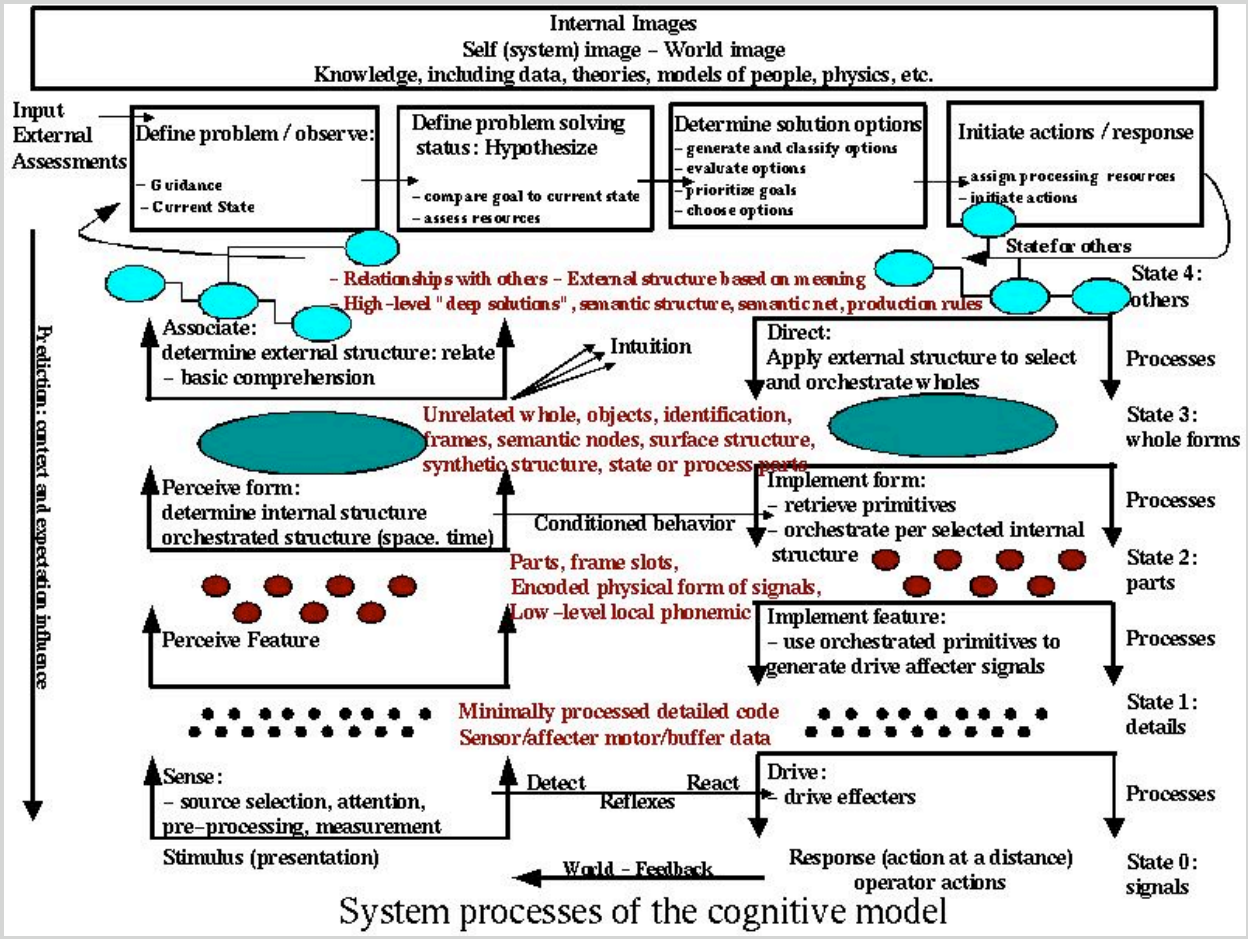
### Lambert's Model of Cognition

This model identifies integers as labels for major brain functions. Within this model, Lambert has created a structure of sub processes identified with behavior in general and deception in particular. This structure is broken down into subsections as follows. In addition to the structural association, Lambert created a detailed mapping of how cognitive function was thought to work. The structure can be interpreted as a stimulus response network but there is an isomorphism to a model-referenced adaptive control system. The components consist of (1) the global executive, (2) a controller with limited processing resources and buffer memory, (3) shot-term memory and working memory which includes visual acoustic, motor, and coded memories, (4) the local manager which does problem solving, learning, and procedures, (5) buffer memories for both input and output, (6) sensors, which include transducers for the senses, (7) affecters, which includes transducers for all outputs, and (8) long-term memory, which includes internal images of the world (knowledge, belief, and situation) and language (sensor data and affector data).

The model provides for specific interconnections between components that appear to occur in humans. Specifically, long term memory is affected only by short term memory but affects short term memory and buffer memories for sensors and affectors. The executive sends information to the local manager and acts in a controlling function over short term memory and the controller. The short term memory interacts with the long-term memory, receives information from sensor buffers, and interacts with the local manager. The local manager receives information from the global executive and interacts with the short term memory. The sensor observes reflections of the world and sends the resulting signals through incoming buffer memory to short and long term memory. Long term memory feeds information to output buffers that then pass the information on to

affectors.

This depiction is reflected in a different structure which models the system processes of cognition.




**Lambert's Model of Cognition**

In this depiction, we see the movement of information from senses through a cognitive process that includes reflexes, conditioned behavior, intuition, and reasoning, and a movement back down to action. Many more details are provided, but this is the general structure of cognition with which Lambert worked. From a standpoint of understanding deception, the notion is that the reflections of the world that reach the senses of the cognition system are interpreted based on its present state. The deception objective is to control those reflections so as to produce the desired changes in the perception of the target so as to achieve compliance. This can be done by inhibiting or inducing cognitive activities within this structure.

The induction of signals at the sense level is relatively obvious, and the resulting reflexive responses are quite predictable in most cases. the problems start becoming considerable as higher levels of the victim's cognitive structure get involved. while the mechanism of deception may involve the perception of feature, any feedback from this can only be seen as a result of conditioned behaviors at the perceive form level or higher level cognitive affects reflected in the ultimate drives of the system. For this reason, while the model may be helpful in understanding internal states, affects at the perceive feature level are aliased as affects at higher levels. Following the earlier depiction of deceptions as consisting of inhibitions and inducements of sensor data we can think of internal effects of deception on cognition in terms of combinations of inhibitions and inducements of internal signals. The objective of a deception might then, for example, be the inhibition of sensed content from being perceived as a feature, perhaps accomplished by a combination of reducing the available signal and distracting focus of attention by inducing the perception of a different form and causing a simultaneous reflexive action to reduce the available signal. This is precisely what is done in the case of the disappearing elephant magic trick. The disappearing elephant trick is an excellent example of the exploitation of the cognitive system and can be readily explained through Lambert's model.

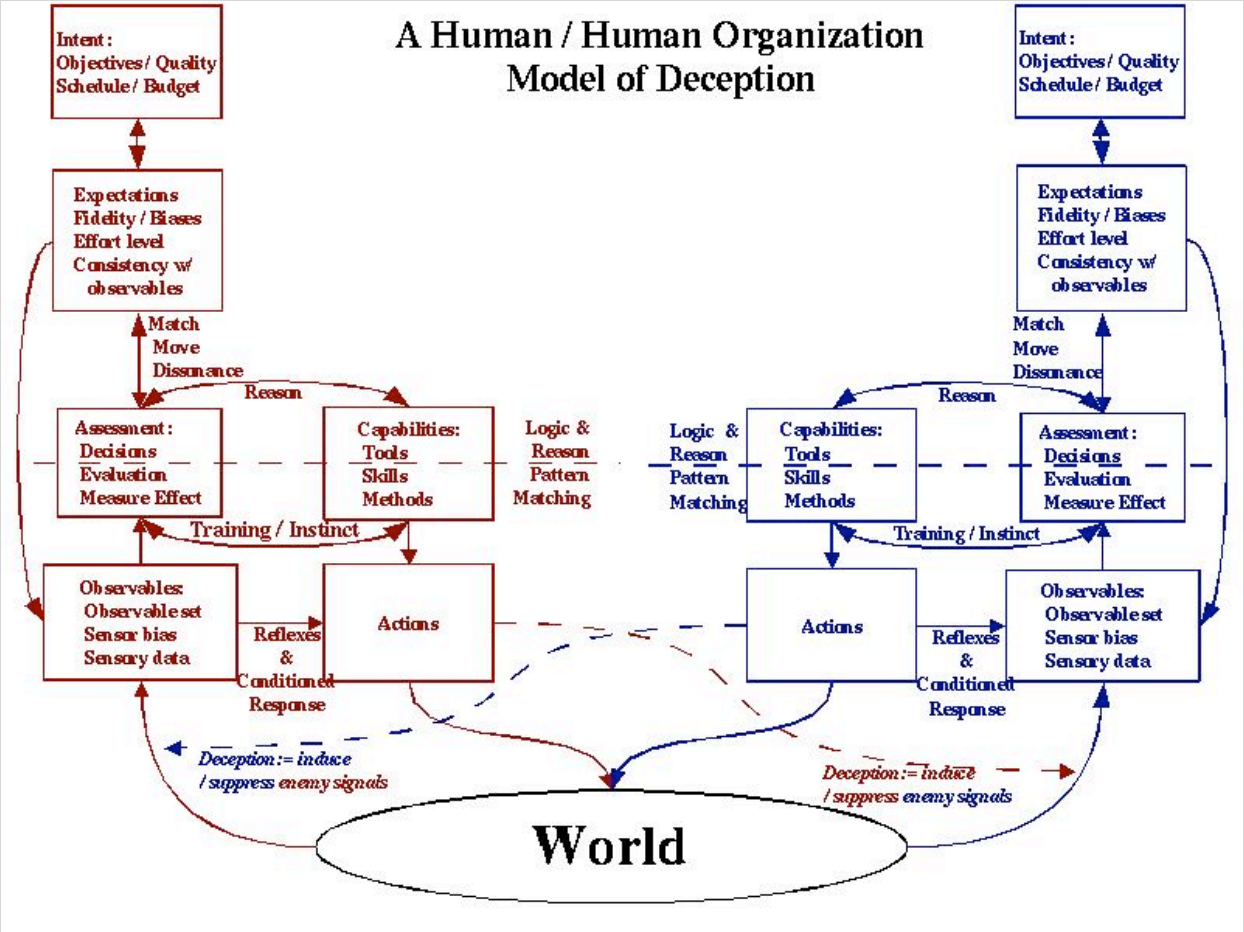
> *Example: This trick is set up by the creation of a rippling black silk curtain behind the elephant, which is gray. The audience is in a fairly close pack staring right at the elephant some distance away. Just before the elephant disappears, a scantily clad woman walks across the front of the crowd and the magician is describing something that is not very interesting with regard to the trick. Then, as eyes turn toward the side the girl is walking toward, a loud crash sound is created to that side of the crowd. They crowd's reflexive response to a crashing sound it to turn toward the sound, which they do. This takes about 1/3 to 1/2 second. As soon as they are looking that way, the magician causes another black silk rippling curtain to rise up in front of the elephant. This takes less than 1/4 second. Because of the low contrast between the elephant and the curtain and the rippling effect of the black back and front curtains, there is no edge line induced in the audience and thus attention is not pulled toward the curtains. By the time the crowd looks back, the elephant is gone and is then moved away while out of sight. The back curtain is lowered, and the front curtain is then raised to prove that only the wall remains behind the curtain.*

For low-level one-step deceptions such as this one, Lambert's model is an excellent tool both for explanation

and for planning. There are a set of known sensors, reflexes, and even well known or trainable conditioned responses that can be exploited almost at will. In some cases it will be necessary to force the cognitive system into a state where these prevail over higher level controlling processes, such as a member of the crowd who is focusing very carefully on what is going on. This can be done by boring them into relaxation, which the magician tries to do with his boring commentary and the more interesting scantily clad woman, but otherwise it is pretty straight forward. Unfortunately, this model provides inadequate structure for dealing with higher level or longer term cognitive deceptions. For these you need to move to another sort of model that, while still consistent with this model, provides added clarity regarding possible moves.

### A Cognitive Model for Higher Level Deceptions

The depiction below attempt to provide additional structure for higher level cognitive deceptions. This model starts to look at how humans interact to create deceptions and how those deceptions can, at a broad level, cause interpretation and behavior in the target that is compliant with the deceiver. It also shows the recursive nature of deception because of the regress induced by both time and symmetry.




**Model of Human Cognition for Deceptions**

The depiction shows interaction between two human or group cognitive systems. The interaction all takes place through the world using human senses (small, taste, hearing, touching, seeing, pheromones, and allergic reactions). Deception is modeled by the induction or suppression of target observables by the deceiver.

> Cognitive processes responding directly to inputs include sensory data which, after sensor bias and the filter of a set of observables, becomes observable. Sensory data, after bias, can trigger reflexive responses which also induce observable internal changes. Other actions can also be generated and expectations actively control everything in this list. Focus of attention can also be affected at this level because of detection mechanisms and their triggering of higher level processes. This paragraph summarizes what we will tentatively call the 'low level' cognitive system.
>
> Cognitive processes in, what we tentatively call, the middle level of cognition include conditioned and other automatic but non reflexive responses, measurement mechanisms and automatic or trained evaluation and decision methods, learned and nearly automated capabilities including skills, tools, and methods that are based on pattern matching, training, instinctual responses, the actions they trigger, and the feedback mechanisms involved in controlling those actions. This level also involves learned patterns of focus of attention.
>
> The remaining cognitive processes are called high level. This includes reason-based assessments and capabilities, expectations, which include biases, fidelity of interest, level of effort, consistency with observables, and high-level focus of attention, and intent, which includes objectives, qualitative evaluation, schedule and budgetary requirements. The link between expectations and the rest of the cognitive structure is particularly important because expectations alter focus of attention sequences, cognitive biases, assessment, intent, and the evaluation of expectations, while changing of expectation

can keep them stable, move them at a limited rate, or cause dissonance.

### Deceptions of Low-level Cognition

In this model, we have collapsed the lower levels (up to conditioned response) of Lambert's model into the bottom two boxes (Observables and Actions) and created a somewhat more specific higher level structure. Details of these deceptions are provided in the sections 6 and 7 of Lambert's data collection. Low-level visual deceptions are demonstrated by Seckel [15] and described by Hoffman [36]. Audio deceptions are demonstrated on an audio CD-ROM by Deutsch [47].

