Chapter 11: UKUSA Deception Management and Cybernetics

In the following chapter the focus of the discussion is on the technology of 'Deception Management' in warfare by the United States and the United Kingdom and by extension all of NATO and other allies, such as Israel. As discussed previously deception management is used to disguise or otherwise camouflage true military intentions from an adversary. Along with management of deception comes the automated and computational nature of contemporary Deception Management, where the management is largely done by automated computational systems using Artificial Intelligence techniques to conduct the 'work'.

Effects Based Operations (EBO), US Reflexive Management

"[EBO]… goes against the very nature of war", - Gen. Mattis
(APDC, 2009)

Cost of Actions and Effects

In western military discourse Reflexive Control became known as Cost of Actions and Effects Based Operations. A quantification of the variables involved in taking an action and a quantization of the effects of those actions is the underlying principle. It is based in Deception, so is a universal military operation repeated whether one is a Red Team or a Blue Team. As the proliferation effect goes on in the Information Warfare battlespace.

> USJFCOM defines EBO as "*a process for obtaining a desired strategic outcome or effect on the enemy through the synergistic and cumulative application of the full range of military and non-military capabilities at all levels of conflict*." EBO places considerable importance on identifying and quantifying specific effects resulting from specific actions against specific targets. This requires vast information on the adversary; an aspect that sometimes draws criticism. It has been suggested that EBO requires unattainable levels of knowledge. (APDC, 2009)

Effects-based operations (EBO) is a concept that emerged during the Persian Gulf War for the planning and conduct of operations combining military and non-military methods to achieve a particular effect. An effects-based approach to operations was first applied in modern times in the design and execution of the Desert Storm air campaign of 1991. The principal author of the daily attack plans—then Lt Colonel, now retired Lt General David A. Deptula—used an effects-based approach in building the actual Desert Storm air campaign targeting plan, we also note that shock and awe strategy was a product of Col. Szafranski covered earlier. The doctrine was developed with an aim of putting desired strategic effects first and then planning from the desired strategic objective back to the possible tactical level actions that could be taken to achieve the desired effect. Contrary to conventional military approaches of force-on-force application that focused on attrition and annihilation, EBO focused on desired outcomes attempting to use a minimum of force. The approach was enabled by advancements in weaponry—particularly stealth and precision weapons—in conjunction with a planning approach based on specific effects rather than absolute destruction. Deptula, defined the goal of EBO; "If we focus on effects, the end of strategy, rather than force-on-force the traditional means to achieve it militarily, that enables us to consider different and perhaps more effective ways to accomplish the same goal quicker than in the past, with fewer resources and most importantly with fewer casualties." Others have postulated that EBO could be interpreted as an emerging understanding that attacking a second-order target may have first order consequences for a variety of objectives, wherein the Commander's intent can be satisfied with a minimum of collateral damage or risk to his own forces.

The main difference between EBO and typical warfare is that EBO involves all aspects of the government not just the military to achieve it's aims, including the use of Information Warfare, but not just targeted on troops in a battlespace but an entire society from head to bottom. The main development in EBO is that of automation which was developed by the United States through Lockheed Martin's Sandia National Labs. To study how automation is used in deception management and EBO we first must start with an understanding of how automation or what is really gamification of a battlespace using deception, which necessarily brings up a discussion of deceptive games as it applies directly to this part of warfare, even if gamified.

Reflexive Control, Deception Management, EBO are all part of leading an adversary away from their goals and is defined as such for deceptive games:

> Deceptive games are games where the reward structure or other aspects of the game are designed to lead the agent away from a globally optimal policy. (Togelius et al, 2018)

The main motivational factor in engineering behaviors is the individuals subjects perceived rewards, we covered rewards and games earlier. In deceptive games the reward function is reversed:

> If we see the reward function as a heuristic function approximating the (inverse) distance from a globally optimal policy, a deceptive reward function is an inadmissible heuristic. (Togelius et al, 2018)

> Deceptive games can be seen as exploiting a specific cognitive bias of the (human or AI) player to trick them into making a suboptimal decision. Withholding or providing false information is a form of deception, and can be very effective at sabotaging a player's performance. (Togelius et al, 2018)

As we have seen earlier this is the very definition of Reflexive Control whether in the Russian parlance or the UKUSA parlance of EBO or Deception Management. It is interesting to note that there are certain traps in deceptive games, which also seem to mimic common matchstick men maneuvers in the civilian world or simply marketing and advertising. These traps are greed trap, smoothness trap and generality trap.

Traps for deceptive games:

> **Greed Trap**: A common problem simplification is to only consider the effect of our actions for a limited future. These greedy algorithms usually aim to maximize some immediate reward and rely on the assumption that the local reward gradient will guide them to a global maximum. One way to specifically exploit this bias (a greedy trap) is to design a game with an accumulated reward and then use some initial small reward to trick the player into an action that will make a later, larger reward unattainable (Togelius et al, 2018)

This trap is basically very effective if you are trying to manipulate someone to do your bidding but have no plans of actually giving them a good payout, a version of a classic con artist. For instance you may receive valuable information for inconsequential bits that lead you to believe you have trust when they are setting you up only to pull the rug out from under you after reinforcing trust with accurate inconsequential information before.

> **Smoothness Trap**: Several AI techniques also rely on the assumption that good solutions are "close" to other good solutions. Genetic Algorithms, for example, assume a certain smoothness of the fitness landscape and MCTS algorithms outperform uninformed random tree search because they bias their exploration towards branches with more promising results. This assumption can be exploited by deliberately hiding the optimal solutions close to many really bad solutions. Since many of the solutions along the dangerous part lead to losses, an agent

> operating with the smoothness bias might be disinclined to investigate this direction further, and would therefore not find the much better solution. (Togelius et al, 2018)

This trap is the pony under the manure trap, you hide a pony under the manure most people pass it by, unless you have an exploratory algorithm that hops around randomly looking for hidden ponies under piles of manure.

> **Generality Trap:** Another way to make decision-making in games more manageable, both for humans and AI agents, is to generalize from particular situations. Rather than learning or determining how to interact with a certain object in every possible context, an AI can be more efficient by developing a generalized rule. For example, if there is a sprite that kills the avatar, avoiding that sprite as a general rule might be sensible. A generality trap can exploit this by providing a game environment in which such a rule is sensible, but for few critical exceptions. Thin Mints aims to realize this, as eating mints gives the AI points unless too many are eaten. So the agent has to figure out that it should eat a lot of them, but then stop, and change its behavior towards the mints. Agents that would evaluate the gain in reward greedily might not have a problem here, but agents that try to develop sophisticated behavioral rules should be weak to this deception.
>
> We should also note that most of the deceptions implemented here are focused on exploiting the reward structure given by the game to trick AIs that are optimized for actual rewards. Consider though, that recent developments in intrinsically motivated AIs have introduced ideas such as curiosity-driven AIs to play games such as Montezuma's Revenge or Super Mario. The internal curiosity reward enhances the AI's gameplay, by providing a gradient in an extrinsic reward landscape, but in itself makes the AI susceptible to deception. One could design a game that specifically punished players for exploration. (Togelius et al, 2018)

With this in mind we shall see how rewards and game play are integrated into military simulations for the purposes of training in war situations.

One other convergence between gaming and Deception Management is that of the need to understand behaviors from a computational stand point. This is one of the reasons the NSA and GCHQ infiltrated video gaming social networks, studying human behaviors. The main reason is that of data sparseness, you need a lot of data to predict behavior:

> We argue that for studying national security issues, where data is sparse, it is difficult to experiment, and behaviors can be complex and varied, online games can serve as a unique and powerful tool to experimentally understand causal relationships. (Epifanovskaya et al, 2018)
>
> Data on millions of actions performed by a large and diverse sample of people lends itself well to statistical analysis, better than surveys and laboratory experiments with much smaller sample sizes taken from a more homogeneous group (e.g., college students, who frequently participate in academic human research studies (Gosling et al., 2010; Henrich et al., 2010). They also offer the opportunity to see how different types of players respond under different circumstances; useful for interrogating differences among players, but also, in wargaming, uncovering novel strategies that would not have occurred to personnel typically involved in these games. (Epifanovskaya et al, 2018)

It is worth pointing out that using data from a ‘game’ world, including studying people while they are immersed in that game world, would lead to some strange bias if applied to society in general. It would also lead to overly stereotypical analysis as data analysis of complex social interactions would tend to over generalize to the mean while ignoring anything it can’t easily classify and cluster, even potentially creating lines of clusters that don’t fit natural reality.

A Blueprint for an Information Warfare Engine

In Information Warfare there is a need to psychologically model the adversary, Lockheed-Martin through it's teams at Sandia National Labs have developed a psychological-cognitive engine which was designed specifically for Counter-Intelligence and **Influence operations** as noted by the researchers, it is also noted that this **work was not funded through government funds so thus remains the sole property of Lockheed-Martin to sell to anyone not on US lists that deny export**s, such as Iran, China, Russia, etc. It could be sold to any civil war party for instance not on the export ban list. Thus, making tracing such engines more difficult from a forensics perspective, unless you can develop fault-proof highly accurate forensics. It is also important to remember as a non-DoD funded private project that it does not face any kind of government oversight from the US Government.

Following up on the study of Terrorist networks, work from which one can engineer anti-terrorism and terrorism, it is important to look at the work of US Homeland Security in this area conducted by Sandia National Labs in influencing individuals towards and away from terrorism. In the work of Backus & Glass 2006 we see how this work is theorized. According to their abstract their work encompasses:

> This document presents the conceptualization of an agent-based, simulation framework that allows the use and testing of various social and behavioral science approaches for understanding the motivation and intent associated with terrorist activities. The framework design provides a LEGO™ -style toolbox that can convert sophisticated SME theses on individual and social behavior into computationally tractable, mathematical representations. Through parameterization, the reconfigurable framework can then simulate the dynamics of any particular group or interacting collection of terrorist groups. (Backus & Glass, 2006)

As noted DHS has sponsored this research in creating and fighting terrorists, for this paper it is specifically carried out by the DHS Motivation & Intent Thrust Area, this work "intends to use computerized models to ultimately improve the efficiency of intelligence analysts as they attempt to assess terrorist threats. A model is a machine, and like a machine, it can only perform the function for which it was built if its design and construction were sufficient. The successful development of an M&I model requires a clear understanding of how to combine modeling methods with the subject matter of interest." (Backus & Glass, 2006) This work is part of the larger Human Factors studies of DHS. The Human Factors/Behavioral Sciences Division of the Department of Homeland Security's Science & Technology Directorate has as its mission to "advance national security by developing and applying the social, behavioral, and physical sciences to improve identification and analysis of threats, to enhance societal resilience, and to integrate human capabilities into the development of technology." Its past mission statement said "know our enemies, understand ourselves; put the human in the equation."