### Deceptions of Mid-level Cognition

The notion is that there are pattern matching and reason-based assessments and capabilities that interact to induce more thoughtful decisions than conditioned response. While pattern matching cognition mechanisms are more thoughtful than conditioned response, they are essentially the programmed behaviors identified by Cialdini [34] and some of the negotiation tactics of Karrass [33]. These include, but are not limited to, reciprocation, authority, contrast, commitment and consistency, automaticity, social proof, liking, and scarcity, and as Karrass formulates it, credibility, message content and appeal, situation setting and rewards, and media choice are all methods.

The potential for decisions to be moved to more logical reasoning exists, but this is limited by the effects identified by Gilovich [14]. Specifically, the notions that people (erroneously) believe that effects should resemble their causes, they misperceive random events, they misinterpret incomplete or unrepresentative data, they form biased evaluations of ambiguous and inconsistent data, they have motivational determinants of belief, they bias second hand information, and they have exaggerated impressions of social support. More content is provided in the sections numbered 1, 2, and some portions of 4 and 8 of Lambert's data collection.

### Deceptions of High-level Cognition

Karrass [33] also provides techniques for affecting influence in high-level thoughtful situations. He explains that change comes from learning and acceptance. Learning comes from hearing and understanding, while acceptance comes from comfort with the message, relevance, and good feelings toward the underlying idea. These are both affected by audience motives and values, the information and language used for presentation, audience attitudes and emotions, and the audience's perception and role in the negotiation. Karrrass [33] provides a three dimensional depiction of goals, needs, and perceptions and asserts that people are predictable. He also provides a set of tactics including timing, inspection, authority, association, amount, brotherhood, and detour that can be applied in a deception context. Handy also provides a set of influence tactics that tend to be most useful at higher levels of reasoning, including physicality, resources, position (which yields information, access, and right to organize), expertise, personal charisma, and emotion. More content is also provided in the sections 4 and 8 of Lambert's data collection.

### Moving from High-Level to Mid-level Cognition

Karrass also augments Cialdini's notions [34] of rush, stress, uncertainty, indifference, distraction, and fatigue leading to less thoughtful and more automatic responses and brings out Maslow's needs hierarchy (basic survival, safety, love, self worth, and self-actualization). By forcing earlier sets of these issues, reasoning can be driven away and replaced by increased automaticity. Tactics of timing can also be used to drive people toward increased automaticity. Thus we can either drive the target toward less thought or use Karrass's methods of negotiation to cause desired change.

### Moving from Mid-Level to High-level Cognition

Cognition moves to higher levels only when there are intent-based forcing factors that lead to deeper analysis, (e.g., when objectives are oriented toward more in-depth thought, quality requirements drive more detailed consideration, schedule availability provides free time to do deeper consideration, or extra budget is available for this purpose) or when expectations are not met (i.e., the fidelity of the deception is inadequate, biases trigger more detailed examination, inconsistencies or errors are above some threshold, or the difference between expectations and observations is so great or changing at so great a rate as to cause dissonance). In these cases, higher levels of reasoning are applied, complete with all of their potential logical fallacies and their special skills, tools, and methods. Higher level reasoning is desired when we wish to change intent or make radical changes in expectations, while we try to drive decisions to lower cognitive levels when we can induce less thoughtful responses in our favor.

### An Example

To get a sense of how the model might be applied to deceptions, we have included a sample analysis of a simple human deception. The deception is an attack with a guard at a gate as the target. It happens many times each day and is commonly called tailgating.

> *The target of this deception is the guard and our method will be to try to exploit a natural overload that takes place during the return from lunch hour on a Thursday. We choose the end of the lunch hour on Thursday because the guard will be as busy as they ever get and because they will be looking forward to the weekend and will probably have a somewhat reduced alertness level. Thus we are intentionally trying to keep processing at a pattern matching level by increased rush, stress, indifference, distraction, and fatigue.*
>
> *We stand casually out of the guard's sight before the crowd comes along, join the crowd as it approaches the entry, hold a notepad where a badge appears on other peoples' attire, and stay away from the guard's side of the group. Our clothing and appearance is such that it avoids dissonance with the guard's expectations and does not affect the*

*guard's intent in any obvious way.*

*We tag along in the third row back near someone that looks generally like us and, when the guard is checking one of the other people, we ease our way over to the other side of the guard, appearing to be in the already checked group. Here we are using automaticity and social proof against the guard and liking by similarity against the group we are tailgaiting with. We are also using similarity to avoid triggering sensory detection and indifference, distraction and fatigue to avoid triggering higher level cognition.*

*As the group proceeds, so do we. After getting beyond the guard's sight, we move to the back of the group and drop out as they round a corner. Here we are using automaticity, liking, and social proof against the group to go along with them, followed by moving slowly out of their notice which exploits slow movement of expectations followed by concealment from observation.*

Team members have used variations on this entry technique in red teaming exercises against facilities from time to time and have been almost universally successful in its use. It is widely published and well known to be effective. It is clearly a deception because if the guard knew you were trying to get past without a badge or authorization they would not permit the entry. While the people who use it don't typically go through this analytical process at a conscious level, they do some part of it at some level and we postulate that this is why they succeed at it so frequently.

As an aside, there should always be a backup plan for such deceptions. The typical tailgaiter, if detected, will act lost and ask the guard how to get to some building or office, perhaps finding out that this is the wrong address in the process. This again exploits elements of the deception framework designed to move the guard away from high level cognition and toward automaticity that would favor letting the attacker go and not reporting the incident.

In the control system isomorphism, we can consider this same structure as attempting to maintain internal consistency and allow change only at a limited rate. The high level control system is essentially oblivious to anything unless change happens at too high a rate or deviations of high level signals from expectations are too high. Similarly, the middle levels operate using Cialdini's rules of thumb unless a disturbance at a lower level prompts obvious dissonance and low-level control decisions (e.g., remain balanced) don't get above the reflexive and conditioned response levels unless their is a control system failure.

---

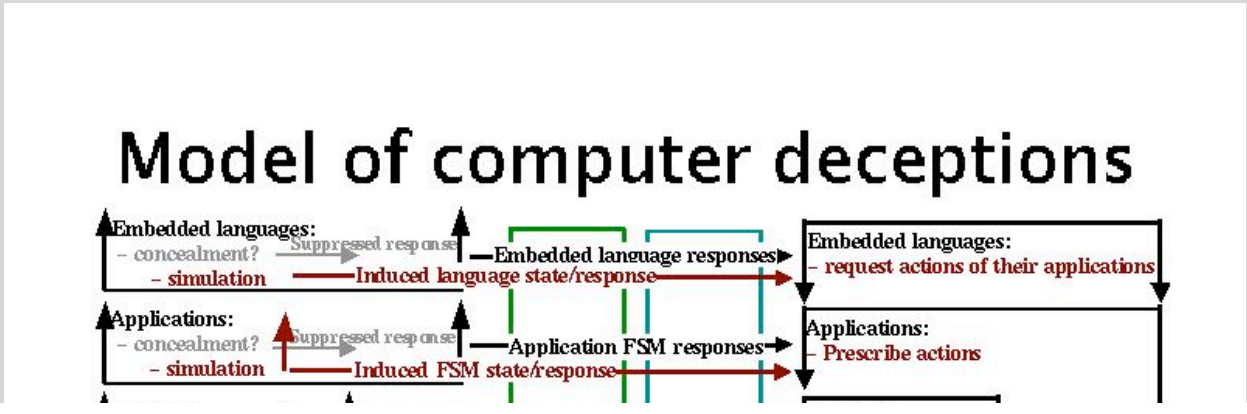
## A Model for Computer Deception

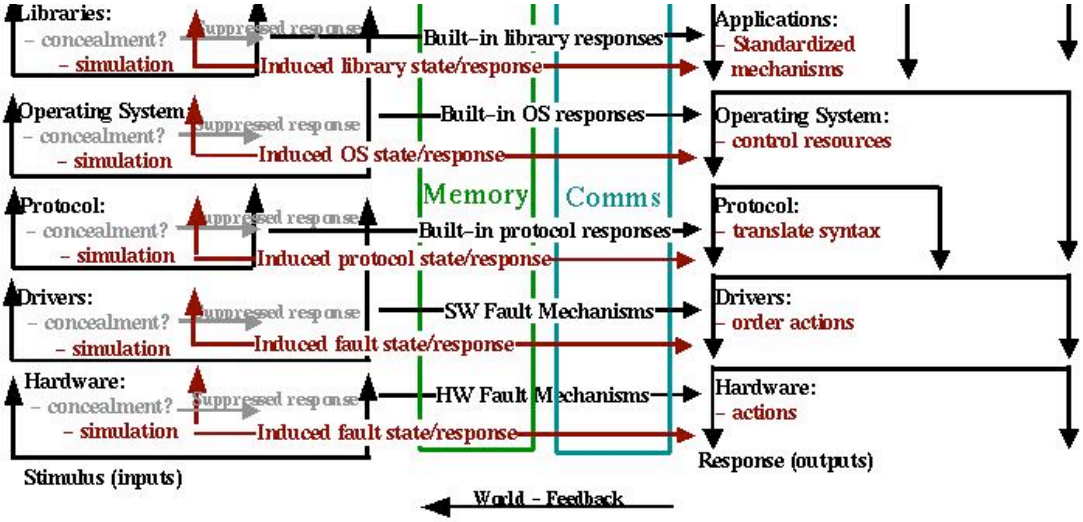
In looking at computer deceptions it is fundamental to understand that the computer is an automaton. Anthropomorphising it into an intelligent being is a mistake in this context - a self-deception. Fundamentally, deceptions must cause systems to do things differently based on their lack of ability to differentiate deception from a non-deception. Computers cannot really yet be called 'aware' in the sense of people. Therefore, when we use a deception against a computer we are really using a deception against the skills of the human(s) that design, program, and use the computer.

In many ways computers could be better at detecting deceptions than people because of their tremendous logical analysis capability and the fact that the logical processes used by computers are normally quite different than the processes used by people. This provides some level of redundancy and, in general, redundancy is a way to defeat corruption. Fortunately for those of us looking to do defensive deception against automated systems, most of the designers of modern attack technology have a tendency to minimize their programming effort and thus tend not to include a lot of redundancy in their analysis.

People use shortcuts in their programs just as they use shortcuts their thinking. Their goal is to get to an answer quickly and in many cases without adequate information to make definitive selections. Computer power and memory are limited just like human brain power and memory are limited. In order to make efficient use of resources, people write programs that jump to premature conclusions and fail to completely verify content. In addition, people who observe computer output have a tendency to believe it. Therefore, if we can deceive the automation used by people to make decisions, we may often be able to deceive the users and avoid in-depth analysis.

Our model for computer deception starts with Cohen's "Structure of Intrusion and Intrusion Detection". [3] In this model, a computer system and its vulnerabilities are described in terms of intrusions at the hardware, device driver, protocol, operating system, library and support function, application, recursive language, and meaning vs. content levels. The levels are all able to interact, but they usually interact hierarchically with each level interacting with the ones just above and below it. This model is depicted in the following graphic:






**Model of computer cognition with deceptions**

This model is based on the notion that at every level of the computer's cognitive hierarchy signals can either be induced or inhibited. The normal process is shown in black, while inhibitions are shown as grey'ed out signals, and induced signals are shown in red. All of these effect memory states and processor activities at other, typically adjacent, levels of the cognitive system. Deception detection and response capabilities are key issues in the ability to defend against deceptions so there is a concentration on the limits of detection in the following discussions.

## Hardware Level Deceptions

If the hardware of a system or network is altered, it may behave arbitrarily differently than expected. While there is a great deal of history of tamper-detection mechanisms for physical systems, no such mechanism is or likely ever will be perfect. The use of intrusion detection systems for detecting improper modifications to hardware today consist primarily of built-in self-test mechanisms such as the power on self test (POST) routine in a typical personal computer (PC). These mechanisms are designed to detect specific sorts of random stochastic fault types and are not designed to detect malicious alterations. Thus deception of these mechanisms is fairly easy to do without otherwise altering their value in detecting fault types they already detect.

Clearly, if the hardware is altered by a serious intruder, this sort of test will not be revealing. Motion sensors, physical seals of different sorts, and even devices that examine the physical characteristics of other devices are all examples of intrusion detection techniques that may work at this level. In software, we may detect alterations in external behavior due to hardware modification, but this is only effective in large scale alterations such as the implanting of additional infrastructure. This is also likely to be ignored in most modern systems because intervening infrastructure is rarely known or characterized as part of intrusion detection and operating environments are intentionally designed to abstract details of the hardware.

Intrusions can also be the result of the interaction of hardware of different sorts rather than the specific use of a particular type of hardware. This type of intrusion mechanism appears to be well beyond the capability of current technology to detect or analyze. Deceptions exploiting these interactions will therefore likely go undetected for extended period of time. Hardware-level deceptions designed to induce desired observables are relatively easy to create and hard to detect. Induction of signals requires only knowledge of protocol and proper design of devices.

The problem with using hardware level deception for defense against serious threat types is that it requires physical access to the target system or logical access with capabilities to alter hardware level functions (e.g., microcode access). This tends to be difficult to attain against intelligence targets, if attempted against insiders it introduces deceptions that could be used against the defenders, and in the case of overrun, it does not seem feasible. That is not to say that we cannot use deceptions that operate at the hardware level against systems, but rather that affecting their hardware level is likely to be infeasible.

## Driver Level Deceptions

Drivers are typically ignored by intrusion detection and other security systems. They are rarely inspected, in modern operating systems they can often be installed from or by applications, and they usually have unlimited hardware access. This makes them prime candidates for exploitations of all sorts, including deceptions.

A typical driver level deception would cause the driver to process items of interest without passing information to other parts of the operating environment or to exfiltrate information without allowing the system to notice that this activity was happening. It would be easy for the driver to cause widespread corruption of arbitrary other elements of the system as well as inhibiting the system from seeing undesired content.

From a standpoint of defensive deceptions, drivers are very good target candidates. A typical scenario is to require that a particular driver be installed in order to gain access to defended sites. This is commonly done with applications like RealAudio. Once the target loads the required driver, hardware level access is granted and arbitrary exploits can be launched. This technique is offensive in nature and may violate rules of engagement in a military setting or induce civil or criminal liability in a civilian setting. Its use for defensive purposes may be

overly aggressive.

### Protocol Level Deceptions

Many protocol intrusions have been demonstrated, ranging from exploitations of flaws in the IP protocol suite to flaws in cryptographic protocols. Except for a small list of known flaws that are part of active exploitations, most current intrusion detection systems do not detect such vulnerabilities. In order to fully cover such attacks, it would likely be necessary for such a system to examine and model the entire network state and effects of all packets and be able to differentiate between acceptable and unacceptable packets.