This work is intended to model human individuals in networks, based on computational algorithms. The main emphasis behind such modeling is the use of Gradient Descent [see Appendix D], which often can lead to over-fitting and under-fitting of data, nonetheless, they believe you can study natural phenomena through such modeling even in highly complex social contexts, which is not what Gradient Descent is designed for, rather very simple mathematical models such as a robot arm moving a screw in a car factory where everything can be measured with exact mechanical precision. It is noted that such work as modeling human behavior can be a 'wicked' or almost intractable problem.

> "A wicked problem usually has the characteristics that many people have tried and failed to solve the problem, that most attempts to solve the problem actually make the problem worse,

> and that nobody really agrees on exactly what the problem actually is! Wicked problems typically arise in feedback systems, and almost any interesting system is a feedback system." Addressing wicked problems requires a recognition that "priors" (preconceptions based on assumption or intuition) can distort the ability to see or accept alternative formulations of, or approaches to, the problem. New methods in Bayesian analysis appear to allow the determination of what causal mechanisms a simulation should include or, conversely, cannot reject in a social-science model. A modern statistical method called co-integration, further allows the determination of those feedback processes that dominate behavior. Sensitivity analysis methods have been successfully applied to combined societal-behavioral and physical systems that are dominated by feedback. Empirically, the feedback assures that only a few mechanisms control behavior under any given condition and these key mechanisms are readily determined. Analysts can than concentrate on those key elements. None of the above noted methods have yet become widely used within the social-psychological (behavioral) community. As such, conclusions drawn about what is not a cause of terrorism, using current statistical practices, might be misleading." (Backus & Glass, 2006, 6)

How this analysis works is that it views each person as a mapable agent in the sense of AI Agents, people do not exist in solitary lives, rather we are social animals that usually operate in herds, in this sense they are talking of herd management like a Farmer. The social existence of people takes place in Groups, through the mathematical representation of Groups these groups can be steered toward terrorism or away from terrorism.

> Within an agent-based simulation, simple individual mechanisms, such those associated with relative-deprivation or frustration, that alone cannot account for the group level action, can interact with other primal behaviors to produce the group behaviors that social movement theory attempts to describe. Agents behave differently in different environments and under the influence of different histories. The varied and seemingly inconsistent histories of various terrorist groups could simply be a consequence of the specific history and environment that shapes the collective agent trajectory. A few simple mechanisms may explain a wide variety of dynamics, and can thereby offer an understanding (and control) of terrorist motivation and intent. Additionally, violence comes from a relatively small element of any society. Agent-based models simulate a distribution of individuals. In essence, the agents reflect the full spectrum of possibilities and can produce the full spectrum of actualizations. (Backus & Glass, 2006, 7)

In this work they view humans as Lego blocks that are easy to reconfigure:

> While groups can be represented as an interacting and evolving set of individuals, individuals (agents) can be represented by a primal set of behavioral mechanisms. The mechanisms are the language that expresses the individual behavior and its connection with other individuals (and the environment). That language must be mapable to/from the language that subject matter experts (SMEs) use to describe behaviors and dynamics. For human behavior to be so complex, it must be composed of a relatively small number of simple mechanisms. If there were a large number of complicated mechanisms, the mechanisms would only work in the situation for which they were designed. Simple mechanisms can fit together and interact in near infinite variety. The concept is no different than that of jig-saw puzzle, where complex pieces can only make one picture, but simple pieces such as colored triangles, can make any picture desirable (albeit, with a few negligible, rough edges). Lego™ pieces can portray very complex shapes because they allow both two and three dimensional constructs. (They even allow four dimensional concepts if the pieces are hinged, and, thus, can vary over time). The Lego™ metaphor seems appropriate for an agent-based framework. The use of a few key (behavioral) building blocks can decompose the logic of a SME. The replicated use of those blocks can then form the representation for any group or interacting set of groups. If the building blocks are well designed, their parameterization with historical data will result in a

realistic and useful representation of the group relative to the future motivation and intent (M&I) dynamics of interest.

They seek to define the Lego blocks view Subject Matter Experts (SME), initially the SMEs were human analysts in later work they are automated SMEs. They believe you can predict behavior from past history based on the blocks understanding. Although they do note that this is only applicable to simple models not complex ones. What they are seeking to formulate are behavioral changes via game theory and optimization. The extensive use of Game theory is noted throughout so called 'anti-terror' operations in an effort to formulate 'intervention' strategies, which can also be flipped around and provide motivational strategies to create terrorist incidents. There are four primal building blocks to formulating the intervention strategies:

> As will be discussed in detail below, the current model design only includes four primal building blocks (components): expectation-formation, response-behavior (choice), coping-capacity, and opinion adjustment. Opinion-adjustment combines coping-skill and behavioral-response dynamics into a single component applicable to a key subset of terrorist dynamics. (Backus & Glass, 2006, 11)

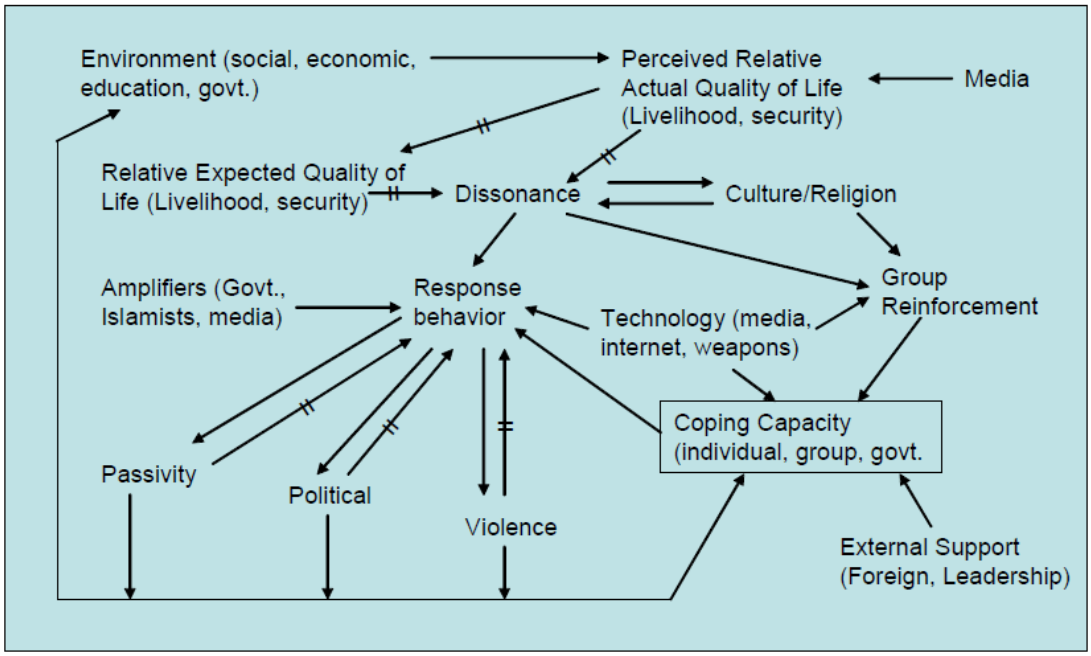


Figure 4: Model Components.

In a visual mapping of this process and how the moving parts of the machine model work Backus & Glass provide the following flow:

> The difference between expectations and perceived reality produces a dissonance.3 The cross-line in the arrows of causality in Figure 4 indicates delays in response. It takes time to gestate and incorporate information. Religion and culture present a form of coping skills. As dissonance occurs, people seek (or withdraw to) religious or cultural traditions that allow a rationalization of conditions. These psychological anchor points act to mitigate the dissonance. The simple representation noted in Figure 4 denotes the choice to seek religious and cultural moral support. Secondary aspects of religions will become clear momentarily. Depending on 1) the ability to act (coping-capacity), 2) the existing effectiveness of violence or political activities, 3) outside influences, and 4) the level of dissonance, the individual makes choices toward violence, passivity, or political activity. Group dynamics can dominate the response-behavior.

> The group amplifies the coping capacity of the individual and reinforces a predilection toward continued association with the group and its function. In this example, the groups represent proponents of the three responses simulated (passivity, political, violence). (The model can include additional responses and distinctions as actual or analogous data allow.) External (exogenous) support from foreign governments/resources and exceptional leadership personalities can dramatically augment the amplifying affect a group has on an individual. The internet makes it easier to join a group both in effort and risk terms. (In economics, this phenomenon is called reducing the transaction and hurdle costs). The media can distort the information an individual uses to make a choice. Weapon technology also amplifies the coping capacity of an individual, just as a power tool amplifies the capability of its user. It is not only the violence or political activism that an individual sees in his social environment, but also the counter (or precipitating) violence or political activity of the government (or government surrogate) that affects the choices made. The reaching out to religious or cultural organizations that resonate with the level of dissonance act primarily to increase the probability of contact with groups that reinforce rather than mitigate the dissonance. Again, the Internet and media also increase the probability of contact of similarly dissonant souls. The added affect of Islamist education will be discussed later. From a sociological and psychological level, dissonance is the measure of motivation and the behavioral response is the measure of intent. From a DHS perspective, the intent is multivariate, and the coping capacity and group-reinforcement qualify as operational measures of motivation. (Backus & Glass, 2006, 12)

One can see the similarity to the work of Tarashenko, covered in a previous chapter, in reconfiguring groups and it's application here in a US national defense contractors version of the same. Influence is about getting the right information into the mind of the target of that information. The work here involves the component of dissonance, which produces alienation:

> The dissonance between perceived conditions and expectations produce alienation. This alienation can lead to antisocial behavior (Sageman). Within the model, dissonance is strictly defined as the proportional difference between perceived conditions and Only when there is a gap between perceived reality and expectations, is there an incentive to act. Many authors indicate that dissonance associated with social and political expectation acts as the starting point on the path to terrorism (Crenshaw; Drummond; Silke) Religion and culture affect dissonance. Embracing either can improve apparent coping capacity (discussed later). The choice to seek religious and cultural-support increases with dissonance (Dennet). The direct impact is to reduce dissonance and temper the actions that dissonance might engender. As will be noted later, large negative indirect impacts are also possible. This "religion" dynamic is included for completeness and to illuminate the multiple impacts such phenomena as religion and media play in terrorist dynamics. As such, only a simplified coping-capacity plus response-behavior representation of religion is included in this part of the model. Because expectation formation is a filtering (long-term) process, even after conditions change, dissonance changes slowly – as experienced in post-war Iraq. (Backus & Glass 2006)

Key to dissonance and alienation is coping capacity which is influenced by social structure and relationships. Coping capacity is defined by Backus et al:

> Anger, or even hate, to the point of wanting to kill somebody, will be void of action if there is no perceived capacity to carry out the act. The psychological coping capacity of individuals determines their ability to respond to conditions. The adaptation to environment comes from changes in coping skills (Helson). If new conditions are excessive compared to coping skills the individual will succumb to the external pressure (flight mode). If the change is within the normal operating range of the coping skills, the individual will counter the new pressure (fight mode). If the new condition is trivial compared to the coping skill, the individual does little and need to do little to (successfully) respond. If the challenge is slightly larger than the average

> coping capacity, the coping skills increase over time to match the environment. (Backus & Glass, 2006, 15)

Obviously coping skills are an important aspect to adaptation, too much adaptation required overloads coping capacity. A key component to the strength of coping capacity is that of group or social reinforcement:

> It is the group dynamics and the behavioral reinforcements that seem to be important to terrorist outcomes...accordingly. The individuals in the group reinforce each other. As noted earlier, the coping grows and declines based on the interaction between the individual and the environment (in this instance, the group). Individuals within the group influence and reinforce each other. (Backus et al, 2006)

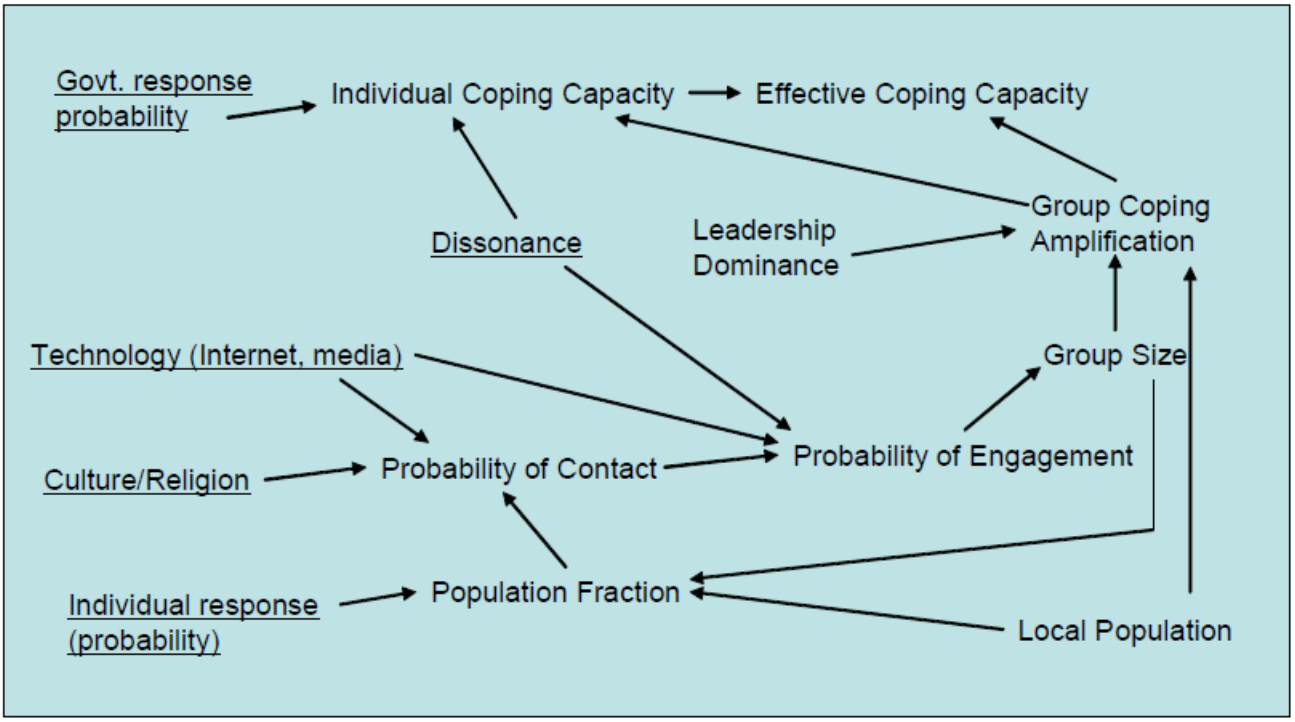


Figure 5: Group Reinforcement Detail

The backing of the group enhances the coping capacity of the individual (e.g., the stereotypically insecure bully being backed by a group of his friends. Group dynamics might instead be the key "amplification" process. (Note that the government is just another agent in the model.)

Backus et al provide an example of 'opinion-adjustment' within a networked agent-based model. This shows how to influence groups, intended to thwart terrorist group formation, which of course can be inverted like all programs.

> The simplified model combines the threshold based approach to the agent activation to violence of Epstein with the dynamics of opinion formation of Weisbuch and coworkers. Epstein represents an agent's action level by its "Grievance", a combination of two agent state variables: perceived "Hardship" and perceived "Legitimacy" of the local regime or status quo. The term Grievance has an interpretation comparable to that of Dissonance or Motivation. Hardship and Legitimacy are perceptions, or opinions, of individual agents and are the result of interactions with other agents, information from the media, environment, events, etc. In

> Epstein's analysis, Hardship or Legitimacy are static variables, however the example implementation allows them to be dynamic and to arise through interaction with other agents via the opinion formation process of Weisbuch. Risk of capture or worse, may temper the intent to perpetrate an act of violence. Risk is a function of the individual's inherent aversion to risk and the probability of being discovered and suffering an undesirable outcome. If the motivation, adjusted for risk, exceeds a threshold, an individual will shift to a new (possibly violent) action state. (Backus et al, 2010)

Opinion formation is very significant in the process of influence it is noted that media plays a key part in this formation. Later, I will cover Memetic Warfare the uses of image media to create groups.

We covered the computational modeling of emotions in an earlier part focusing on the work of Tarashenko. Here we see how the US Government labs tackle the issue of modeling human cognition and emotion into their systems. In work dating back to 2010 Backus et al developed a computational framework for automating influence operations and deception:

> "...key features of the SNL psychological engine. The engine is designed to be a generic presentation of cognitive entities interacting among themselves and with the external world. The engine combines the most accepted theories of behavioral psychology with those of behavioral economics to produce a unified simulation of human response from stimuli through executed behavior. The engine explicitly recognizes emotive and reasoned contributions to behavior and simulates the dynamics associated with cue processing, learning, and choice selection. Most importantly, the model parameterization can come from available media or survey information, as well subject-matter-expert information. The framework design allows the use of uncertainty quantification and sensitivity analysis to manage confidence in using the analysis results for intervention decisions." (Backus et al, 2010)

As mentioned the engine is a simulator of human emotion based on cue processing. Below we cover cues and contexts. The framework is succinctly put as:

> "...a computational framework for analyzing the behaviors of individuals and populations, over time, in response to information operations, diplomacy, and other intercessions."

It is important to note that it is a feedback system that it creates responses from actions done in areas such as state diplomacy, etc. One of the differences between this software and that of usual US MilDef is that they replace 'course of action' with 'intervention':

> We use the concept of "intervention" rather than "course of action" because we only emphasize those endeavors that intervene to affect the behaviors of the system and individuals of interest. (Backus, 2010, 7)

Fundamentally, the software is about steering an individual and collectively a society. It tries to influence decisions by presenting different scenarios with different trajectories:

> The framework is based on first principles that can encompass an unlimited number of entities with any number of alternative decisions, and with any level of interrelationship complexity. Because we only allowed the use of theories that 1) were mutually self consistent, 2) would integrate into a complete representation of behavior from stimuli through to action, 3) would translate to a unique set of computational equations, and 4) could be instantiated, tested, and verified using accessible date, we can 1) readily use available data on individual or regions to calibrate the model, 2) use Subject Matter Expert (SME) data to augment data sparsity, 3) test hypotheses about alternative interventions and behavioral responses, 4) quantify the uncertainty (risk) that an intervention will produce the desired results, and 5) follow the time-

dependent consequential counter-responses from an intervention. Most importantly, the framework naturally captures the implications of new (even unique) information flows such as may be considered in information operations or other interventions. (Backus, 2010, 8)

Some of the principles involved in it's design are Bounded Rationality (Simon) Qualitative Choice (McFadden) Imperfect Information (Stiglitz) Risk Asymmetry, Stock & Flow Cointegration. We shall cover some of these below. What we are talking about here is the previously discussed 'Influence Machine', a machine built to shape public opinion through automated processes. It creates a model of individuals ('entities') then uses a cognitive map to influence the individual ('entity').

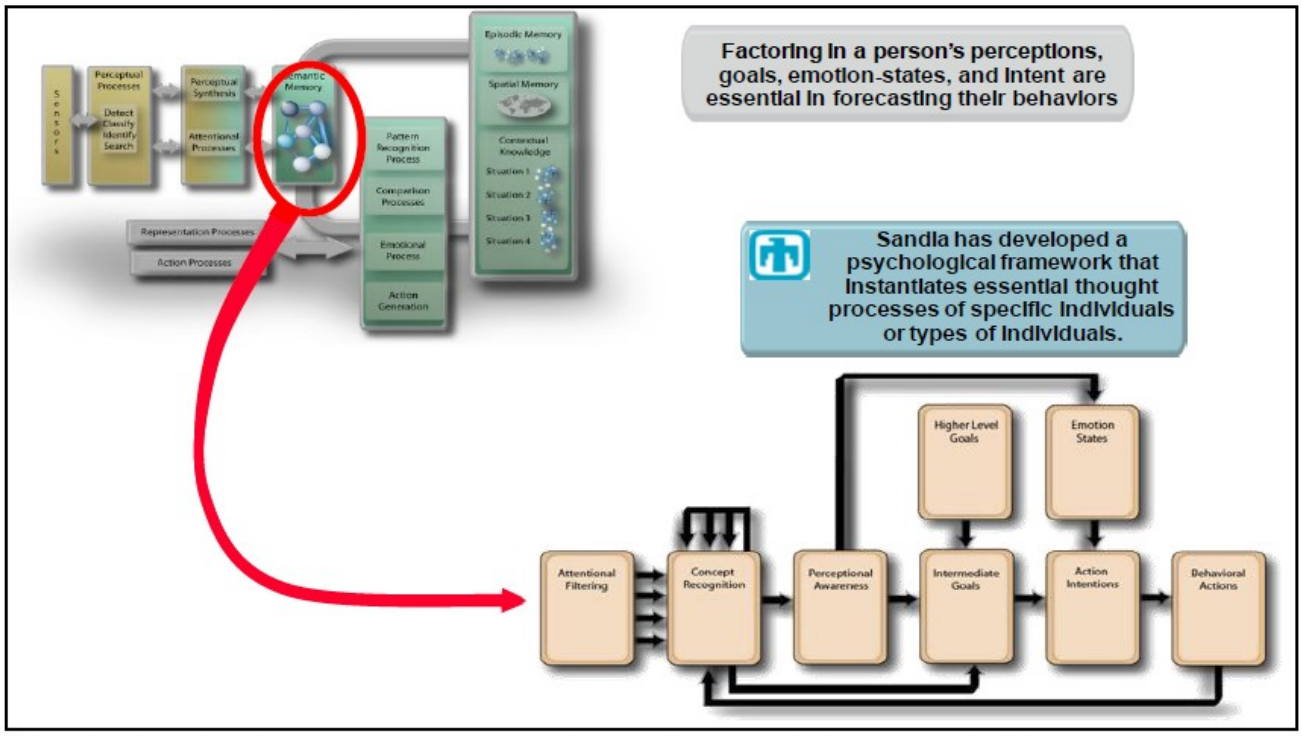


Figure 3: Sandia High-Fidelity Cognitive Modeling A Modeled Entity

The mathematical representation of these processes is largely a re-application of the Backus and Glass (2006) to match the specific needs of influence operations. Here we are talking about influence agents, and agents in general in the sense of computer game programming. Backus has a sepcific model of an Agent as shown in the following diagram:

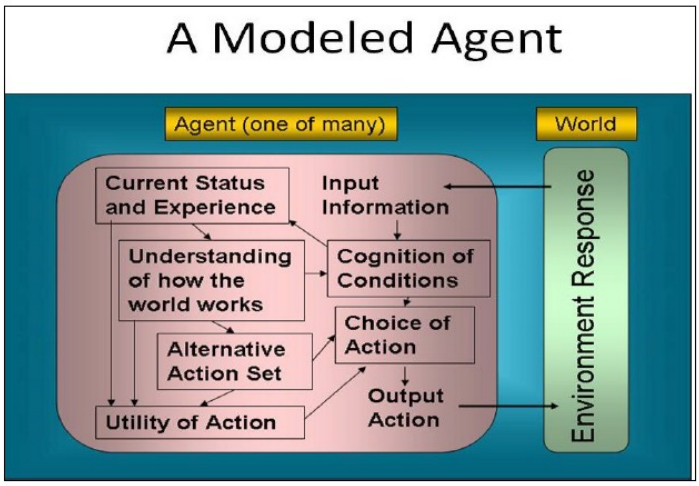


Figure 4: Agent-Based Modeling

(Backus 2010, 12)

We can see the typical interaction between sensors to collect environment variables then the processing of inputs in the Agent software with costs being represented by Utility of Action which is informed by such tings as a knowledge base, 'understanding how the world works', eventually leading to an output action based on the algorithmic processing of inputs. The computational model used by Backus et al is as follows:

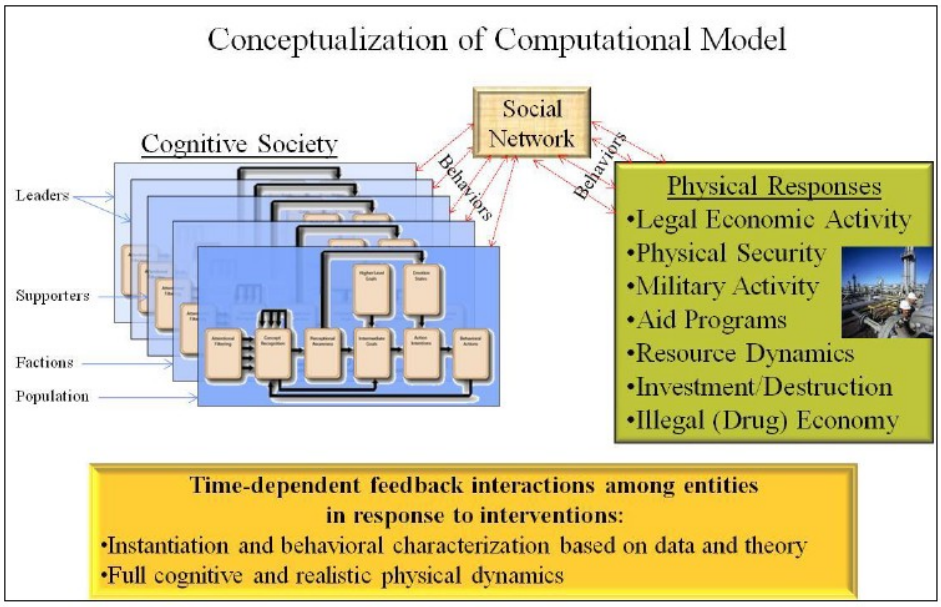


Figure 5: Structural Overview of the Unified Psychological Model

It is interesting to note that the cognitive society is divided into various layers: leaders, supporters, factions, populations. Which we see also studied in understanding how online gaming clans work as studied by the NSA and GCHQ in efforts to understand these layers mechanics. All interacting in social networks and the behaviors associated with different social networks mapped out for computation. All of which has a feedback interaction among the entities (people) in response to 'cost-of-actions' or 'interventions' dished out by automated algorithms.

Elements of Psychological Profiling for Computation:

The basic components of the SNL are comprised of psychological concepts such as attitudes, notions, expectations, passivity, incongruity and cognitive resources. All of which form a complex which an action is molded by, behaviors. The following is an overview as presented by Backus et al:

:

> Attitudes are pattern derived from cognitive resources. Attitudes can be affective or rational. Some types of attitudes have evolution-based and cultural based components. In essence, attitudes are an internalized notion affecting the interpretation of external cues (Backus, 2010, 23)
>
> Notions: A collection of Cues forms a Notion. Notions are then a collection of selected stimuli relevant to specific intents and behavior. Notion can be affective or reasoned and influence the utility an intent (choice) or behavior has. The sensory aspect of notion formation is identical in

construction to attitudes but is composed of the cuing stimuli rather than cognitive resources. Nonetheless, attitudes (by altering the β of the pattern strength equation) can affect how cues produce a specific notion. Elaboration Likelihood Theory considers notions as based on differing patterns of perception that then feed into the utility of actions whose components may include higher cognitive considerations as well as emotive elements for support biasing of intentions. (Backus, 2010, 24)

Backus et al drills down into the elements which drive this process:

- **Notion Assimilation** Psychological studies and experiments indicate that individuals only remember the peak and end value of sensory input (Fredrickson, Schwarz). Further, the peak values from multiple sensory episodes are not additive. Other studies indicate that cuing frequency and recency play a role in behavioral responses (Perugini). As noted previously, the lingering aspect of sensory notions can have an emotive content called "mood." The speed at which a cue becomes a notion is faster for an emotive notion than a reasoned notion. The lingering effect is longer for an emotive notion than reasoned notion. Hence, there can be mood-congruent behavior where the phasing of reasoned and affective notions play off each other.

- **Expectations** Incongruity comes from the difference between perceived conditions and expected conditions. Expectations come from the memory of prior conditions. Nonetheless, the expectation for the future may not coincide with the exact memory of the past conditions if there is anticipation of change, such as raise or a job promotion.

  The psychologist Hogarth notes that almost all decisions are based on expectations. Further, the shortcomings of human cognition and information assimilation mean that decisions are based on limited information and selective perceptions. Additionally, new work shows that humans use information to build up “priors” that are then used as a reference bias for subsequent decisions under the normally existing conditions of uncertain and deficient data.

  Humans form expectations and make decisions about issues that are worth the effort. In the current conception of the model, separate considerations of security and livelihood act as surrogates for the key issues driving individual and societal evolution. Having adequate food, housing, income, and employment (livelihood) means little if you won’t live long enough to enjoy it (security). An education brings with it expectations of a better livelihood. Lawlessness and excess government repression add to a sense of insecurity. Weighted sums of socioeconomic and security indicators produce livelihood and security indices, respectively. A multiplicative combination of security and livelihood indices produce a quality of life (QOL) index that captures the key dynamics characteristics important to decision making. Over time, the perceived current QOL evolves into the expected QOL. (Backus & Glass, 2006, 13)

- **Passivity** Passivity is simply an attitude that affects the offset associated with incongruity. In a sense, passivity is an attitude toward incongruity. A high degree of passivity means that there needs to be a large disparity between existing and "normal" conditions before the individual recognizes a need to act. Passivity determines the changing sensitivity to incongruity. Estimating the parameters of passivity would require an extended data time-series, but the impact of passivity is secondary and can be neglected for most studies

- **Incongruity** Incongruity is the proportional change between perceived conditions and expected conditions. It is the primal dissonance driving the reaction to external stimuli (cues). It has a dead-zone response using an offset (as discussed above) to capture the threshold effect. Figure 16 shows how incongruity changes as the proportional difference between actual conditions and expectations vary.

- **Cognitive Resources** Cognitive resources are the accumulation of experiential learning tempered by evolutionary constraints. They can contain emotive conditioning (such the fear of dark alleys) or the acquisition of a physical capability (such as playing a musical instrument). Incongruity initiates learning. Learning changes the process for coping with the environment. If the incongruity is small, it means that individual is well suited to respond to existing conditions and has probably controlled the existing condition to correspond to expectations -- through previous behaviors.