Although this might be feasible in some circumstances, the more common approach is to differentiate between protocols that are allowed and those that are not. Increasing granularity can be used to differentiate based on location, time, protocol type, packet size and makeup, and other protocol-level information. This can be done today at the level of single packets, or in some circumstances, limited sequences of packets, but it is not feasible for the combinations of packets that come from different sources and might interact within the end systems. Large scale effects can sometimes be detected, such as aggregate bandwidth utilization, but without a good model of what is supposed to happen, there will always be malicious protocol sequences that go undetected. There are also interactions between hardware and protocols. For example, there may be an exploitation of a particular hardware device which is susceptible to a particular protocol state transition, resulting in a subtle alteration to normal timing behaviors. This might then be used to exfiltrate information based on any number of factors, including very subtle covert channels.

Defensive protocol level deceptions have proven relatively easy to develop and hard to defeat. Deception ToolKit [6] and D-WALL [7] both use protocol level deceptions to great effect and these are relatively simplistic mechanisms compared to what could be devised with substantial time and effort. This appears to be a ripe area for further work. Most intelligence gathering today starts at the protocol level, overrun situations almost universally result in communication with other systems at the protocol level, and insiders generally access other systems in the environment through the protocol level.

### Operating System Level Deceptions

At the operating system (OS) level, there are a very large number of intrusions possible, and not all of them come from packets that come over networks. Users can circumvent operating system protection in a wide variety of ways. For a successful intrusion detection system to work, it has to detect this before the attacker gains the access necessary to disable the intrusion detection mechanisms (the sensors, fusion, analysis, or response elements or the links between them can be defeated to avoid successful detection). In the late 1980s a lot of work was done in the limitations of the ability of systems to protect themselves and integrity-based self defense mechanisms were implemented that could do a reasonable job of detecting alterations to operating systems. [51] These systems are not capable of defeating attacks that invade the operating system without altering files and reenter the operating system from another level after the system is functioning. Process-based intrusion detection has also been implemented with limited success. Thus we see that operating system level deceptions are commonplace and difficult to defend against.

Any host-based IDS and the analytical part of any network-based IDS involves some sort of operating environment that may be defeatable. But even if defeat is not directly attainable, denial of services against the components of the IDS can defeat many IDS mechanisms, replay attacks may defeat keep-alive protocols used to counter these denial of service attacks, selective denial of service against only desired detections are often possible, and the list goes on and on. If the operating systems are not secure, the IDS has to win a battle of time in order to be effective at detecting things it is designed to detect. Thus we see that the induction or suppression of signals into the IDS can be used to enhance or cover operating system level deceptions that might otherwise be detected.

Operating systems can have complex interactions with other operating systems in the environment as well as between the different programs operating within the OS environment. For example, variations in the timing of two processes might cause race conditions that are extremely rare but which can be induced through timing of otherwise valid external factors. Heavy usage periods may increase the likelihood of such subtle interactions, and thus the same methods that would not work under test conditions may be inducible in live systems during periods of high load. An IDS would have to detect this condition and, of course, because of the high load the IDS would be contributing to the load as well as susceptible to the effects of the attack. A specific example is the loading of a system to the point where there are no available file handles in the system tables. At this point, the IDS may not be able to open the necessary communications channels to detect, record, analyze, or respond to an intrusion.

Operating systems may also have complex interactions with protocols and hardware conditions, and these interactions are extremely complex to analyze. To date, nobody has produced an analysis of such interactions as far as we are aware. Thus deceptions based on mixed levels including the OS are likely to be undetected as deceptions.

Of course an IDS cannot detect all of the possible OS attacks. There are systems which can detect known attacks, detect anomalous behavior by select programs, and so forth, but again, a follow-up investigation is required in order for these methods to be effective, and a potentially infinite number of attacks exist that do not trip anomaly detection methods. If the environment can be characterized closely enough, it may be feasible to detect the vast majority of these attacks, but even if you could do this perfectly, there is then the library and support function level intrusion that must be addressed.

Operating systems are the most common point of attack against systems today largely because they afford a tremendous amount of cover and capability. They provide cover because of their enormous complexity and capability. They have unlimited access within the system and the ability to control the hardware so as to yield arbitrary external effects and observables. They try to control access to themselves, and thus higher level

programs do not have the opportunity to measure them for the presence of deceptions. They also seek to protect themselves from the outside world so that external assessment is blocked. While they are not perfect at either of these types of protection, they are effective against the rest of the cognitive system they support. As a location for deception, they are thus prime candidates.

To use defensive deception at the target's operating system level requires offensive actions on the part of the deceiver and yields only indirect control over the target's cognitive capability. This has to then be exploited in order to affect deceptions at other levels and this exploitation may be very complex depending on the specific objective of the deception.

### Library and Support Function Level Intrusions

Libraries and support functions are often embedded within a system and are largely hidden from the programmer so that their role is not as apparent as either operating system calls or application level programs. A good example of this is in languages like C wherein the language has embedded sets of functions that are provided to automate many of the functions that would otherwise have to be written by programmers. For example the C strings library includes a wide range of widely used functions. Unfortunately, the implementations of these functions are not standardized and often contain errors that become embedded in every program in the environment that uses them. Library-level intrusion detection has not been demonstrated at this time other than by the change detection methodology supported by the integrity-based systems of the late 1980s and behavioral detection at the operating system level. Most of the IDS mechanisms themselves depend on libraries.

An excellent recent example is the use of leading zeros in numerical values in some Unix systems. On one system call, the string -08 produces an error, while in another it is translated into the integer -8. This was traced to a library function that is very widely used. It was tested on a wide range of systems with different results on different versions of libraries in different operating environments. These libraries are so deeply embedded in operating environments and so transparent to most programmers that minor changes may have disastrous effects on system integrity and produce enormous opportunities for exploitation. Libraries are almost universally delivered in loadable form only so that source codes are only available through considerable effort. Trojan horses, simple errors, or system-to-system differences in libraries can make even the most well written and secure applications an opportunity for exploitation. This includes system applications, commonly considered part of the operating system, service applications such as web servers, accounting systems, and databases, and user level applications including custom programs and host-based intrusion detection systems.

The high level of interaction of libraries is a symptom of the general intrusion detection problem. Libraries sometimes interact directly with hardware, such as the libraries that are commonly used in special device functions like writing CD-rewritable disks. In many modern operating systems, libraries can be loaded as parts of device drivers that become embedded in the operating system itself at the hardware control level. A hardware device with a subtle interaction with a library function can be exploited in an intrusion, and the notion that any modern IDS would be able to detect this is highly suspect. While some IDS systems might detect some of the effects of this sort of attack, the underlying loss of trust in the operating environments resulting from such an embedded corruption is plainly outside of the structure of intrusion detection used today.

Using library functions for defensive deceptions offers great opportunity but, like operating systems, there are limits to the effectiveness of libraries because they are at a level below that used by higher level cognitive functions and thus there is great complexity in producing just the right effects without providing obvious evidence that something is not right.

### Application Level Deceptions

Applications provide many new opportunities for deceptions. The apparent user interface languages offer syntax and semantics that may be exploited while the actual user interface languages may differ from the apparent languages because of programming errors, back doors, and unanticipated interactions. Internal semantics may be in error, may fail to take all possible situations into account, or there may be interactions with other programs in the environment or with state information held by the operating environment. They always trust the data they receive so that false content is easily generated and efficient. These include most intelligence tools, exploits, and other tools and techniques used by severe threats. Known attack detection tools and anomaly detection have been applied at the application level with limited success. Network detection mechanisms also tend to operate at the application level for select known application vulnerabilities.

As in every other level, there may be interactions across levels. The interaction of an application program with a library may allow a remote user to generate a complex set of interactions causing unexpected values to appear in inter-program calls, within programs, or within the operating system itself. It is most common for programmers to assume that system calls and library calls will not produce errors, and most programming environments are poor at handling all possible errors. If the programmer misses a single exception - even one that is not documented because it results from an undiscovered error in an interaction that was not anticipated - the application program may halt unexpectedly, produce incorrect results, pass incorrect information to another application, or enter an inconsistent internal state. This may be under the control of a remote attacker who has analyzed or planned such an interaction. Modern intrusion detection systems are not prepared to detect this sort of interaction.

Application level defensive deceptions are very likely to be a major area of interest because applications tend to be driven more by time to market than by surety and because applications tend to directly influence the decision processes made by attackers. For example, a defensive deception would typically cause a network scanner to make wrong decisions and report wrong results to the intelligence operative using it. Similarly, an application level deception might be used to cause a system that is overrun to act on the wrong data. For systems administrators the problem is somewhat more complex and it is less likely that application-level deceptions will work against them.

## Recursive Languages in the Operating Environment

In many cases, application programs encode Turing Machine capable embedded languages, such as a language interpreter. Examples include Java, Basic, Lisp, APL, and Word Macros. If these languages can interpret user-level programs, there is an unlimited possible set of embedded languages that can be devised by the user or anybody the user trusts. Clearly an intrusion detection system cannot anticipate all possible errors and interactions in this recursive set of languages. This is an undecidable problem that no IDS will ever likely be able to address. Current IDS systems only address this to the extent that anomaly detection may detect changes in the behavior of the underlying application, but this is unlikely to be effective.

These recursive languages have the potential to create subtle interactions with all other levels of the environment. For example, such a language could consume excessive resources, use a graphical interface to make it appear as if it were no longer operating while actually interpreting all user input and mediating all user output, test out a wide range of known language and library interactions until it found an exploitable error, and on and on. The possibilities are literally endless. All attempts to use language constructs to defeat such attacks have failed to date, and even if they were to succeed to a limited extent, any success in this area would not be due to intrusion detection capabilities.

It seems that no intrusion detection system will ever have a serious hope of detecting errors induced at these recursive language levels as long as we continue to have user-defined languages that we trust to make decisions affecting substantial value. Unless the IDS is able to 'understand' the semantics of every level of the implementation and make determinations that differentiate desirable intent from malicious intent, the IDS cannot hope to mediate decisions that have implications on resulting values. This is clearly impossible,

Recursive languages are used in many applications including many intelligence and systems administration applications. In cases where this can be defined or understood or cases where the recursive language itself acts as the application, deceptions against these recursive languages should work in much the same manner as deceptions against the applications themselves.

## The Meaning of the Content versus Realities

Content is generally associated with meaning in any meaningful application. The correspondence between content and realities of the world cannot reasonably be tracked by an intrusion detection system, is rarely tracked by applications, and cannot practically be tracked by other levels of the system structure because it is highly dependent on the semantics of the application that interprets it. Deceptions often involve generating human misperceptions or causing people to do the wrong thing based on what they see at the user interface. In the end, if this wrong thing corresponds to a making a different decision than is supposed to be made, but still a decision that is a feasible and reasonable one in a slightly different context, only somebody capable of making the judgment independently has any hope of detecting the error.

Only certain sorts of input redundancy are known to be capable of detecting this sort of intrusion and this becomes cost prohibitive in any large-scale operation. This sort of detection is used in some high surety critical applications, but not in most intelligence applications, most overrun situations, or by most systems administrators. The programmers of these systems call this "defensive programming" or some such thing and tend to fight against its use.

Attackers commonly use what they call 'social engineering' (a.k.a., perception management) to cause the human operator to do the wrong thing. Of course such behavioral changes can ripple through the system as well, ranging from entering wrong data to changing application level parameters to providing system passwords to loading new software updates from a web site to changing a hardware setting. All of the other levels are potentially affected by this sort of interaction.

Ultimately, deception in information systems intended to affect other systems or people will cause results at this level and thus all deceptions of this sort are well served to consider this level in their assessments.

## Commentary

Unlike people, computers don't typically have ego, but they do have built-in expectations and in some cases automatically seek to attain 'goals'. If those expectations and goals can be met or encouraged while carrying out the deception, the computers will fall prey just as people do.

In order to be very successful at defeating computers through deception, there are three basic approaches. One approach is to create as high a fidelity deception as you can and hope that the computer will be fooled. Another is to understand what data the computer is collecting and how it analyzes the data provided to it. The third it to alter the function of the computer to comply with your needs. The high fidelity approach can be quite expensive but should not be abandoned out of hand. At the same time, the approach of understanding enemy tools can never be done definitively without a tremendous intelligence capability. The modification of cognition approach requires an offensive capability that is not always available and is quite often illegal, but all three avenues appear to be worth pursuing.

> **High Fidelity:** High fidelity deception of computers with regard to their assessment, analysis, and use against other computers tends to be fairly easy to accomplish today using tools like D-WALL [7] and the IR effort associated with this project. D-WALL created high fidelity deception by rerouting attacks toward substitute systems. The IR does a very similar process in some of its modes of operation. The notion is that by providing a real system to attack, the attacker is suitably entertained. While this is effective in the generic sense, for specific systems, additional effort must be made to create the internal system conditions indicative of the desired deception environment. This can be quite costly. These deceptions tend to operate at a protocol level and are augmented by other technologies to effect other levels of

deception.

**Defeating Specific Tools:** Many specific tools are defeated by specific deception techniques. For example, nmap and similar scans of a network seeking out services to exploit are easily defeated by tools like the Deception ToolKit. [6] More specific attack tools such as Back Orafice (BO) can be directly countered by specific emulators such as "NoBO" - a PC-based tool that emulates a system that has already been subverted with BO. Some deception systems work against substantial classes of attack tools.

**Modifying Function:** Modifying the function of computers is relatively easy to do and is commonly used in attacks. The question of legality aside, the technical aspects of modifying function for defense falls into the area of counterattack and is thus not a purely defensive operation. The basic plan is to gain access, expand privileges, induce desired changes for ultimate compliance, leave those changes in place, periodically verify proper operation, and exploit as desired. In some cases privileges gained in one system are used to attack other systems as well. Modified function is particularly useful for getting feedback on target cognition.

The intelligence requirements of defeating specific tools may be severe, but the extremely low cost of such defenses makes them appealing. Against off-the-Internet attack tools, these defenses are commonly effective and, at a minimum, increase the cost of attack far more than they affect the cost of defense. Unfortunately, for more severe threats, such as insiders, overrun situations, and intelligence organizations, these defenses are often inadequate. They are almost certain to be detected and avoided by an attacker with skills and access of this sort. Nevertheless, from a standpoint of defeating the automation used by these types of attackers, relatively low-level deceptions have proven effective. In the case of modifying target systems, the problems become more severe in the case of more severe threats. Insiders are using your systems, so modifying them to allow for deception allows for self-deception and enemy deception of you. For overrun conditions you rarely have access to the target system, so unless you can do very rapid and automated modification, this tactic will likely fail. For intelligence operations this requires that you defeat an intelligence organization one of whose tasks is to deceive you. The implications are unpleasant and inadequate study has been made in this area to make definitive decisions.