  Components of Learning inhibition Excitation Conditioning improves the level of cognitive resource through the conditioning response.... The conditioning can improve a cognitive resource until behaviors bring incongruity levels to within acceptable ranges. The level of a cognitive resource grows to slightly exceed the level needed to accommodate external stimuli. This phenomena has a basis in evolution where 31 there needs to be a contingency if allows the individual to tolerate conditions that exceed previously experienced values. In brief, repetitive tolerable, stress producing (incongruity) events modify cognitive resources to cope with that environment. (Backus, 2010, 30-1)

  Cognitive resource is a broad term reflecting a learned capability for responding to notions (patterns of relevant cues). A pattern of cognitive resources represent an attitude. The attitude may have a reasoned or emotive basis and it can reflect a propensity for response or perceptions of which an entity is not consciously aware. Anything learned (knowledge, belief, emotional response, intuition) other than pure memory of past conditions for making expectations, is a cognitive resource. The model dynamics indicate that “motivation” is the circumstance whereby perceptions are large enough to offer a challenge, yet small enough to ensure adequate response with readily achievable effort (Grossberg, Yerkes). This aspect is reflected as the excitatory and inhibitory components of conditioning. The Cognitive Resources include belief, knowledge, experience, and emotive levels of memory to decisions. The current logic is based on coping skill dynamics (Backus 2006). (Backus, 2010, 89)

Interventions interact with all the above elements, along with these elements that occur within or internally to an agent or entity or person it is also necessary to understand how they interact socially:

> To model the consequence of interventions, it is necessary to not only model the initial behaviors of affected individuals, but to also determine how interactions with other individuals and the physical world, over time, can alter the outcome. The changes over time are called dynamics. The feedback processes among individuals and the physical world unfold dynamically and cause the outcome of an intervention to, for example, start off going in the desired direction, but in the long term lead to counter-responses that generate new concerns without improving the original issue. The delay between behaviors and impacts can cause secondary dynamics that make it extremely difficult to know whether the ups and downs of behavioral responses and counter-responses will ultimately lead to the desired outcome. (Backus, 2010, 13)
>
> The process for developing a psychological model using the system dynamics methodology starts with a description of the psychological theories the model must simulate. These theories need to encompass all the salient considerations needed to make a comprehensive system‟s model describing the problems of interest. Note that there is no attempt of model the entire

> system, but only those aspects of the system relevant to the problems to be addressed/analyses. The next step is to develop a causal-loop diagram the causally relates all the interactions embodied in the theories. The casual loop diagram is next mapped to a stock-and-flow diagram that explicitly details the flow of information and physical quantities through the system. A key feature is the designation of stocks that represent the accumulation of information, experience, monetary, or physical quantities. These stocks are called “state variables” and they largely characterize the nature of the system and its responses. The difference in the value of stocks over time increments is the “differential” part of the differential-equation approach to computational modeling. The exact mathematical expression of the theory is anchored in the accumulation of flow into and out of the stocks. The mathematical expression of the flows comes from a causal interpretation of the theory into the language of mathematics. The key equations will be described later in this report. Only those theories that have a measurable meaning, supportable, a least in principle, by historical or experimental data, are included in the model. The data determines the parameters that control the progression of the simulated values through time. (Backus, 2010, 13)

Management of a situation is always the prime concern of military intelligence. The management of military assets is not just considered in this framework, rather it is more general then a specific implementation for a specific behavior and specific tasks. The management of ‘flows’ whether monetary, natural resources, or human populations is the main objective of system dynamics method, which is anchored in the accumulation of flow into and out of the stocks. How are flows affected is a matter of human actions and the steering of human actions through cues:

> Stimuli are the physical realization of world conditions and of human action. When an individual places these stimuli in context, they become cues that inform or affect behaviors. The grouping of cues forms a pattern. For example, the observation of asphalt, cars, sidewalks, and buildings act as cues, giving you the notion that you are on a city street. We use the term “notion” rather than “perception” because the term "perception" can often denote a higher level of cognition than the recognition of simple physical stimuli, such as, the higher-level perception that quantum mechanics better explain atomic phenomena than thermodynamics and opposed to primal sensation of "that pin is sharp!" Notions typically take on importance when they are incongruous with (different from) expectations. Expectations are often the memory of the status-quo or the anticipation of future conditions. Cognitive resources are our learned attitude toward a condition (the condition being a perceived notion or incongruity) or our learned ability to respond to a condition. Our cognitive resources and perceptions of a situation (via notions and incongruities) act together to help us evaluate the choices we have to respond to those conditions. The result represents our intentions. The execution of those intentions further depends on the level of the incongruity and our attitudes toward that behavior. Once we initiate a behavior, it takes time before it becomes an action affecting the external world (including other individuals). Depending on the proximity or our social network, the realized consequence of our actions becomes the cues to some individuals but not to others. (Backus, 2010, 15)

Proximity within social networks amplifies the actions of an individual, those closer will tend to mimic the person that is a center of action. Expectations are a central part of the whole process of behavior intervention by manipulating intentions. “Notions typically take on importance when they are incongruous with (different from) expectations. Expectations are often the memory of the status-quo or the anticipation of future conditions.” Weight is added to a notion, an engineered notion or a free-will notion, by the manipulation of expectations, the more unexpected the more important the notion.

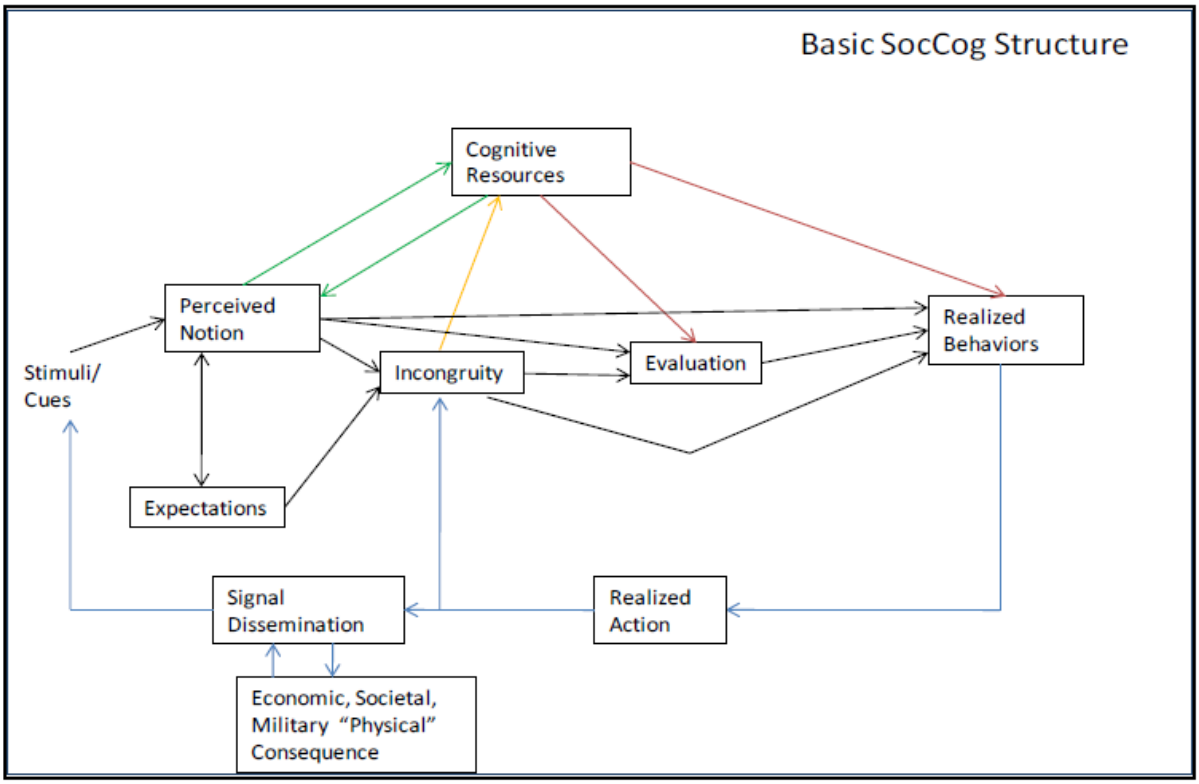


Figure 6: Computation Elements of the Behavioral model.

The formation of notions is dependent on the information being fed into the system. One can use Subject Matter Experts (SME) or an automated process, which is favored today. Based on the background information behaviors are formulated and a process of influence is calculated:

> ...feedback logic of one entity"s behavior becoming another entity"s stimuli (cues), possibly through the intermediation of external physical processes, explicitly captures the social network considerations that are often the domain of more-abstract agent-based modeling. An entity is an individual or a group. The approach for this modeling is made possible by assuming a fixed set of potential behaviors embodied in a representation of the individual. The representation contains the preferences and personality characteristics pertinent to the relevant decision-making. It is called the "blueprint" and it fully characterizes a specific individual or group of individuals. While the magnitude of interactions may change, the model does not produce new paths of cognition. All potential interactions are determined via initial parameterization of the model. Over the time, frame of the model simulation (at most a couple of years and often on the order of weeks), there should be little possibility, and there is little predictive capability for modeling, that entities would change their behaviors outside the domain of their historical experience and habits. (See Appendix 15 for an expanded discussion on relaxing the assumptions of a "blueprint" approach.) The mathematical expression of what stimuli cause cues and what choices or behaviors those cues can invoke has to be determined a priori thought the use of subject matter experts (SMEs) and available data. SMEs can hypothesize notions and perception that are not reflected in the data. Analytical methods can allow an estimate of how those hypothesized behaviors could occur based on the knowledge of an individual"s behaviors in other circumstances. The singular personality of an individual has a large affect on all his or her decisions. Uncertainty analysis could determine the potential for

such a behavior to affect the policy selection of external security interventions. (The assessment of these interventions is the actual purpose of the model, not the pin-point prediction of individual behavior.) The model cannot generate potential behaviors that are beyond the imagination of SME‟s and are not reflected in available data. These unknown unknowns are a limitation in all realms of physical and social science. Nonetheless, the “blueprint” for an individual within the model is the best representation of that individual‟s behavior characteristics available. Any other representations would be less valid and less reliable. Figure 7 [see below] show an exploded view of model structure of Figure 6 [see above]. Each block in the diagram contains only one to two equations. Each of these equations has a simple theoretical construct that will be discussed briefly below and fully elucidated in the Appendices. . Note however that each block can process large flows of data. There are typically a large number of stimuli generating a large number of notions, leading to a large number of potential choices and behaviors, across a large number of individuals. Some of the differences to note between Figure 6 and Figure 7 are the decomposition of the "Perceived Notion" in Figure 6 into several subcomponents in Figure 7, such as the sensory and assimilated notion. It takes time to cognitively recognize a set of cues. Cues can also produce emotive notions that characteristically occur faster than cognitive notions and use minimal information. The emotive notions can set the “mood” for processing the cognitive information, often adding a risk aversion element to the choice invoked by the cognitive information . Research shows that emotive and non-motive components are both part of the normal processing that leads to behavior. The model explicitly recognizes and uses both these categories of information flow. (Backus, 2010)

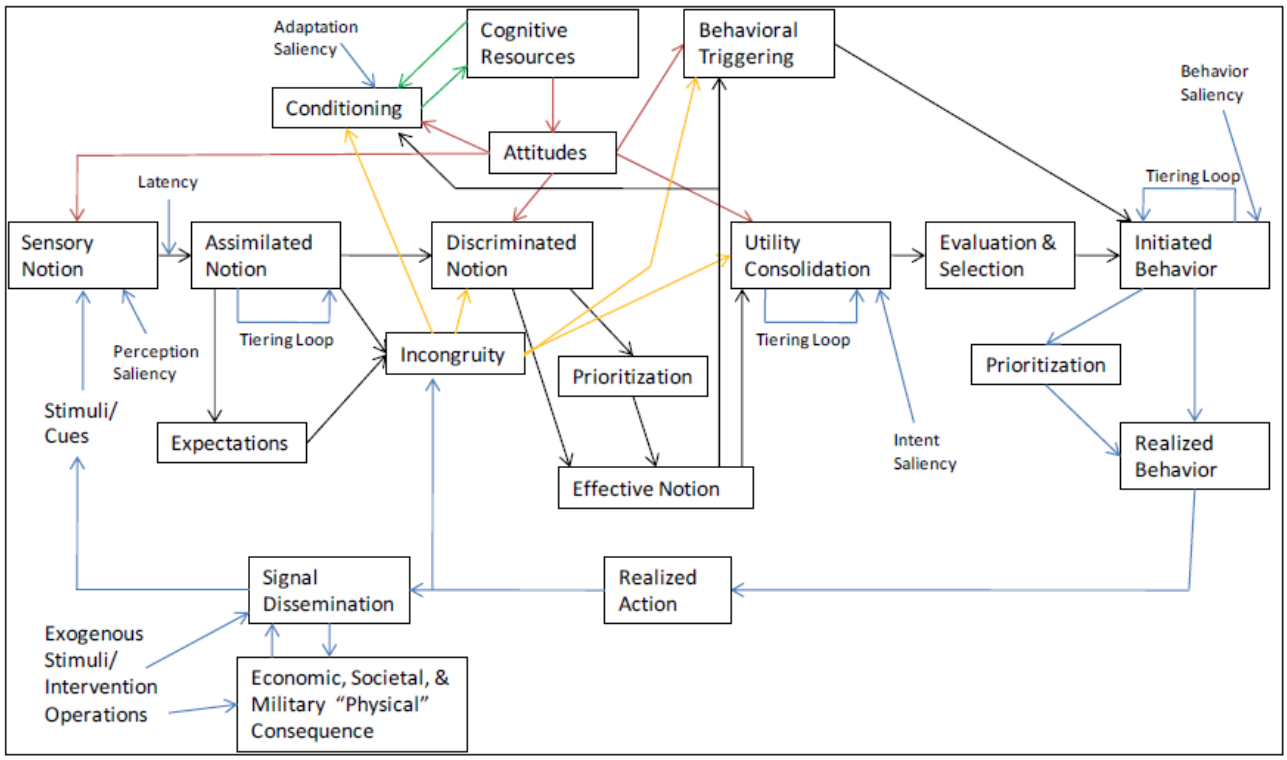


Figure 7: Detailed Model Cognitive components.

Related to notions is that of attitudes help determine the utility of choice:

> Specific notions (such as you realizing there is a fire in your house), can dramatically amplify your realization of other notion/cues such as the location of doors and other occupants of the house. Similarly, making one decision may affect your selection of a related decision. The same is true for executing behaviors. Attitudes affect the importance you may place on information. Attitudes are explicitly calculated in the model and are based on cognitive resources (experiences, abilities, and beliefs). Learning is noted as conditioning in the model and is an effort to reduce an incongruity by developing the ability to accommodate or effectively respond in the presence of a notion. Attitudes, emotive content, and cognitive information all act to determine the utility of a choice. These utilities come together to shape the probability of making a specific choice. Limitations in mental processing and physical response mean the individual must prioritize notions and behaviors when either becomes potentially excessive. For example, changing the radio station when you hear a song you dislike is quickly neglected when you see the car ahead of you hit another car. (Backus et al, 2010)

Another component, as noted above in choice is that of emotions, or moods:

> Moods are essentially lingering emotive notions. Altering conditions can cause moods to change over time, but typically not instantaneously. Therefore, decisions based solely on objective information may be different than those made in the presence of a specific mood (Rusting, Mellers). Because moods can arise quickly, decisions later in time, when moods have subsided, may be significantly different from those made when the individual is in a highly emotional state (Tiedens). The terms Saliency and Latency in Figure 7 note the parameterizations that capture the importance the individuals place on information (facts or feelings). The individual can adapt the importance he/she places on information as a result of conditioning and modified cognitive resources. This adaptation reflects itself as strengthened or weakened attitudes (Backus, 2010).

Moods affect decision making, in that it can affect the timing of decisions:

> Emotional conditioning can force snap decisions against one's own best-interest:
> The Assimilation process takes time with affective notion realization occurring faster than for reasoned notions. The rise in affective notion can exceed the threshold for a response and act to trigger behavior that might not otherwise occur if cognitive process dominated or timing were different. The lingering of an affective notion due the "afterimage" phenomenon of the assimilation delay in essence sets the mood of the entity. Because incongruity is the relationship between the assimilated notion and its expected values, the assimilation process can cause a delayed buildup of incongruity. The incongruity can reach a behavioral threshold some time after the actual initiating stimuli. (Backus, 2010, 82)

> A key feature is that some notions arrive sooner than others (such as affective ones), and if decisions need to be made promptly they may be different than what would occur without the time pressure. The assimilation process automatically capture the ideas of recency by remembering the previous events consistent with theory; it does not remember the duration and only remembers the maximum intensity. The notion takes time to die away and can affect future decisions. For affective notions, this response reflects the concept of moods. In the model, a discriminated notion is essentially a mood. (Backus, 2010, 83-4)

So that overall we see that the Influence machine involves creating a choice within an entity or agent or in reality a person or population of peoples (groups, tribes). This involves molding notions that are

acted upon with emotional reinforcement, and learning of behaviors connected to those notions and emotions that break previous expectations.

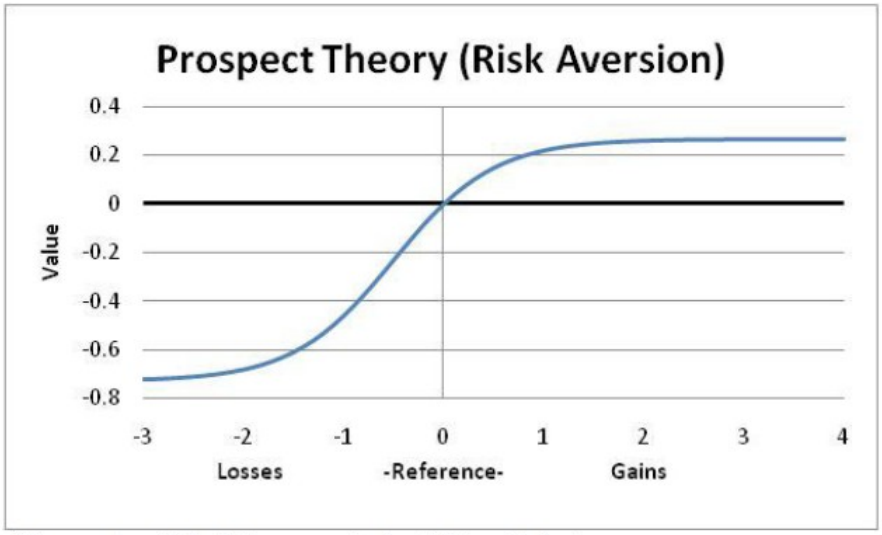


Figure 8: Model generated utility of choices.

Figure 8: Model generated utility of choices. Notions and expectations need to be significantly different from their normal values before an individual recognizes that incongruity as a “concern.” The level of discrepancy needs to evoke a response in line with “importance” assigned to information (notion). Because the perceived level of incongruity, intensified or diminished by attitude, contributes to the evaluation of choice, a perceived negative association may have a much larger impact on choice than an equal sized perceived benefit. This phenomenon is called loss or risk aversion and is exemplified in Prospect Theory or Risk Asymmetry (Backus et al, 2010)

Decisions are based on what is known as Qualitative Choice Thoery (QCT),

> The basic equation that weighs information, be it simply sensory input or the multifaceted utility of alternative decisions, is shown below. It is based on Qualitative Choice Theory (QCT) whose foundation comes from psychology (Luce) and from economics. The first term (numerator) on the right hand side of the equation is the relative value of the utility (U) for a collection of information.1 The second term (the denominator) is its comparison with all other relevant information. The result may be a choice (C), a simple recognition of sensory input, or an incongruity with a remembered condition and an existing condition. Subtleties of the equation reflect the probability that a choice is correct or useful in the context of perceived conditions.
>
> The first term (numerator) on the right hand side of the equation is the relative value of the utility (U) for a collection of information.1 The second term (the denominator) is its comparison with all other relevant information. The result may be a choice (C), a simple recognition of sensory input, or an incongruity with a remembered condition and an existing condition. Subtleties of the equation reflect the probability that a choice is correct or useful in the context of perceived conditions.

Choice Evaluation:

$$C(j) = e^{U(j)} / \sum_{i} e^{U(i)}$$

Equation 1.

The equation is also used for triggering learning and triggering behaviors by reflecting excitation and inhibitory responses. Conditions that are deemed too trivial to recognize or counter cause little excitatory reaction. Conversely, a condition may be so intense that sensory channels are saturated or that certain behaviors are too ineffective to execute. These situations cause extreme inhibitory effects.

The next equation below determines the pattern strength of cues forming notions or the cognitive resources forming attitudes. It is directly derivable from the choice equation above when the utility is proportion to the logarithm of the cues, such as with the Weber–Fechner law. As a specific manifestation of the choice equation, the notion (P), for example is a combination (z) of relevant cues (S). The β are the weights of each cue. and generally sum to unity. Pattern Strength:

Pattern Strength:

$$P = \alpha \times \prod_{z} S(z)^{\beta(z)}$$

Equation 2.

The next equation is an asymptotically exact, approximation of choice response when utility is based on a single consideration. It captures the incongruity between actual (perceived) conditions and expected (remembered or anticipated) conditions using an offset to avoid responding to insignificant discrepancies.

Incongruity:

$$D = \frac{Actual - Expected}{Expected} \pm Offset$$

Equation 3.

The last equation, below, is almost tautological in nature as the accumulation (R) of some quantity such as experience, memory, or capability that can atrophy over time (η) in the absence of continued activity (Input). This equation is used in simulating notion assimilation, cognitive resource conditioning, and expectation formation. Its theoretical basis is found in the “stock and flow” constructs developed in the field of System Dynamics (Sterman), but its statistical estimation and validation comes from the economics approach called Cointegration.

Conditioning and Fading activity:

$$R(t) = \int_{t_0}^{t} (Input(t) - R(t)/\tau) \times dt$$

Equation 4.