There is a general method of deception against computer systems being used to launch fully automated attacks against other computer systems. The general method is to analyze the attacking system (the target) in terms of its use of responses from the defender and create sequences of responses that emulate the desired responses to the target. Because all such mechanisms published or widely used today are quite finite and relatively simplistic, with substantial knowledge of the attack mechanism, it is relatively easy to create a low-quality deception that will be effective. It is noteworthy, for example, that the Deception ToolKit[6], which was made publicly available in source form in 1998, is still almost completely effective against automated intelligence tools attempting to detect vulnerabilities. It seems that the widely used attack tools are not yet being designed to detect and counter deception.

That is not to say that red teams and intelligence agencies are not beginning to start to look at this issue. For example, in private conversations with defenders against select elite red teams the question often comes up of how to defeat the attackers when they undergo a substantial intelligence effort directed at defeating their attempts at deceptive defense. The answer is to increase the fidelity of the deception. This has associated costs, but as the attack tools designed to counter deception improve, so will the requirement for higher fidelity in deceptions.

## Deception Mechanisms for Information Systems

This content is extracted from a previous paper on attack mechanisms [48] and is intended to summarize the attack mechanisms that are viable deception techniques against information systems - in the sense that they induce or inhibit cognition at some level. All of the attack techniques in the original paper may be used as parts of overall deception processes, but only these are specifically useful as deception methods and specifically oriented toward information technology as opposed to the people that use and control these systems. We have explicitly excluded mechanisms used for observation only and included examples of how these techniques affect cognition and thus assist in deception and added information about deception levels in the target system.

| Mechanism | Levels |
|---|---|
| Cable cuts | HW |
| Fire | HW |
| Flood | HW |
| Earth movement | HW |
| Environmental control loss | HW |
| System maintenance | All |
| Trojan horses | All |
| Fictitious people | All |
| Resource availability manipulation | HW, OS |
| Spoofing and masquerading | All |
| Infrastructure interference | HW |
| Insertion in transit | All |
| Modification in transit | All |
| Sympathetic vibration | All |
| Cascade failures | All |

| | |
|---|---|
| Invalid values on calls | OS and up |
| Undocumented or unknown function exploitation | All |
| Excess privilege exploitation | App, Driver |
| Environment corruption | All |
| Device access exploitation | HW, Driver |
| Modeling mismatches | App and up |
| Simultaneous access exploitations | All |
| Implied trust exploitation | All |
| Interrupt sequence mishandling | Driver, OS |
| Emergency procedure exploitation | All |
| Desychronization and time-based attacks | All |
| Imperfect daemon exploits | Lib, App |
| Multiple error inducement | All |
| Viruses | All |
| Data diddling | OS and up |
| Electronic interference | HW |
| Repair-replace-remove information | All |
| Wire closet attacks | HW |
| Process bypassing | All |
| Content-based attacks | Lib and up |
| Restoration process corruption or misuse | Lib and up |
| Hangup hooking | HW, Lib, Driver, OS |
| Call forwarding fakery | HW |
| Input overflow | All |
| Illegal value insertion | All |
| Privileged program misuse | App, OS, Driver |
| Error-induced misoperation | All |
| Audit suppression | All |
| Induced stress failures | All |
| False updates | All |
| Network service and protocol attacks | HW, Driver, Proto |
| Distributed coordinated attacks | All |
| Man-in-the-middle | HW, Proto |
| Replay attacks | Proto, App, and up |
| Error insertion and analysis | All |
| Reflexive control | All |
| Dependency analysis and exploitation | All |
| Interprocess communication attacks | OS, Lib, Proto, App |
| Below-threshold attacks | All |
| Peer relationship exploitation | Proto, App, and up |
| Piggybacking | All |
| Collaborative misuse | All |
| Race conditions | All |
| Kiting | App and up |
| Salami attacks | App and up |
| Repudiation | App and up |

---

## Models of Deception of More Complex Systems

Larger cognitive systems can me modeled as being built up from smaller cognitive subsystems through some composition mechanism. Using these combined models we may analyze and create larger scale deceptions. To date there is no really good theory of composition for these sorts of systems and attempts to build theories of composition for security properties of even relatively simple computer networks have proven rather difficult. We can also take a top-down approach, but without the ability to link top-level objectives to bottom-level capabilities and without metrics for comparing alternatives, the problem space grows rapidly and results cannot be meaningfully compared.

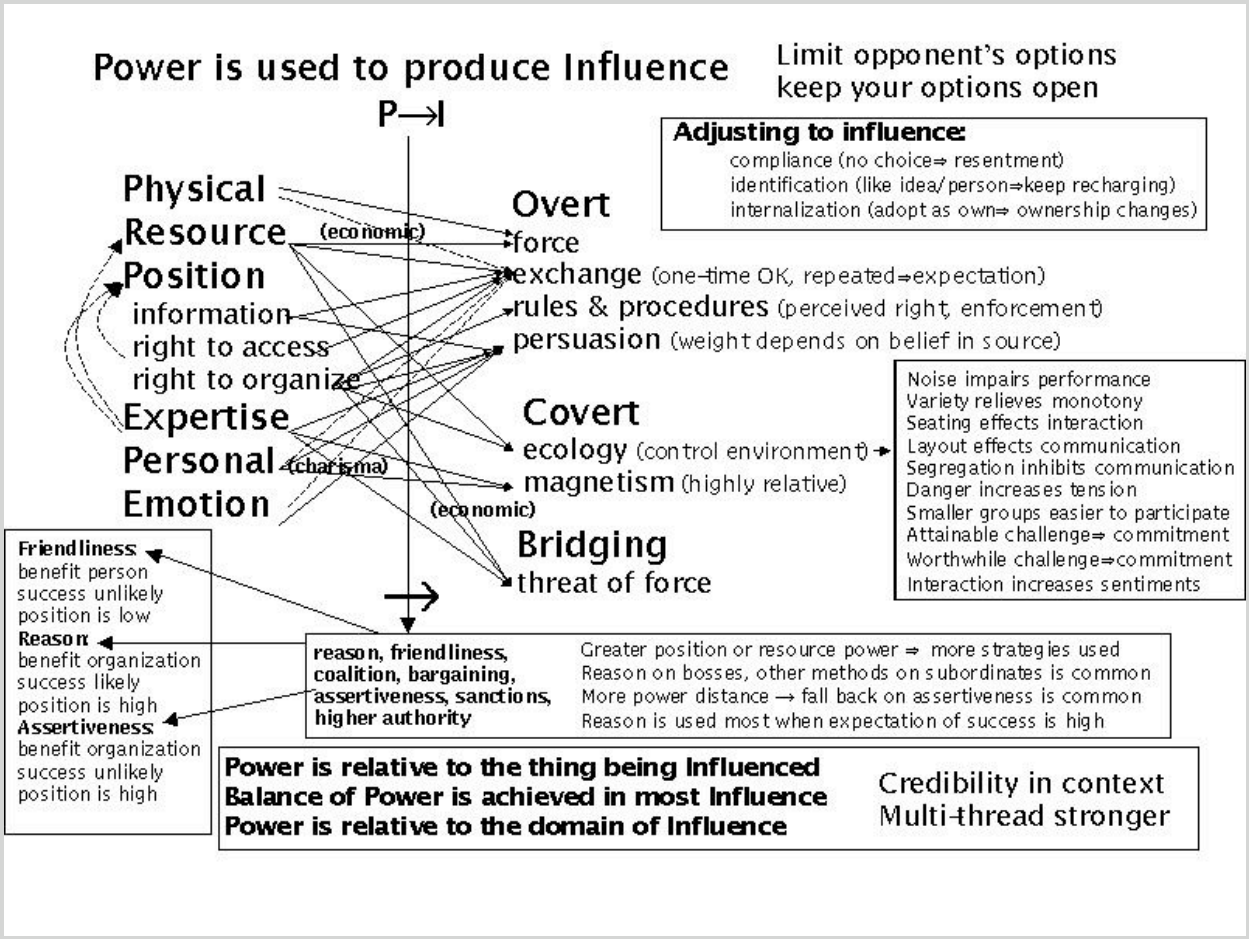
### Human Organizations

Humans operating in organizations and groups of all sorts have been extensively studied, but deception results

in this field are quite limited. The work of Karrass [33] (described earlier) deals with issues of negotiations involving small groups of people, but is not extended beyond that point. Military intelligence failures make good examples of organizational deceptions in which one organization attempts to deceive another. Hughes-Wilson describes failures in collection, fusion, analysis, interpretation, reporting, and listening to what intelligence is saying as the prime causes of intelligence blunders, and at the same time indicates that generating these conditions generally involved imperfect organizationally-oriented deceptions by the enemy. [54] John Keegan details a lot of the history of warfare and along the way described many of the deceptions that resulted in tactical advantage. [55] Dunnigan and Nofi detail many examples of deception in warfare and, in some cases, detail how deceptions have affected organizations. [8] Strategic military deceptions have been carried out for a long time, but the theory of how the operations of groups lead to deception has never really been worked out. What we seem to have, from the time of Sun Tzu [28] to the modern day, [57] is sets of rules that have withstood the test of time. Statements like *"It is far easier to lead a target astray by reinforcing the target's existing beliefs" [57, p42]* are stated and restated without deeper understanding, without any way to measure the limits of its effectiveness, and without a way to determine what beliefs an organization has. It sometimes seems we have not made substantial progress from when Sun Tzu originally told us that "All warfare is based on deception.

The systematic study of group deception has been under way for some time. In 1841, Mackay released his still famous and widely read book titled "Extraordinary Popular Delusions and the Madness of Crowds" [56] in which he gives detailed accounts of the history of the largest scale deceptions and financial 'bubbles' of history to that time. It is astounding how relevant this is to modern times. For example, the recent bubble in the stock market related to the emergence of the Internet is incredibly similar to historical bubbles, as are the aftermaths of all of these events. The self-sustaining unwarranted optimism, the self fulfilling prophecies, the participation even by the skeptics, the exit of the originators, and the eventual bursting of the bubble to the detriment of the general public, all seem to operate even though the participants are well aware of the nature of the situation. While Mackay offers no detailed psychological accounting of the underlying mechanisms, he clearly describes the patterns of behavior in crowds that lead to this sort of group insanity.

Charles Handy [37] describes how power and influence work in organizations. This leads to methods by which people with different sorts of power create changes in the overall organizational perspective and decision process. In deceptions of organizations, models of who stands where on which issues and methods to move them are vital to determining who to influence and in what manner in order to get the organization to move.




**Power and Influence in Human Organizations**

These principles have been applied without rigor and with substantial success for a long time.

> *Example: In World War II Germany, Hitler was the target of many of the allied strategic deceptions because the German organs of state were designed to grant him unlimited power. It didn't matter that Romel believed that the allies would attack at Normandy because Hitler was convinced that they would strike at Pas de Calais. All dictatorial regimes tend to be swayed by influencing the mind of a single key decision maker. At the same time we should not make the mistake of believing that this works at a tactical level. The German military in World War II was highly*

*skilled at local decision making and field commanders were trained to innovate and take command when in command.*

Military hierarchies tend to operate this way to a point, however, most military Juntas have a group decision process that significantly complicates this issue. For example, the FARC in Colombia have councils that make group decisions and cannot be swayed by convincing a single authority figure. Swaying the United States is very a complex process, while swaying Iraq is considerably easier, at least from a standpoint of identifying the target of deceptions. The previously cited works on individual human deception certainly provide us with the requisite rational for explaining individual tendencies and the creation of conditions that tend to induce more advantageous behaviors in select circumstances, but how this translates into groups is a somewhat different issue.

Organizations have many different structures, but those who study the issue [37] have identified 4 classes of organizational structure that are most often encountered and which have specific power and influence associations: hierarchy, star, matrix, and network. In hierarchies orders come from above and reporting is done from lower level to higher level in steps. Going "over a supervisor's head" is considered bad form and is usually punished. These sorts of organizations tend to be driven by top level views and it is hard to influence substantial action except at the highest levels. In a star system all personnel report to a single central point. In small organizations this works well, but the center tends to be easily overloaded as the organization grows or as more and more information is fed into it. Matrix organizations tend to cause all of the individuals to have to serve more than one master (or at least manager). In these cases there is some redundancy, but the risk of inconsistent messages from above and selective information below exists. In a network organization, people form cliques and there is a tendency for information not to get everywhere it might be helpful to have it. Each organizational type has it features and advantages, and each has different deception susceptibility characteristics resulting from these structural features. Many organizations have mixes of these structures within them.

Deceptions within a group typically include; (1) members deceive other members, (2) members deceive themselves (e.g., "group think"), and (3) leader deceives members. Deception between groups typically include (1) leader deceives leader and (2) leader deceives own group members. Self deception applies to the individual acting alone.

> *Example: "group think", in which the whole organization may be mislead due to group processes/social norms. Many members of the German population in World War II became murderous even though under normal circumstances they never would have done the things they did.*

Complex organizations require more complex plans for altering decision processes. An effective deception against a typical government or large corporation may involve understanding a lot about organizational dynamics and happens in parallel with other forces that are also trying to sway the decision process in other directions. In such situations, the movement of key decision makers in specific ways tends to be critical to success, and this in turn depends on gaining access to their observables and achieving focus or lack of focus when and where appropriate. This can then lead to the need to gain access to those who communicate with these decision makers, their sources, and so forth.

> *Example: In the roll-up to the Faulkland Islands war between Argentina and the United Kingdom, the British were deceived into ignoring signs of the upcoming conflict by ignoring the few signs they say, structuring their intelligence mechanisms so as to focus on things the Argentines could control, and believing the Argentine diplomats who were intentionally asserting that negotiations were continuing when they were not. In this example, the Argentines had control over enough of the relevant sensory inputs to the British intelligence operations so that group-think was induced.*

Many studies have shown that optimal group sizes for small tightly knit groups tend to be in the range of 4-7 people. For tactical situations, this is the typical human group size. Whether the group is running a command center, a tank, or a computer attack team, smaller groups tend to lack cohesion and adequate skills, while larger groups become harder to manage in tight situations. It would seem that for tactical purposes, deceptions would be more effective if they could be successful at targeting a group of this size. Groups of this sort also have a tendency to have specialties with cross limited training. For example, in a computer attack group, a different individual will likely be an expert on one operating system as opposed to another. A hardware expert, a fast systems programmer / administrator, appropriate operating system and other domain experts, an information fusion person, and a skilled Internet collector may emerge. No systematic testing of these notions has been done to date but personal experience shows it to be true. Recent work in large group collaboration using information technology to augment normal human capabilities have show limited promise. Experiments will be required to determine whether this is an effective tool in carrying out or defeating deceptions, as well as how such a tool can be exploited so as to deceive its users.

The National Research Council [38] discusses models of human and organizational behavior and how automation has been applied in the modeling of military decision making. This includes a wide range of computer-based modeling systems that have been developed for specific applications and is particularly focused on military and combat situations. Some of these models would appear to be useful in creating effective models for simulation of behavior under deceptions and several of these models are specifically designed to deal with psychological factors. This field is still very new and the progress to date is not adequate to provide coverage for analysis of deceptions, however, the existence of these models and their utility for understanding military organizational situations may be a good foundation for further work in this area.