Analytics, Measuring Effectiveness:

In all computational systems, especially those with automation like AI itself, it is necessary to have analytics to measure the effects of actions on a target entity. The usual AI techniques are used, which it should be born in mind were not initially developed to model highly complex systems such as psychological persuasion as also used in neuromarketing (commercial warfare), not just information warfare. The interventions are modeled by probability of success:

> With this statistical knowledge, we can provide confidence intervals on the results of the model analyses that test interventions. By simultaneously performing uncertainty quantification for model parameters and potential interventions, the framework can determine the portfolio of interventions that have the highest (quantified) probability of success despite uncertainty. It can also quantify the risk associated with the intervention not performing as anticipated. Additionally, as will be discussed shortly, the framework can perform sensitivity analyses to determine what minimal additional information is needed to maximally reduce uncertainty and further assure the proposed interventions produce the desired outcome throughout the time horizon of interest. (Backus, 2010, 43)

Below is a diagram outlining the analytical engine:

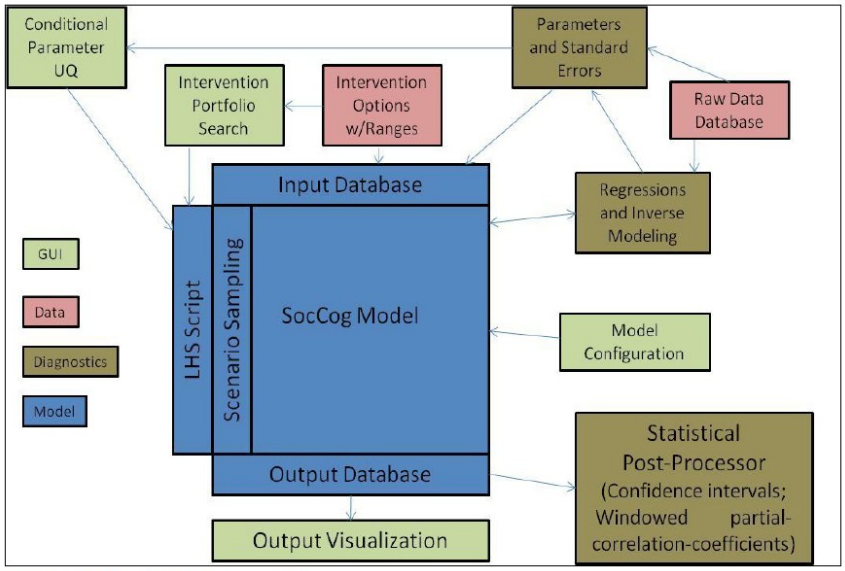


Figure 30: Analysis Framework

Finally, for software engineers we can see how the entire framework is put together by walking through the program code, originally written in COBOL, as given by Backus et al:

> Coding the Model of the Psych/Cog Engine:
> >
> **#Define Procedure Model**
> * If Lmax is gt 1, then we are doing UQ, SA, or search.
> * First run is always with best estimate values.
> Select Lrun(1-Lmax)
> * Start outer Stochastic (K+LHS) loop on
> Do Lrun
> * Determine values of model parameters.
> Do LHS (Algorithms developed be V&V team)
> Select Mrun(1-Rmax)
> Do Mrun

```
    * Select time range from Starting moment to Ending moment.
    * This allows restart from any point in a stored past simulation.
    Select Moment (SMoment-EMoment)
    * All differential equations need state variables initialized
      Do Initialize
      * Start march over time for simulation
        Do Moment
        * Current is the active moment in the loop
        Current=Moment:s
        Prior=XMAX(SMOMENT,Current-1)
        Next=XMIN(LMOMENT,Current+1)
        * (Social network and physical entities) Map physical and entity actions to all affected
entities
        Do Stimuli
        * Determine impact of entity action on external environment.
        Do External
        * Calculate Attitudes based on cognitive resources
        Do Attitudes
       * Calculate passivity and Offsets based on cognitive resources
      Do Passivity
      * Calculate Dissonance between current notions and expectation of    conditions
      Do Incongruity
      * Map patterns of Cue Stimuli into Notions
      Do Notion
      * Create memory of notions for expectations of “normal” values
      Do Expectation
      * Reinforce referent memory intensity based on schema and perceptions
      Do CogRes
      * Decide choice intent
      Do EvalSel
      * Affect behavior based on Referents (Norms) and dissonance
      Do Behavior
      * Transform Behavior in physical consequence via the associated   physical action
      Do Action
      * Map action of entities to external environment (Social Network)
      Do Stimuli
    End Moment
  End Mrun
End Lrun
End Procedure Model
```
(Backus, 2010, 73)

As we can see the framework can be run again and again iterating over a specific intent, choice, trying to persuade entities to select a particular behavior. The SNL of Sandia is also applicable to other forms of simulation and can be added to other areas of research such as military simulations, or geo-political simulations.

Simulated Soldiering

We covered military simulations architecture in Appendix D. Here we drill down further into the ideals in the defense industry regarding simulations and automated control of soldiers, which was research begun in the Soviet Union using Reflexive Management. The research of Behzadan (2017)

focuses on using Sandia Labs technology to create military simulations using Agent-Based-Models (ABM). The work is designed as a means of destabilization, again like Tarashenko's research from earlier.

> Terrorist organizations have social networks that enable them to recruit and operate around the world. This paper presents a novel computational framework for derivation of optimal destabilization strategies against dynamic social networks of terrorists. We develop a game-theoretic model to capture the distributed and complex dynamics of terrorist organizations, and introduce a technique for estimation of such dynamics from incomplete snapshots of target networks. Furthermore, we propose a mechanism for devising the optimal sequence of actions that drive the internal dynamics of targeted organizations towards an arbitrary state of instability. The performance of this framework is evaluated on a model of the Al-Qaeda network in 2001, verifying the efficacy of our proposals for counter-terrorism applications. (Behzadan, 2017)

They study the microeconomics of terrorist networks as Agent-Based Models (ABM), thereby allowing the analysis and potentially control of terrorist organizations by considering the micro-scale models of such systems, this control is made to destabilize the group under pressure.

> The proposed framework consists of a dynamics estimation method for inference of payoffs from a game theoretic model of the network, complemented with a technique based on reinforcement learning for
> derivation of optimal action policies. The main contributions of this paper are:
> (i) We present a game-theoretic model that captures the complex self-organizing dynamics of terrorist organizations in the settings of strategic network formation.
> (ii) We introduce a technique for the estimation of network dynamics based on incomplete snapshot observations.
> (iii) By adopting a reinforcement learning approach, we develop a mechanism for devising the optimal sequence of actions that drive the internal dynamics of targeted organizations towards an arbitrary state of instability.
> (iv) We propose a methodology for extraction of personal and relational attributes from unstructured text using cloudbased services. (Behzadan, 2017)

The main device for destabilization is based on game-theoretic techniques, which is to say deceptive games or reflexive control. The first target of destabilization is by minimization of connectivity in the network. "Such strategies seek the set of nodes or links whose removal from the network maximizes the number of isolated groups (known as connected components). Approaches for selection of critical nodes or edges in this type of strategy are largely based on graph centrality metrics (Behzadan, 2017)." They adopt inverse game theory:

> With the potential of algorithms developed for inverse game theory, the problem of destabilization can be approached with enhanced confidence in the accuracy of the network model, at the expense of further complicating the problem due to the increased complexity in the dynamic network model. Manipulation of such highly complex networks of autonomous agents under nonlinear dynamics is the subject of a novel area in complexity science, known as guided self-organization. This area investigates the controllability of self-organizing systems, characterized by the emergence of order and pattern from uncoordinated actions of autonomous agents. The necessity of guidance in such systems appears when there is a need to hasten or perturb the natural evolution of the system towards a desired state. (Behzadan, 2017)

So what we see is an automated algorithm that allows for a group to become destabilized, the optimal control inputs that perturb the complex dynamic network toward a state or goal is considered a promising method by them. “In this application, the desired state is the goal of destabilization, and control inputs are exerted in the form of perturbations on terrorist network’s structure and topology (Behzadan, 2017). The algorithm is presented by Behazadan in the following:

> We adopt a dynamic systems viewpoint to interpret this objective, where destabilization refers to driving the target system away from undesired steady-states (equilibria). In the remainder, we consider a further level of detail by specifying the objective as minimizing the desire of terrorist agents to remain affiliated to the organization. It must be noted that this choice of objective will only serve as a representative example, and may be replaced by any arbitrary goal in the presented framework. Having determined the criteria of objective, the problem of optimal destabilization can be formally stated as follows:
>
> $$\pi_O^*(G(t))$$
>
> Given the observations of relational links in a terrorist network:
> devise the optimal policy function (G(t)) that for any observed network G(t) determines the counter-terrorism action with maximum cumulative (i.e. long-term) reward R according to an objective O. The notion of cumulative reward in this problem formulation necessitates the prediction of future states that emerge in the target network as a result of the actions implemented by the counter-terrorism entity, which is dependent on the dynamics of the target. The problem can be seen as comprising of two sub-tasks: (i) estimation of target network’s dynamics from (noisy) observations, and (ii) determination of optimal actions given the estimated dynamics.(Backus et al, 2006)

The question of how to measure the dynamics of the complex networks is addressed, through a utility function of each agent:

> Having the game theoretic model of network’s dynamics, the problem of estimation can be mapped to the domain of inverse game theory. Assuming that the observed network is at equilibrium (i.e. the topology is not changing), the objective of this problem thus becomes to estimate the utility functions of each agent such that the resulting equilibrium matches the observed topology. The majority of approaches proposed for utility estimation from network topology observations require multiple observations of changes in network over time. Yet this is commonly not feasible for terrorist networks due to their covert nature. Approaches proposed for estimating utilities from a single network observation are also mostly developed for network formation games of complete information. (Backus et al, 2006)

From the estimation of agent utility one can then start to formulate an optimal policy of destabilization:

> The estimated dynamics of the terrorist network provides the opportunity to simulate the responses of targeted organization to counter-terrorism actions, thereby enabling the employment of exploratory methods for determination of the optimal policy.
> Accordingly, we propose Reinforcement Learning algorithms as a promising approach to the problem of Policy Optimization. Reinforcement learning techniques are described by the Markov Decision Process tuple MDP = (S; A; P;R), where S is the set of reachable states in the process, A is the set of available actions, R is the mapping of transitions to the immediate reward, and P represents the transition probabilities (i.e. system dynamics). At any given time-step t, the Markov decision process is at a state st 2 S, which can represent the

current topology of the network. The reinforcement learning agent's choice of action at time t, at 2 A causes a transition from st to a state st+1 according to the transition probability Pat st;st+a . The agent receives a reward rt = R(st; at) 2 R for choosing the action at at state st. Interactions of the agent with Markov decision process are captured in a policy . When such interactions are deterministic, the policy : S ! A is a mapping between the states and their corresponding actions. A stochastic policy (s; a) represents the probability of optimality for action a at state s.