### Computer Network Deceptions

Computer network deceptions essentially never exist without people involved. The closest thing we see to purely computer to computer deceptions have been feedback mechanisms that induce livelocks or other denial of service impacts. These are the result of misinformation passing between computers.

*Examples include the electrical cascade failures in the U.S. power grid, [58] telephone system cascade failures causing widespread long distance service outages, [59] and inter-system cascades such as power failures bringing down telephone switches required to bring power stations back up. [59]*

But the notion of deception, as we define it, involves intent, and we tend to attribute intent only to human actors at this time. There are, of course, programs that display goal directed behavior, and we will not debate the issue further except to indicate that, to date, this has not been used for the purpose of creating network deceptions without human involvement.

Individuals have used deception on the Internet since before it became the Internet. In the Internet's predecessor, the ARPAnet, there were some rudimentary examples of email forgeries in which email was sent under an alias - typically as a joke. As the Internet formed and become more widespread, these deceptions continued in increasing numbers and with increasing variety. Today, person to person and person to group deception in the Internet is commonplace and very widely practiced as part of the notion of anonymity that has pervaded this media. Some examples of papers in this area include:

*"Gender Swapping on the Internet" [67] was one of the original "you can be anyone on the Internet" descriptions. It dealt with players in MUDs (Multi-User Dungeon), which are multiple-participant virtual reality domains. Players soon realized that they could have multiple online personalities, with different genders, ages, and physical descriptions. The mind behind the keyboard often chooses to stay anonymous, and without violating system rules or criminal laws, it is difficult or impossible for ordinary players to learn many real-world identities.*

*"Cybernetic Fantasies: Extended Selfhood in a Virtual Community" by Mimi Ito, from 1993, [60] is a first-person description of a Multi-User Dungeon (MUD) called Farside, which was developed at a university in England. By 1993 it had 250 players. Some of the people using Farside had characters they maintained in 20 different virtual reality MUDs. Ito discusses previous papers, in which some people went to unusual lengths such as photos of someone else, to convince others of a different physical identity.*

*"Dissertation: A Chatroom Ethnography" by Mark Peace, [61] is a more recent study of Internet Relay Chat (IRC), a very popular form of keyboard to keyboard communication. This is frequently referred to as Computer Mediated Communication (CMC). Describing first-person experiences and observation, Peace believes that many users of IRC do not use false personalities and descriptions most of the time. He also provides evidence that IRC users do use alternate identities.*

*Daniel Chandler writes, "In a 1996 survey in the USA, 91% of homepage authors felt that they presented themselves accurately on their web pages (though only 78% believed that other people presented themselves accurately on their home pages!) [62]*

Criminals have moved to the Internet environment in large numbers and use deception as a fundamental part of their efforts to commit crimes and conceal their identities from law enforcement. While the specific examples are too numerous to list, there are some common threads, among them that the same criminal activities that have historically worked person to person are being carried out over the Internet with great success.

Identity theft is one of the more common deceptions based on attacking computers. In this case, computers are mined for data regarding an individual and that individual's identity is taken over by the criminal who then commits crimes under the assumed name. The innocent victim of the identity theft is often blamed for the crimes until they prove themselves innocent.

One of the most common Internet-based deceptions is an old deception of sending a copier supply bill to a corporate victim. In many cases the internal controls are inadequate to differentiate a legitimate bill from a fraud and the criminal gets paid illegitimately.

Child exploitation is commonly carried out by creating friends under the fiction of being the same age and sex as the victim. Typically a 40 year old pedophile will engage a child and entice them into a meeting outside the home. In some cases there have been resulting kidnappings, rapes, and even murders.

During the cyber conflict between the Palestinian Liberation Organization (PLO) and a group of Israeli citizens that started early in 2001, one PLO cyber terrorist lured an Israeli teenager into a meeting and kidnapped and killed the teen. In this case the deception was the simulation of a new friend made over the Internet:

*The Internet "war" assumed new dimensions here last week, when a 23-year-old Palestinian woman, posing as an American tourist, apparently used the Internet to lure a 16-year-old Israeli boy to the Palestinian Authority areas so he could be murdered.* - Hanan Sher, The Jerusalem Report, 2001/02/10

Larger scale deceptions have also been carried out over the Internet. For example, one of the common methods is to engage a set of 'shills' who make different points toward the same goal in a given forum. While the forum is generally promoted as being even handed and fair, the reality is that anyone who says something negative about a particular product or competitor will get lambasted. This has the social effect of causing distrust of the dissenter and furthering the goals of the product maker. The deception is that the seemingly independent members are really part of the same team, or in some cases, the same person. In another example, a student at a California university made false postings to a financial forum that drove down the price of a stock that the student had invested in derivatives of. The net effect was a multi-million dollar profit for the student and the near collapse of the stock.

The largest scale computer deceptions tend to be the result of computer viruses. Like the mass hysteria of a financial bubble, computer viruses can cause entire networks of computers to act as a rampaging group. It turns out that the most successful viruses today use human behavioral characteristics to induce the operator to foolishly run the virus which, on its own, could not reproduce. They typically send an email with an infected program as an attachment. If the infected program is run it then sends itself in email to other users this user

communicates with, and so forth. The deception is the method that convinces the user to run the infected program. To do this, the program might be given an enticing name, or the message may seem like it was really from a friend asking the user to look at something, or perhaps the program is simply masked so as to simulate a normal document.

> *In one case a computer virus was programmed to silently dial out on the user's phone line to a telephone number that generated revenues to the originator of the virus (a 900 number). This example shows how a computer system can be attacked while the user is completely unaware of the activity.*

These are deceptions that act across computer networks against individuals who are attached to the network. They are targeted at the millions of individuals who might receive them and, through the viral mechanism, distribute the financial burden across all of those individuals. They are a form of a "Salami" attack in which small amounts are taken from many places with large total effect.

## Implications

These examples would tend to lead us to believe that effective defensive deceptions against combinations of humans and computers are easily carried out to substantial effect, and indeed that appears to be true, if the only objective is to fool a casual attacker in the process of breaking into a system from outside or escalating privilege once they have broken in. For other threat profiles, however, such simplistic methods will not likely be successful, and certainly not remain so for long once they are in widespread use. Indeed, all of these deceptions have been oriented only toward being able to observe and defend against attackers in the most direct fashion and not oriented toward the support of larger deceptions such as those required for military applications.

There have been some studies of interactions between people and computers. Some of the typical results include the notions that people tend to believe things the computers tell them, humans interacting through computers tend to level differences of stature, position, and title, that computer systems tend to trust information from other computer systems excessively, that experienced users to interact differently than less experienced ones, the ease of lying about identities and characteristics as demonstrated by numerous stalking cases, and the rapid spread viruses as an interaction between systems with immunity to viruses (by people) for limited time periods. The Tactical Decision Making Under Stress (TADMUS) program is an example of a system designed to mitigate decision errors caused by cognitive overload, which have been documented through research and experimentation. [65]

Sophisticated attack groups tend to be small, on the order of 4-7 people in one room, or operate as a distributed group perhaps as many as 20 people can loosely participate. Most of the most effective groups have apparently been small cells of 4 to 7 people or individuals with loose connections to larger groups. Based on activities seen to date, but without a comprehensive study to back these notions up, less than a hundred such groups appear to be operating overtly today, and perhaps a thousand total groups would be a good estimate based on the total activities detected in openly available information. A more accurate evaluation would require additional research, specifically including the collection of data from substantial sources, evaluation of operator and group characteristics (e.g., times of day, preferred targets, typing characteristics, etc.), and tracking of modus operandi of perpetrators. In order to do this, it would be prudent to start to create sample attack teams and do substantial experiments to understand the internal development of these team, team characteristics over time, team makeup, develop capabilities to detect and differentiate teams, and test out these capabilities in a larger environment. Similarly, the ability to reliably deceive these groups will depend largely on gaining understanding about how they operate.

We believe that large organizations are only deceived by strategic application of deceptions against individuals and small groups. While we have no specific evidence to support this, ultimately is must be true to some extent because groups don't make decisions without individuals making decisions. While there may be different motives for different individuals and groups insanity of a sort may be part of the overall effect, there nevertheless must be specific individuals and small groups that are deceived in order for them to begin to convey the overall message to other groups and individuals. Even in the large-scale perception management campaigns involving massive efforts at propaganda, individual opinions are affected first, small groups follow, and then larger groups become compliant under social pressures and belief mechanisms.

Thus the necessary goal of creating deceptions is to deceive individuals and then small groups that those individuals are part of. This will be true until targets develop far larger scale collaboration capabilities that might allow them to make decisions on a different basis or change the cognitive structures of the group as a whole. This sort of technology is not available at present in a manner that would reduce effectiveness of deception and it may never become available.

Clearly, as deceptions become more complex and the systems they deal with include more and more diverse components, the task of detailing deceptions and their cognitive nature becomes more complex. It appears that there is regular structure in most deceptions involving large numbers of systems of systems because the designers of current widespread attack deceptions have limited resources. In such cases it appears that a relatively small number of factors can serve to model the deceptive elements, however, large scale group deception effects may be far more complex to understand and analyze because of the large number of possible interactions and complex sets of interdependences involved in cascade failures and similar phenomena. If deception technology continues to expand and analytical and implementation capabilities become more substantial, there is a tremendous potential for highly complex deceptions wherein many different systems are involved in highly complex and irregular interactions. In such an environment, manual analysis will not be capable of dealing with the issues and automation will be required in order to both design the deceptions and counter them.

## Experiments and the Need for an Experimental Basis

One of the more difficult things to accomplish in this area is meaningful experiments. While a few authors have

published experimental results in information protection, far fewer have attempted to use meaningful social science methodologies in these experiments or to provide enough testing to understand real situations. This may be because of the difficulty and high cost of each such experiment and the lack of funding and motivation for such efforts. We have identified this as a critical need for future work in this area.

If one thing is clear from our efforts it is the fact that too few experiments have been done to understand how deception works in defense of computer systems and, more generally, too few controlled experiments have been done to understand the computer attack and defense processes and to characterize them. Without a better empirical basis, it will be hard to make scientific conclusions about such efforts.

While anecdotal data can be used to produce many interesting statistics, the scientific utility of those statistics is very limited because they tend to reflect only those examples that people thought worthy of calling out. We get only *"lies, damned lies, and statistics."*

**Experiments to Date**

From the time of the first published results on honeypots, the total number of published experiments performed in this area appear to be very limited. While there have been hundreds of published experiments by scores of authors in the area of human deception, articles on computer deception experiments can be counted on one hand.

Cohen provided a few examples of real world effects of deception, [6] but performed no scientific studies of the effects of deception on test subjects. While he did provide a mathematical analysis of the statistics of deception in a networked environment, there was no empirical data to confirm or refute these results. [7]

The HoneyNet Project [43] is a substantial effort aimed at placing deception system in the open environment for detection and tracking of attack techniques. As such, they have been largely effective at luring attackers. These lures are real systems placed on the Internet with the purpose of being attacked so that attack methods can be tracked and assessed. As deceptions, the only thing deceptive about them is that they are being watched more closely than would otherwise be apparent and known faults are intentionally not being fixed to allow attacks to proceed. These are highly effective at allowing attackers to enter because they are extremely high fidelity, but only for the purpose they are intended to provide. They do not, for example, include any user behaviors or content of interest. They are quite effective at creating sites that can be exploited for attack of other sites. For all of the potential benefit, however, the HoneyNet project has not performed any controlled experiments to understand the issues of deception effectiveness.

Red teaming (i.e., finding vulnerabilities at the request of defenders) [64] has been performed by many groups for quite some time. The advantage of red teaming is that it provides a relatively realistic example of an attempted attack. The disadvantage is that it tends to be somewhat artificial and reflective of only a single run at the problem. Real systems get attacked over time by a wide range of attackers with different skill sets and approaches. While many red teaming exercises have been performed, these tend not to provide the scientific data desired in the area of defensive deceptions because they have not historically been oriented toward this sort of defense.

Similarly, war games played out by armed services tend to ignore issues of information system attacks because the exercises are quite expensive and by successfully attacking information systems that comprise command and control capabilities, many of the other purposes of these war games are defeated. While many recognize that the need to realistically portray effects is important, we could say the same thing about nuclear weapons, but that doesn't justify dropping them on our forces for the practice value.

The most definitive experiments to date that we were able to find on the effectiveness of low-quality computer deceptions against high quality computer assisted human attackers were performed by RAND. [24] Their experiments with fairly generic deceptions operated against high quality intelligence agency attackers demonstrated substantial effectiveness for short periods of time. This implies that under certain conditions (i.e., short time frames, high tension, no predisposition to consider deceptions, etc.) these deceptions may be effective.

The total number of controlled experiments to date involving deception in computer networks appear to be less than 20, and the number involving the use of deceptions for defense are limited to the 10 or so from the RAND study. Clearly this is not enough to gain much in the way of knowledge and, just as clearly, many more experiments are required in order to gain a sound understanding of the issues underlying deception for defense.

**Experiments We Believe Are Needed At This Time**

In this study, a large set of parameters of interest have been identified and several hypotheses put forth. We have some anecdotal data at some level of detail, but we don't have a set of scientific data to provide useful metrics for producing scientific results. In order for our models to be effective in producing increased surety in a predictive sense we need to have more accurate information.

The clear solution to this dilemma is the creation of a set of experiments in which we use social science methodologies to create, run, and evaluate a substantial set of parameters that provide us with better understanding and specific metrics and accuracy results in this area. In order for this to be effective, we must not only create defenses, but also come to understand how attackers work and think. For this reason, we will need to create red teaming experiments in which we study both the attackers and the effects of defenses on the attackers. In addition, in order to isolate the effects of deception, we need to create control groups, and experiments with double blinded data collection.

---

# Analysis and Design of Deceptions

A good model should be able to explain, but a good scientific model should be able to predict and a good model for our purposes should help us design as well. At a minimum, the ability to predict leads to the ability to design by random variation and selective survival with the survival evaluation being made based on prediction. In most cases, it is a lot more efficient to have the ability to create design rules that are reflective of some underlying structure.

Any model we build that is to have utility must be computationally reasonable relative to the task at hand. Far more computation is likely to be available for a large-scale strategic deception than for a momentary tactical deception, so it would be nice to have a model that scales well in this sense. Computational power is increasing with time, but not at such a rate that we will ever be able to completely ignore computational complexity in problems such as this.

A fundamental design problem in deception lies in the fact that deceptions are generally thought of in terms of presenting a desired story to the target, while the available techniques are based on what has been found to work. In other words, there is a mismatch between available deception techniques and technologies and objectives.

## A Language for Analysis and Design of Deceptions

Rather than focus on what we wish to do, our approach is to focus on what we can do and build up 'deception programs' from there. In essence, our framework starts with a programming language for human deception by finding a set of existing primitives and creating a syntax and semantics for applying these primitives to targets. We can then associate metrics with the elements of the programming language and analyze or create deceptions that optimize against those metrics.

The framework for human deception then has three parts:

- **A set of primitive techniques:** The set of primitive techniques is extensive and is described hierarchically based on the model shown above, with each technique associated with one or more of Observables, Actions, Assessments, Capabilities, Expectations, and Intent and causing an effect on the situation depicted by the model.
- **Properties of those techniques:** Properties of techniques are multi-dimensional and include all of the properties discussed in this report. This includes, but is not limited to, resources consumed, effect on focus of attention, concealment, simulation, memory requirements and impacts, novelty to target, certainty of effect, extent of effect, timeliness of effect, duration of effect, security requirements, target system resource limits, deceiver system resource limits, the effects of small changes, organizational structure, knowledge, and constraints, target knowledge requirements, dependency on predisposition, extent of change in target mind set, feedback potential and availability, legality, unintended consequences, the limits of modeling, counterdeception, recursive properties, and the story to be told. These are the same properties of deception discussed under "The Nature Of Deception" earlier.
- **A syntax and semantics for applying and optimizing the properties:** This is a language that has not yet been developed for describing, designing, and analyzing deceptions. It is hoped that this language and the underlying database and simulation mechanism will be developed in subsequent efforts.

The astute reader will recognize this as the basis for a computer language, but it has some differences from most other languages, most fundamentally in that it is probabilistic in nature. While most programming languages guarantee that when you combine two operators together in a sequence you get the effect of the first followed by the effect of the second, in the language of deception, a sequence of operators produces a set of probabilistic changes in perceptions of all parties across the multi-dimensional space of the properties of deception. It will likely be effective to "program" in terms of desired changes in deception properties and allow the computer to "compile" those desired changes into possible sequences of operators. The programming begins with a 'firing table' of some sort that looks something like the following table, but with many more columns filled in and many more details under each of the rows. Partial entries are provided for technique 1 which, for this example, we will choose as 'audit suppression' by packet flooding of audit mechanisms using a distributed set of previously targeted intermediaries.

| Deception Property | Technique 1 | ... | Technique n |
|---|---|---|---|
| name | Audit Suppression | | |
| general concept | packet flooding of audit mechanisms | | |
| means | using a distributed set of intermediaries | | |
| target type | computer | | |
| resources consumed | reveals intermediaries which will be disabled with time | | |
| effect on focus of attention | induces focus on this attack | | |
| concealment | conceals other actions from target audit and analysis | | |
| simulation | n/a | | |
| memory requirements and impacts | overruns target memory capacity | | |
| novelty to target | none - they have seen similar things before | | |
| certainty of effect | 80% effective if intel is right | | |

| | | | |
|---|---|---|---|
| extent of effect | reduces audits by 90% if effective | | |
| timeliness of effect | takes 30 seconds to start | | |
| duration of effect | until ended or intermediaries are disabled | | |
| security requirements | must conceal launch points and intermediaries | | |
| target system resource limits | memory capacity, disk storage, CPU time | | |
| deceiver system resource limits | number of intermediaries for attack, pre-positioned assets lost with attack | | |
| the effects of small changes | nonlinear effect on target with break point at effectiveness threshold | | |
| organizational structure and constraints | Going after known main audit server which will impact whole organization audits | | |
| target knowledge | OS type and release | | |
| dependency on predisposition | Must be proper OS type and release to work | | |
| extent of change in target mind set | Large change - it will interrupt them - they will know they are being attacked | | |
| feedback potential and availability | Feedback apparent in response behavior observed against intermediaries and in other fora | | |
| legality | Illegal except at high intensity conflict - possible act of war | | |
| unintended consequences | Impacts other network elements, may interrupt other information operations, may result in increased target security | | |
| the limits of modeling | Unable to model overall network effects | | |
| counterdeception | If feedback known or attack anticipated, easy to deceive attacker | | |
| recursive properties | only through counter deception | | |
| possible deception story | We are concealing something - they know this - but they don't know what | | |

Considering that the total number of techniques is likely to be on the order of several hundred and the vast majority of these techniques have not not been experimentally studied, the level of effort required to build such a table and make it useful will be considerable.

### Attacker Strategies and Expectations

For a moment, we will pause from the general issue of deception and examine more closely the situation of an attacker attempting to exploit a defender through information system attack. In this case, there is a commonly used attack methodology that subsumes other common methodologies and there are only three known successful attack strategies identified by simulation and verified against empirical data. We start with some background.

The pathogenesis of diseases has been used to model the process of breaking onto computers and it offers an interesting perspective. [63] In this view, the characteristics of an attack are given in terms of the survival of the attack method.

Table 7.1 from "Emerging Viruses"

| | |
|---|---|
| 1 | Stability in environment |
| 2 | Entry into host - portal of entry |
| 3 | Localization in cells near portal of entry |
| 4 | Primary replication |
| 5 | Non-specific immune response |
| 6 | Spread from primary site (blood, Nerves) |
| 7 | Cells and tissue tropism |
| 8 | Secondary replication |
| 9 | Antibody and cellular immune response |
| 10 | Release from host |

"Pathogenesis of Computer Viruses"

| | |
|---|---|
| 1 | Stability in environment |
| 2 | Entry into host - portal of entry |
| 3 | Localization in software near portal of entry |
| 4 | Primary replication |
| 5 | Non-specific immune response |
| 6 | Spread from primary site (disk, comms) |
| 7 | Program and data tropism |
| 8 | Secondary replication |
| 9 | Human and program immune response |
| 10 | Release from host |

"Pathogenesis of Manual Attacks"

| | |
|---|---|
| 1 | Stability in environment |
| 2 | Entry into host - portal of entry |
| 3 | Localization near portal of entry |
| 4 | Primary modifications |
| 5 | Non-specific immune response |
| 6 | Spread from primary site (privilege expansion) |
| 7 | Program and data tropism (hiding) |
| 8 | Secondary replication |
| 9 | Human and program immune response |
| 10 | Release from host (spread on) |

This particular perspective on attack as a biological process ignores one important facet of the problem, and that is the preparation process for an intentional and directed attack. In the case of most computer viruses, targeting is not an issue. In the case of an intelligent attacker, there is generally a set of capabilities and an intent behind the attack. Furthermore, survival (stability in the environment) would lead us to the conclusion that a successful attacker who does not wish to be traced back to their origin will use an intelligence process including personal risk reduction as part of their overall approach to attack. This in turn leads to an intelligence process that precedes the actual attack.

The typical attack methodology consists of:

(1) intelligence gathering, securing attack infrastructure, tool development, and other preparations
(2) system entry (beyond default remote access),
(3) privilege expansion,
(4) subversion, typically involving planting capabilities and verifying over time, and
(5) exploitation.

There are loops from higher numbers to lower numbers so that, for example, privilege expansion can lead back to intelligence and system entry or forward to subversion, and so forth. In addition, attackers have expectations throughout this process that adapt based on what has been seen before this attack and within this attack. Clean up, observation of effects, and analysis of feedback for improvement are also used throughout the attack process.

Extensive simulation has been done to understand the characteristics of successful attacks and defenses. [5] Among the major results of this study were a set of successful strategies for attacking computer systems. It is particularly interesting that these strategies are similar to classic military strategies because the simulation methods used were not designed from a strategic viewpoint, but were based solely on the mechanisms in use and the times, detection, reaction, and other characteristics associated with the mechanisms themselves. Thus the strategic information that fell out of this study was not biased by its design but rather emerged as a result of the metrics associated with different techniques. The successful attack strategies identified by this study included:

(1) speed,
(2) stealth, and
(3) overwhelming force.

Slow, loud attacks tend to be detected and reacted to fairly easily. A successful attacker can use combinations of these in different parts of an attack. For example, speed can be used for a network scan, stealth for system entry, speed for privilege expansion and planting of capabilities, stealth for verifying capabilities over time, and overwhelming force for exploitation. This is a typical pattern today.

Substantial red teaming and security audit experience has led to some speculations that follow the general notions of previous work on individual deception. It seems clear from experience that people who use computers in attacks:

(1) tend to trust what the computers tell them unless it is far outside normal expectations,
(2) use the computer to automate manual processes and not to augment human reasoning, and
(3) tend to have expectations based on prior experience with their tools and targets.

If this turns out to be true, it has substantial implications for both attack and defense. Experiments should be undertaken to examine these assertions as well as to study the combined deception properties of small groups of people working with computers in attacking other systems. Unfortunately, current data is not adequate to thoroughly understand these issues. There may be other strategies developed by attackers, other attack processes undertaken, and other tendencies that have more influence on the process. We will not know this until extensive experimentation is done in this area.

## Defender Strategies and Expectations

From the deceptive defender's perspective, there also seem to be a limited set of strategies.

- **Computer Only:** If the computer is being used for a fully automated attack, analysis of the attack tool or relatively simply automated response mechanisms are highly effective at maintaining the computer's expectations, dazzling the computer to induce unanticipated processing and results, feeding false information to the computer, or in some cases, causing the computer to crash. We have been able to easily induce or suppress signal returns to an attacking computer and have them seen as completely credible almost no matter how ridiculous they are. Whether this will continue and to what extent it will continue in the presence of a sophisticated hostile environment remain to be seen.

- **People Only:** Manual attack is very inefficient so it is rarely used except in cases where very specific targets are involved. Because humans do tend to see what they expect to see, it is relatively easy to create high fidelity deceptions by redirecting traffic to a honey pot or other such system. Indeed, this transition can even be made fairly early in an attack without most human attackers noticing it. In this case there are three things we might want to do:

    - (1) maintain the attackers expectations to consume their time and effort,

    - (2) slowly change their expectations to our advantage at a rate that is not noticeable by typical humans (e.g., slow the computer's response minute by minute till it is very slow and the attacker is wasting lots of time and resources), and

    - (3) create cognitive dissonance to force them to think more deeply about what is going on, wonder if they have been detected, and induce confusion in the attacker.

- **People With Poorly Integrated Computers:** This is the dominant form of efficient widespread attack today. In this form, people use automated tools combined with short bursts of human activity to carry out attacks.

  The intelligence process is almost entirely done by scanning tools which (1) can be easily deceived and (2) tend to be believed. Such deceptions will only be disbelieved if inconsistencies arise between tools, in

which case the tools will initially be suspected.

System entry is either automated with the intelligence capability or automated at a later time when the attacker notices that an intelligence sweep has indicated a potential vulnerability. Results of these tools will be believed unless they are incongruous with normal expectations.

Privilege expansion is either fully automated or has a slight manual component to it. It typically involves the loading of a toolkit for the job followed by compilation and/or execution. This typically involves minimal manual effort. Results of this effort are believed unless they are incongruous with normal expectations.

Planting capabilities is typically nearly automated or fully automated. Returning to verify over time is typically automated with time frames substantially larger than attack times. This will typically involve minimal manual effort. Results of this effort will be believed unless they are incongruous with normal expectations.

Exploitation is typically done under one-shot or active control. A single packet may trigger a typical exploit, or in some cases the exploit is automatic and ongoing over an extended period of time. This depends on whether speed, stealth, or force is desired in the exploitation phase. This causes observables that can be validated by the attacker. If the observables are not present it might generate deeper investigation by the attacker. If there are plausible explanations that can be discovered by the attacker they will likely be believed.

- **People With Well Integrated Computers:** This has not been observed to date. People are not typically augmenting their intelligence but rather automating tasks with their computers.

As in the case with attacker strategies, few experiments have been undertaken to understand these issues in detail, but preliminary experiments appear to confirm these notions.

### Planning Deceptions

Several authors have written simplistic analyses and provided rules of thumb for deception planning. There are also some notions about planning deceptions under the present model using the notions of low, middle, and high level cognition to differentiate actions and create our own rules of thumb with regard to our cognitive model. But while notions are fine for contemplation, scientific understanding in this area requires an experimental basis.

According to [10] a 5-step process is used for military deception. (1) Situation analysis determines the current and projected enemy and friendly situation, develops target analysis, and anticipates a desired situation. (2) Deception objectives are formed by desired enemy action or non-action as it relates to the desired situation and friendly force objectives. (3) Desired [target] perceptions are developed as a means to generating enemy action or inaction based on what the enemy now perceives and would have to perceive in order to act or fail to act - as desired. (4) The information to be conveyed to or kept from the enemy is planned as a story or sequence, including the development and analysis of options. (5) A deception plan is created to convey the deception story to the enemy.

These steps are carried out by a combination of commander and command staff as an embedded part of military planning. Because of the nature of military operations, capabilities that are currently available and which have been used in training exercises and actual combat are selected for deceptions. This drives the need to create deception capabilities that are flexible enough to support the commander's needs for effective use of deceptions in a combat situation. From a standpoint of information technology deceptions, this would imply that, for example, a deceptive feint or movement of forces behind smoke screens with sonic simulations of movement should be supported by simulated information operations that would normally support such action and concealed information operations that would support the action being covered by the feint.

Deception maxims are provided to enhance planner understanding of the tools available and what is likely to work: [10]

**Magruder's principles - the exploitation of perceptions:** It is easier to maintain an existing belief than to change it or create a new one.

**Limitations of human information processing:** The law of small numbers (once you see something twice it is taken as a maxim), and susceptibility to conditioning (the cumulative effect of small changes). These are also identified and described in greater detail in Gilovich [14].

**Cry-Wolf:** This is a variant on susceptibility to conditioning in that, after a seeming threat appears again and again to be innocuous, it tends to be ignored and can be used to cover real threats.

**Jones' Dilemma:** Deception is harder when there are more information channels available to the target. On the other hand, the greater the number of 'controlled channels', the better it is for the deception.

**A choice among deception types:** In "A-type" deception, ambiguity is introduced to reduce the certainty of decisions or increase the number of available options. In "M-type" deception, misdirection is introduced to increase the victim's certainty that what they are looking for is their desired (deceptive) item.

**Axelrod's contribution - the husbanding of assets:** Some deceptions are too important to reveal through their use, but there is a tendency to over protect them and thus lose them by lack of application. Some deception assets become useless once revealed through use or overuse. In cases where strategic goals are greater than tactical needs, select deceptions should be held in reserve until they can be used with greatest effect.

**A sequencing rule:** Sequence deceptions so that the deception story is portrayed as real for as long as possible. The most clear indicators of deception should be held till the last possible moment. Similarly, riskier elements of a deception (in terms of the potential for harm if the deception is discovered) should be done later rather than earlier so that they may be called off if the deception is found to be a failure.

**The importance of feedback:** A scheme to ensure accurate feedback increases the chance of success in deception.

**The Monkey's Paw:** Deceptions may create subtle and undesirable side effects. Planners should be sensitive to such possibilities and, where prudent, take steps to minimize these effects.

**Care in the designed and planned placement of deceptive material:** Great care should be used in deceptions that leak notional information to targets. Apparent windfalls are subjected to close scrutiny and often disbelieved. Genuine leaks often occur under circumstances thought improbable.

Deception failures are typically associated with (1) detection by the target and (2) inadequate design or implementation. Many examples of this are given. [10]

As a doctrinal matter, Battlefield deception involves the integration of intelligence support, integration and synchronization, and operations security. [10]

**Intelligence Support:** Battlefield deceptions rely heavily on timely and accurate intelligence about the enemy. To make certain that deceptions are effective, we need to know (1) how the target's decision and intelligence cycles work, (2) what type of deceptive information they are likely to accept, (3) what source they rely on to get their intelligence, (4) what they need to confirm their information, and (5) what latitude they have in changing their operations. This requires both advanced information for planning and real-time information during operations.

**Integration and Synchronization:** Once we know the deception plan we need to synchronize it with the true combat operations for effect. History has shown that for the greatest chance of success, we need to have plans that are: (1) flexible, (2) doctrinally consistent with normal operations, (3) credible as to the current situation, and (4) simple enough to not get confused during the heat of battle. Battlefield deceptions almost always involve the commitment of real forces, assets, and personnel.

**Operations Security:** OPSEC is the defensive side of intelligence. In order for a deception to be effective, we must be able to deny access to the deceptive nature of the effort while also denying access to our real intentions. Real intentions must be concealed, manipulated, distorted, and falsified though OPSEC.

> *"OPSEC is not an administrative security program. OPSEC is used to influence enemy decisions by concealing specific, operationally significant information from his intelligence collection assets and decision processes. OPSEC is a concealment aspect for all deceptions, affecting both the plan and how it is executed"* [10]

In the DoD context, it must be assumed that any enemy is well versed in DoD doctrine. This means that anything too far from normal operations will be suspected of being a deception even if it is not. This points to the need to vary normal operations, keep deceptions within the bounds of normal operations, and exploit enemy misconceptions about doctrine. Successful deceptions are planned from the perspective of the targets.

The DoD has defined a set of factors in deceptions that should be seriously considered in planning [10]. It is noteworthy that these rules are clearly applicable to situations with limited time frames and specific objectives and, as such, may not apply to situations in information protection where long-term protection or protection against nebulous threats are desired.

**Policy:** Deception is never an end in itself. It must support a mission.

**Objective:** A specific, realistic, clearly defined objective is an absolute necessity. All deception actions must contribute to the accomplishment of that objective.

**Planning:** Deception should be addressed in the commander's initial guidance to staff and the staff should be engaged in integrated deception and operations planning.

**Coordination:** The deception plan must be in close coordination with the operations plan.

**Timing:** Sufficient time must be allowed to: (1) complete the deception plan in an orderly manner, (2) effect necessary coordination, (3) promulgate tasks to involved units, (4) present the deception to the enemy decision-maker through their intelligence system, (5) permit the enemy decision maker to react in the desired manner, including the time required to pursue the desired course of action.

**Security:** Stringent security is mandatory. OPSEC is vital but must not prevent planning, coordination, and timing from working properly.

**Realism:** It must look realistic.

**Flexibility:** The ability to react rapidly to changes in the situation and to modify deceptive action is mandatory.

**Intelligence:** Deception must be based on the best estimates of enemy intelligence collection and decision-making processes and likely intentions and reactions.

**Enemy Capabilities:** The enemy commander must be able to execute the desired action.

**Friendly Force Capabilities:** Capabilities of friendly forces in the deception must match enemy estimates of capabilities and the deception must be carried out without unacceptable degradation in friendly capabilities.

**Forces and Personnel:** Real forces and personnel required to implement the deception plan must be provided. Notional forces must be realistically portrayed.

**Means:** Deception must be portrayed through all feasible and available means.

**Supervision:** Planning and execution must be continuously supervised by the deception leader. Actions must be coordinated with the objective and implemented at the proper time.

**Liaison:** Constant liaison must be maintained with other affected elements to assure that maximum effect is attained.

**Feedback:** A reliable method of feedback must exist to gage enemy reaction.

Deception of humans and automated systems involves interactions with their sensory capabilities. [10] For people, this includes (1) visual (e.g., dummies and decoys, camouflage, smoke, people and things, and false vs. real sightings), (2) Olfactory (e.g., projection of odors associated with machines and people in their normal activities at that scale including toilet smells, cooking smells, oil and gas smells, and so forth), (3) sonic (e.g., directed against sounding gear and the human ear blended with real sounds from logical places and coordinated to meet the things being simulated at the right places and times) (4) electronic (i.e., manipulative electronic deception, simulative electronic deception, and imitative electronic deception).

Resources (e.g., time, devices, personnel, equipment, materiel) are always a consideration in deceptions as are the need to hide the real and portray the false. Specific techniques include (1) feints, (2) demonstrations, (3) ruses, (4) displays, (5) simulations, (6) disguises, and (7) portrayals. [10]

## A Different View of Deception Planning Based on the Model from this Study

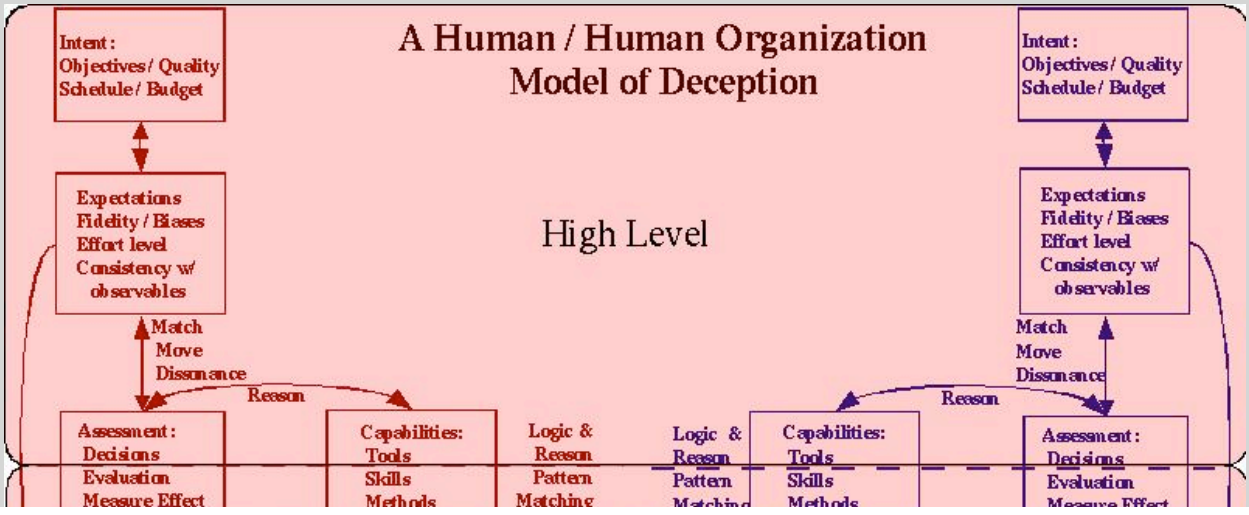
A typical deception is carried out by the creation and invocation of a deception plan. Such a plan is normally based on some set of reasonably attainable goals and time frames, some understanding of target characteristics, and some set of resources which are made available for use. It is the deception planner's objective to attain the goals with the provided resources within the proper time frames. In defending information systems through deception our objective is to deceive human attackers and defeat the purposes of the tools these humans develop to aid them in their attacks. For this reason, a framework for human deception is vital to such an undertaking.

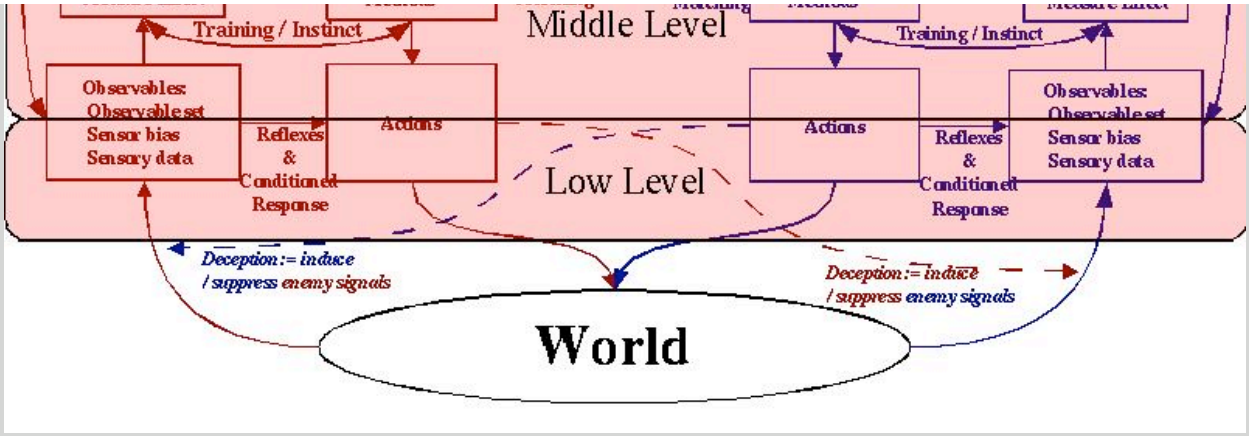
All deception planning starts with the objective. It may work its way back toward the creation of conditions that will achieve that objective or use that objective to 'prune' the search space of possible deception methods. While it is tempting for designers to come up with new deception technologies and turn them into capabilities; (1) Without a clear understanding of the class of deceptions of interest, it will not be clear what capabilities would be desirable; and (2) Without a clear understanding of the objectives of the specific deception, it will not be clear how those capabilities should be used. If human deception is the objective, we can begin the planning process with a model of human cognition and its susceptibility to deception.

The skilled deception planner will start by considering the current and desired states of mind of the deception target in an attempt to create a scenario that will either change or retain the target's state of mind by using capabilities at hand. State of mind is generally only available when (1) we can read secret communications, (2) we have insider access, or (3) we are able to derive state of mind from observable outward behavior. Understanding the limits of controllable and uncontrollable target observables and the limits of intelligence required to assure that the target is getting and properly acting (or not acting) on the information provided to them is a very hard problem.

**Deception Levels**

In the model depicted above and characterized by the diagram below, three levels can be differentiated for clearer understanding and grouping of available techniques. They are characterized here by mechanism, predictability, and analyzability:






**Human Deception Levels**

| Level | Mechanism | Predictability | Analysis | Summary |
|---|---|---|---|---|
| Low-level | Low-level deceptions operate at the lower portions of the areas labeled observables and actions. They are designed to cause the target of the deception to be physically unable to observe signals or to cause the target to selectively observe signals. | Low-level deceptions are highly predictable based on human physiology and known reflexes. | Low-level deceptions can be analyzed and very clearly characterized through experiments that yield numerical results in terms of parameters such as detection thresholds, response times, recovery times, edge detection thresholds, and so forth. | Except in cases where the target has sustained physiological damage, these deceptions operate very reliably and predictably. The time frames for these deceptions tend to be in the range of milliseconds to seconds and they can be repeated reliably for ongoing effect. |
| Mid-Level | Mid-level deceptions operate in the upper part of the areas labeled Observables and Actions and in the lower part of the areas marked Assessment and Capabilities. They are generally designed to either: (1) cause the target to invoke trained or pattern matching based responses and avoid deep thought that might induce unfavorable (to us) actions; or (2) induce the target to use high level cognitive functions, thus avoiding faster pattern matching responses. | Mid-level deceptions are usually predictable but are affected by a number of factors that are rather complex, including but not limited to socialization processes and characteristics of the society in which the person was brought up and lives. | Analysis is based on a substantial body of literature. Experiments required for acquiring this knowledge are complex and of limited reliability. There are a relatively small number of highly predictable behaviors. These relatively small number of behaviors are common and are invoked under predictable circumstances. | Many mid-level deceptions can be induced with reasonable certainty through known mechanisms and will produce predictable results if applied with proper cautions, skills, and feedback. Some require social background information on the subject for high surety of results. The time frame for these deceptions tends to be seconds to hours with lasting residual effects that can last for days to weeks. |
| High-level | High-level deceptions operate from the upper half of the areas labeled Assessment and Capabilities to the top of the chart. They are designed to cause the subject to make a series of reasoned decisions by creating sequences of circumstances that move the individual to a desired mental state. | High-level deceptions are reasonably controlled if adequate feedback is provided, but they are far less certain to work than lower level deceptions. The creation and alteration of expectations has been studied in detail and it is clearly a high skills activity where greater skill tends to prevail. | High-level deception requires a high level of feedback when used against a skilled adversary and less feedback under mismatch conditions. There is a substantial body of supporting literature in this area but it is not adequate to lead to purely analytical methods for judging deceptions. | High level deception is a high skills game. A skilled and properly equipped team has a reasonable chance of carrying out such deceptions if adequate resources are applied and adequate feedback is available. These sorts of deceptions tend to operate over a time frame of hours to years and in some cases have unlimited residual effect. |

**Deception Guidelines**

This structuring leads to general guidelines for effective human deception. In essence, they indicate the situations in which different levels of deception should be used and rules of thumb for their use.

| | |
|---|---|
| Low-Level | - Higher certainty can be achieved at lower levels of perception.<br>- Deception should be carried out at as low a level as feasible.<br>- If items are to be hidden and can be made invisible to the target's sensors, this is preferred.<br>- If a perfect simulation of a desired false situation can be created for the enemy sensors, this is preferred.<br>- Do not invoke unnecessary mid-level responses and pattern matching<br>- Try to avoid patterns that will create dissonance or uncertainty that would lead to deeper inspection. |
| Mid-Level | - If a low-level deception will not work, a mid-level deception must be used.<br>- Time pressure and high stress combine to keep targets at mid-level cognitive activities.<br>- Activities within normal situational expectations tend to be handled by mid-level decision processes.<br>- Training tends to generate mid-level decision processes.<br>- Mid-level deceptions require feedback for increased assurance.<br>- Remain within the envelope of high-level expectations to avoid high level analysis.<br>- Exceed the envelope of high-level expectations to trigger high level analysis. |
| High-Level | - If the target cannot be forced to make a mid-level decision in your favor, a high-level deception must be used.<br>- It is easiest to reinforce existing predispositions.<br>- To alter predisposition, high-level deception is required.<br>- Movement from predisposition to new disposition should be made at a pace that does not create dissonance.<br>- If target confusion is desired, information should be changed at a pace that creates dissonance.<br>- In high-level deceptions, target expectations must be considered at all times.<br>- High-level deceptions require the most feedback to measure effect and adapt to changing situations. |

Just as Sun Tzu created guidelines for deception, there are many modern pieces of advice that probably work pretty well in many situations. And like Sun Tzu, these are based on experience in the form of anecdotal data. As someone once said: *The plural of anecdote is statistics.*

### Deception Algorithms

As more and more of these sorts of rules of thumb based on experience are combined with empirical data from experiments, it is within the realm of plausibility to create more explicit algorithms for decision planning and evaluation. Here is an example of the codification of one such algorithm. It deals with the issue of sequencing of deceptions with different associated risks identified above.

Let's assume you have two deceptions, A (low risk) and B (high risk). Then, if the situation is such that the success of either means the mission is accomplished, the success of both simply raises the quality of the success (e.g. it costs less), and the discovery of either by the target will increase the risk that the other will also fail, then you should do A first to assure success. If A succeeds you then do B to improve the already successful result. If A fails, you either do something else or do B out of desperation. On the other hand, if the situation is such that the success of both A and B are required to accomplish the mission and if the discovery of either by the target early in execution will result in substantially less harm than discovery later in execution, then you should do B first so that losses are reduced if, as is more likely, B is detected. If B succeeds, you then do A. Here this is codified into a form more amenable to computer analysis and automation:

```
GIVEN: Deception A (low risk) and Deception B (high risk).
IF [A Succeeds] OR [B Succeeds] IMPLIES [Mission Accomplished, Good Quality/Sched/Cost]
AND [A Succeeds] AND [B Succeeds] IMPLIES [Mission Accomplished, Best Quality/Sched/Cost]
AND [A Discovered] OR [B Discovered ] IMPLIES [A (higher risk) AND B (higher risk)]
THEN    DO B [comment: Do high-risk B first to insure minimal loss in case of detection]
        IF [B Succeeds] DO A (Late) [comment: Do low-risk A second to improve outcome]
                        ELSE DO Out #1 [comment: Do higher-risk A because you're desperate.]
                        OR ELSE DO Out #n [comment: Do something else instead.]

IF [A Succeeds] OR [B Succeeds] IMPLIES [Mission Accomplished, Good Quality/Sched/Cost]
AND [A Detected] OR [B Detected] IMPLIES [Mission Fails]
AND [A Discovered Early] OR [B Discovered Early] IMPLIES [Mission Fails somewhat]
AND [A Discovered Late] OR [B Discovered Late] IMPLIES [Mission Fails severely]
THEN    DO B [comment: Do high-risk B first to test and advance situation]
        IF [B Early Succeeds] DO A (Late) [comment: Do low-risk A second for max chance of success]
                IF [A Late Succeeds (likely)] THEN MISSION SUCCEEDS.
                ELSE [A Late Fails (unlikely)] THEN MISSION FAILS/in real trouble.
        ELSE [B Early Fails] [Early Failure]
                DO Out #1 [comment: Do successful retreat as pre-planned.]
                OR DO Out #m [comment: Do another pre-planned contingency instead.]
```

We clearly have a long way to go in codifying all of the aspects of deception and deception sequencing in such a form, but just as clearly, there is a path to the development of rules and rule-based analysis and generation methods for building deceptions that have effect and reduce or minimize risk, or perhaps optimize against a wide range of parameters in many situations. The next reasonable step down this line would be the creation of a set of analytical rules that could be codified and experimental support for establishing the metrics associated

with these rules. A game theoretical approach might be one of the ways to go about analyzing these types of systems.

---

## Summary, Conclusions, and Further Work

This paper has summarized a great deal of information on the history of deception in general and the historical, current, and emerging use of deception for information protection in specific. While there is a great deal to know about how deception has been used in the past, it seems quite clear that there will be far more to know about deception in the future. The information protection field has an increasingly pressing need for innovations that change the balance between attack and defense. It is clear from what we already know that deception techniques have the demonstrated ability to increase attacker workload and reduce attacker effectiveness while decreasing defender effort required for detection and providing substantial increases in defender understanding of attacker capabilities and intent.

Modern defensive computer deceptions are in their infancy, but they are moderately effective, even in this simplistic state. The necessary breakthrough that will turn these basic deception techniques and technologies into viable long-term defenses is the linkage of social sciences research with technical development. In specifics, we need to measure the effects and known characteristics of deceptions on the systems comprising of people and their information technology to create, understand, and exploit the psychological and physiological bases for the effectiveness of deceptions. The empirical basis for effective deception in other arenas is simply not available in the information protection arena today, and in order to attain it, there is a crying need for extensive experimentation in this arena.

To a large extent this work has been facilitated by the extensive literature on human and animal deception that has been generated over a long period of time. In recent years, the experimental evidence has accumulated to the point where there is a certain degree of general agreement in the part of the scientific community that studies deception about many of the underlying mechanisms, the character of deception, the issues in deception detection, and the facets that require further research. These same results and experimental techniques need to be applied to deception for information protection if we are to become designers of effective and reliable deceptions.

The most critical work that must be done in order to make progress is the systematic study of the effectiveness of deception techniques against combined systems with people and computers. This goes hand in hand with experiments on how to counter deceptions and the theoretical and practical limits of deceptions and deception technologies. In addition, codification of prior rules of engagement, the creation of simulation systems and expert systems for analysis of deceptions sequences, and a wide range of related work would clearly be beneficial as a means to apply the results of experiments once empirical results are available.

---

## References


[1] OODA Loop Citation - Boyd Cycle

[2] David Lambert, "A Cognitive Model for Exposition of Human Deception and Counter-deception" (NOSC Technical Report 1076 - October, 1987).

[3] Fred Cohen, "The Structure of Intrusion and Intrusion Detection", May 16, 2000, http://all.net/ (InfoSec Baseline Studies)

[4] Fred Cohen, "A Theory of Strategic Games with Uncommon Objectives"

[5] Fred Cohen, "Simulating Cyber Attacks, Defenses, and Consequences", IFIP TC-11, Computers and Security, 1999.

[6] F. Cohen, "A Note on the Role of Deception in Information Protection", Computers and Security 1999.

[7] F. Cohen, "A Mathematical Structure of Simple Defensive Network Deceptions", 1999, http://all.net (InfoSec Baseline Studies).

[8] James F. Dunnigan and Albert A. Nofi, "Victory and Deceipt: Dirty Tricks at War", William Morrow and Co., New York, NY, 1995.

[9] F. Cohen, "Managing Network Security: What does it do behind your back?", July, 2000, Network Security Management Magazine.

[10] Field Manual 90-02: Battlefield Deception, 1998.

[11] Bart Whaley, "Stratagem: Deception and Surprise in War", Cambridge: MIT Center for International Studies. 1969

[12] Chuck Whitlock, "Scam School", MacMillan, 1997.

[13] Bob Fellows, "Easily Fooled", Mind Matters, PO Box 16557, Minneapolis, MN 55416, 2000

[14] Thomas Gilovich, "How We Know What Isn't So: The fallibility of human reason in everyday life", Free Press, NY, 1991

[15] Al Seckel, "The Art of Optical Illusions", Carlton Books, 2000.

[16] Colonel Michael Dewar, "The Art of Deception in Warfare", David and Charles Military Books, 1989.

[17] William L. Griego, "Deception - A 'Systematic Analytic' Approach", (slides from 1978, 1983)

[18] Scott Gerwehr, Jeff Rothenberg, and Robert H. Anderson, "An Arsenal of Deceptions for INFOSEC (OUO)", PM-1167-NSA, October, 1999, RAND National Defense Research Institute Project Memorandum.

[19] Fred Cohen, "Deception Toolkit", March, 1998

[20] Bill Cheswick, Steve Bellovin, Diana D'Angelo, and Paul Glick, "An Evening with Berferd" - followed by S. M. Bellovin. "There Be Dragons". Proceedings of the Third Usenix UNIX Security Symposium. Baltimore (September 1992).

[21] F. Cohen, "Internet Holes - Internet Lightning Rods", Network Security Magazine, July, 1996.

[22] F. Cohen, Operating System Protection Through Program Evolution Computers and Security 1992.

[23] F. Cohen, A Note On Distributed Coordinated Attacks, Computers and Security, 1996.

[24] Scott Gerwehr, Robert Weissler, Jamison Jo Medby, Robert H. Anderson, Jeff Rothenberg, "Employing Deception in Information Systems to Thwart Adversary Reconnaissance-Phase Activities (OUO)", PM-1124-NSA, Novermber 2000, RAND National Defense Research Institute.

[25] Robert E. Huber, "Information Warfare: Opportunity Born of Necessity", News Briefs, September-October 1983, Vol. IX, Num. 5, "Systems Technology" (Sperry Univac) pp 14-21.

[26] Knowledge Systems Corporation, "C3CM Planning Analyzer: Functional Description (Draft) First Update", RADC/COAD Contract F30602-87-C-0103, December 12, 1987.

[27] John J. Ratey, M.D., "A User's Guide to the Brain", Pantheon Books, 2001. [In contrast, the auditory nerve only has about 25,000 nerve fibers. Information must be assessed beginning in the ear itself, guided by the brain. "Evidence that our brains continually shape what we hear lies in the fact that there are more neuronal networks extending from the brain to the ears than there are coming from the ears to the brain." [27] (p. 93)]

[28] Sun Tzu, "The Art of War", (Translated by James Clavell), Dell Publishing, New York, NY 10036 (1983).

[29] Gordon Stein, "Encyclopedia of Hoaxes", Gale Research, Inc, 1993, p. 293.

[30] Fay Faron, "Rip-Off: a writer's guide to crimes of deception", Writers Digest Books, 1998, Cinn, OH.

[31] Richard J. Robertson and William T. Powers, Editors, "Introduction to Modern Psychology, The Control-Theory View". The Control Systems Group, Inc., Gravel Switch, Kentucky, 1990.

[32] Charles K. West, "The Social and Psychological Distortion of Information", Nelson-Hall, Chicago, 1981.

[33] Chester R. Karrass, "The Negotiating Game", Thomas A. Crowell, New York, 1970.

[34] Robert B. Cialdini, "Influence: Science and Practice", Allyn and Bacon, Boston, 2001.

[35] Robert W. Mitchell and Nicholas S. Thompson, "DECEPTION: Perspectives on human and nonhuman deceipt", SUNY Press, 1986, NY.

[36] Donald D. Hoffman, "Visual Intelligence: How We Create What We See", Norton, 1998, NY.

[37] Charles Handy, "Understanding Organizations", Oxford University Press, NY, 1993. img35.jpg

[38] National Research Council, "Modeling Human and Organizational Behavior", National Academy Press, Washington, DC, 1998.

[39] Bill Cheswick, An Evening with Berferd, 1991.

[40] Fred Cohen, "The Unpredictability Defense", Managing Network Security, April, 1998.

[41] David Kahn, "The Code Breakers", Macmillan Press, New York, 1967

[42] Norbert Weiner, "Cybernetics", 1954?

[43] The HoneyNet Project web site (www.honeynet.org).

[44] Tom Keaton, "A History of Warfare", Vintage Books, NY, 1993

[45] Andrew Wilson, "The Bomb and The Computer", Delacorte Press, NY, 1968.

[46] Robert Greene, "The 48 Laws of Power", Penguin Books, New York 1998

[47] Diana Deutsch, "Musical Illusions and Paradoxes", Philomel, La Jolla, CA 1995.

[48] Fred Cohen Cynthia Phillips, Laura Painton Swiler, Timothy Gaylor, Patricia Leary, Fran Rupley, Richard Isler, and Eli Dart, "A Preliminary Classification Scheme for Information System Threats, Attacks, and Defenses; A Cause and Effect Model; and Some Analysis Based on That Model", The Encyclopedia of Computer Science and Technology, 1999.

[49] Richards J. Heuer, Jr., "Psychology of Intelligence Analysis", History Staff Center for the Study of Intelligence

Central Intelligence Agency 1999.

[50] Aldert Vrij, "Detecting Lies and Deceipt", Wiley, New York, NY, 2000.

[51] National Technical Baseline, "Intrusion Detection and Response", Lawrence Livermore National Laboratory, Sandia National Laboratories, December, 1996

[52] Various documents, A list of documents related to MKULTRA can be found over the Internet.

[53] Kalbfleisch, Pamela J. The language of detecting deceit. Journal of Language & Social Psychology, Dec94, Vol. 13 Issue 4, p469, 28p, 1 chart [Provides information on the study of language strategies that are used to detect deceptive communication in interpersonal interactions. Classification of the typology; Strategies and implementation tactics; Discussions on deception detection techniques; Conclusion.]

[54] Colonel John Hughes-Wilson, "Military Intelligence Blunders", Carol & Graf, NY, 1999

[55] John Keegan, "A History of Warfare", Vintage Books, NY 1993.

[56] Charles Mackay, "Extraordinary Popular Delusions and the Madness of Crowds", Templeton Publications, 1989 (originally Richard Bently Publishers, London, 1841)

[57] Donald Danial and Katherine Herbig, ed. "Strategic Military Deception", Pergamon Books, 1982.

[58] Western Systems Coordinating Council WSCC Preliminary System Disturbance Report Aug 10, 1996 - DRAFT [This report details the August 10, 1996 major system disturbance that separated the Western Systems Coordinating Council system into 4 islands, interrupting service to 7.5 million customers for periods ranging from several minutes to nearly six hours.]

[59] Bob Pekarske. Restoration in a Flash---Using DS3 Cross-connects, Telephony. September 10, 1990. [This paper describes the techniques used to compensate for network failures in certain telephone switching systems in a matter of a millisecond. The paper points out that without this rapid response, the failed node would cause other nodes to fail, causing a domino effect on the entire national communications networks.]

[60] Mimi Ito, "Cybernetic Fantasies: Extended Selfhood in a Virtual Community", 1993.

[61] Mark Peace, "Dissertation: A Chatroom Ethnography", May 2000

[62] Daniel Chandler, "Personal Home Pages and the Construction of Identities on the Web", 2001

[63] Fred Cohen, "Understanding Viruses Bio-logically", Network Security Magazine, Aug, 2000.

[64] Fred Cohen, "Red Teaming and Other Agressive Auditing Techniques", Managing Network Security", March, 1998.

[65] SSCSD Tactical DecisionMaking Under Stress, SPAWAR Systems Center.

[66] Fred Cohen, "Method and Aparatus for Network Deception/Emulation", International Patent Application No PCT/US00/31295, Filed Octoboer 26, 2000.

[67] Heidi Vanderheiden, Boston University "Gender swapping on the Net?", http://web.aq.org/~tigris/loci-virtualtherapy.html