Algorithm 1: Optimal Destabilization Algorithm

Input : observed initial topology G_0, set of profile vectors X, objective O, set of actions A
Data: reward R, current topology G, policy π
Output: action $a = \pi(G)$

```
R ← 0
G ← G_0
while R < O do
    U ← EstimateDynamics(G, X)
    R, π ← QLearning(SimulateDynamics(.), G, U, X)
    Implement a ← π(G)
    Update G
end
```

The selected action is then implemented in the target simulation, which responds by rearranging the topology towards the best achievable state of equilibrium. The reinforcement learning process observes and measures the resulting state according to the desired metric of stability, which provides a form of feedback for the selected action, enabling the Qlearning [AI reinforcement learning] process to learn the efficiency of its choice and adjust its future actions accordingly. The iterations of this cycle continues until consecutive measurements of the simulated target's state satisfy the intended criteria of instability, at which time the learned policy is declared to be optimal and ready for implementation on the real target network (Backus et al, 2006)

From which we now have the optimal destabilization, but it comes with limits, not to mention MDP bias issues, in that estimations tend to be inaccurate, due to available information on target networks. Enhancing the accuracy of these methods may be achieved through an iterative update of observations. Once an optimal policy is obtained, the result of its implementation on the target network can be utilized as a new observation of the network. Consequently, running the dynamic estimator on the new observation can increase its precision through fine-tuning on new data points, which will lead to enhanced accuracy in the derivation of optimal policy. Algorithm 1 details this iterative process (Backus 2006). This work has offered us a look into the ideal of destabilization of terrorist networks and networks in general, and of course the inverse settings allow for stabilization or creation of networks.

Umbra and SCREAM

The initial steps to setting up automated systems that are based on Deception Management and US Cost of Actions (COA) (Reflexive Control) is to develop models that can be enacted by AI systems. One of the unique attributes of the US Mil-Intel Simulations is the creation of Emotional Agents to simulate real world figures. To trace out the development how deception and influencing work we need to start with the issue of technologies developed by Lockheed-Martin, sometimes through their 'national' private run government labs. We can look at training and simulations developed by their research teams to understand how AI agents can be used as emotional agents to train humans from warfare to disaster relief. In Xavier et al (2017) we can look at their architecture to better understand how in a later section how emotions are used for deception management and influence operations using emotional agents. Dante Agent Architecture is the framework developed by the research team, in an effort to provide simulations for typical Red vs. Blue exercises dealing with physical security, such as Nuclear sites (energy and weapons), for which this study was generated. Dante is an extension of Lockheed's Umbra software package the provides simulation capabilities across varying platforms and modules also developed by Lockheed as plug-and-play modular components of varying functionalities within National Defense, do to compartmentalization in covert programs each silo of the covert program creates an API so that other groups can easily use their computer engineering in their project. As covered in the Automation and Simulations chapter this system uses Finite State Machines (FSM) which manage behaviors, while the Decision Agents are based on utility theory.

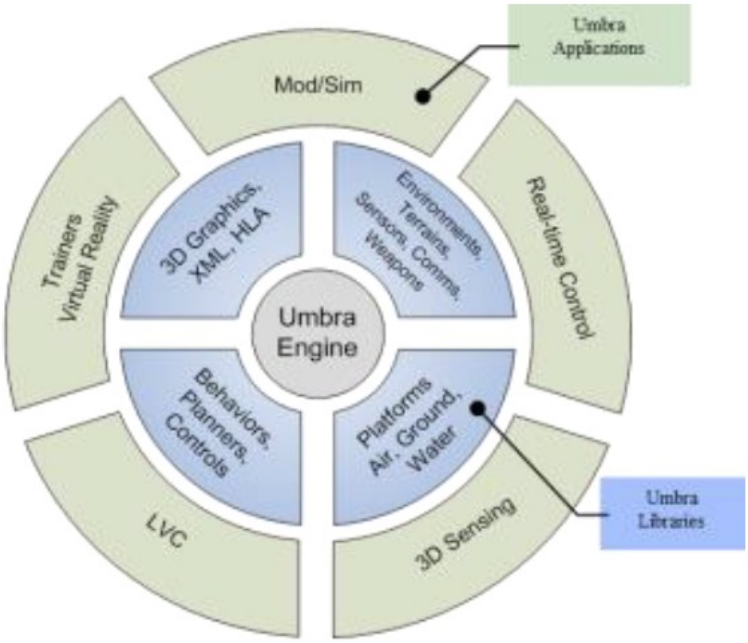


Figure 1: Umbra Simulation Framework & Applications

When a character is issued a command, they execute a priority queue of behaviors. The queue is sorted on the priority and activation of the behaviors available, calculated through a utility-theory AI approach. Behaviors can run concurrently or preempt other behaviors to take control of the character. Later, preempted behaviors can be restarted to accomplish the specified command. (Xavier et al, 2017)

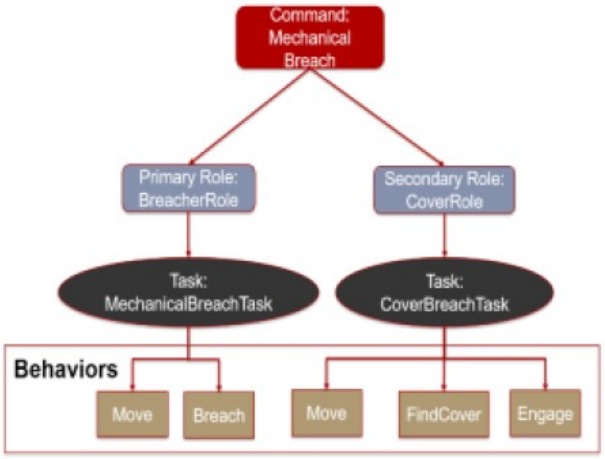


Figure 3: Example of MechanicalBreach command issued to team

Within the framework the usage of different functions is developed Roles which translates a Command into behaviors, these roles allow characters to differentiate how they react to the same Command. A Task represents a collection of behaviors (FSMs). In addition to Roles, Tasks, etc. there are Triggers, which determine if a behavior should run. The Activation of a Behavior, A is represented as, A = T * p, T represent the set of triggers associated with a behavior, and p representing the priority for the behavior. Triggers may have p ositive evidence or negative evidence (inverse) for a behavior to run. The Triggers are stored in a 'blackboard' is regular video games, a datastructure, character perception produces many triggers including ones that note the presence of threats, the trigger values are not binary rather a floating point number between 0-1, which is used for nuancing. Triggers are stored into the TriggerManager which is queried when a new threat is detected the CharacterMemory, then a behavior is triggered. These weighted triggers exist on a priority queue which can result in behavior jitter as competing perceptions affect the weighting and re-organization of the behavior list. Path planning is based on A*. One interesting component of the system is the Scenario Editor which allows the importing of various scenarios in Red. vs. Blue training. Users construct an agent's team plan by chaining together commands, with the completion of one command leading to the execution of the next, which can also be automated for different scenarios. One issue to deal with in all software development are exceptions and failures. If there is no branching or exception handling then the simulation would become stuck. Dante uses a fail forward mechanism to address failure:

In the case when a command has failed, the RunManager will look for any chained commands that should be called on failure. This mechanism allows *Dante* scenario to branch when things are not going according to the original plan. There is a special mechanism in *Dante* called "fail forward." If there is not a failure branch for a failed command to follow, then it still signals the success branch actions. This allows the plan to continue forward, but likely it will encounter further problems that eventually result in a Terminate action. "Fail forward" aids users in not having to provide explicit detail how agents should respond if any portion of the plan fails, and safeguard against faulty user-defined plans where failed commands may not be necessary in accomplishing the scenario goal(s).

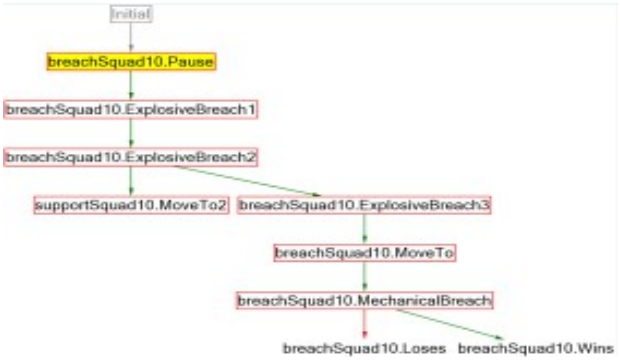


Figure 7: Example of a plan formed within Dante Scenario Editor Figure 7 provides a snapshot from a *Dante* Scenario Editor displaying a red-team plan for breaking into a building within a secured facility, requiring the team to demolish multiple barriers between their start point and the building door. (Xavier et al, 2017)

In a more tangible example of using emotional based agents in training can be found in Emergency Response training using gaming in a simulation. As previously discussed in the Chapter 4 'Lessons from an American Weapons Developer' Sandia National Labs (Lockheed-Martin) has built a cognitive computing framework to load agents with emotional behaviors. The framework, Sandia Cognitive Runtime Engine with Active Memory (SCREAM) can be extended upon and a Sandia research team has expanded on that with the Sandia Human Embodiment and Representation Cognitive Architecture (SHERCA). In Djordjevich of Sandia Livermore Labs has done research using SHERCA in emergency response training. The reason for studying emotions in emergency response is that according to the researchers 34% of decisions are derived from emotions when utility cannot be seen (Djordjevich, 2008). It is also noted that emotions can override reason and vice versa in later studies in this chapter, although people respond to emotional stimuli faster then reason. "A large amount of research supports the notion that attitudes, norms, emotions, goals, and the perception of control helps drive actual behaviors. In fact, the theories that support this research have been successfully used to predict a wide range of behaviors, such as voting, shoplifting, gun-related violent acts, and other moral and ethical decisions" (Bernard et al, 2006, 21). Again we find the Theory of Planned Behavior referenced here which is behaviors are influenced by 1) attitudes towards a specific behavior, 2) the subjective norms associated with acting out that behavior, and 3) the perception that this behavior is within a person's control, forming an action intention state. (Djordjevich, 2008). The framework is based on emotions which distinguishes it from other cognitive architectures like ACT-R and SOAR, it is interesting to note that Lockheed-Martin based SCREAM on SOAR and that Dr. John Norseen cited SOAR in his work. "As noted in a prior section, emotions tie into a person's motivation for deciding what actions to pursue. To computationally represent the role of emotion in motivation, researchers have found mapping perceptions to pre-defined emotional states as an effective method for virtual [cognitive] characters [NPCs] in these environments" (Djordjevich, 2008). Some aspects of the SCREAM framework that supports the psychological framework, SHERCA, is that emphasis on

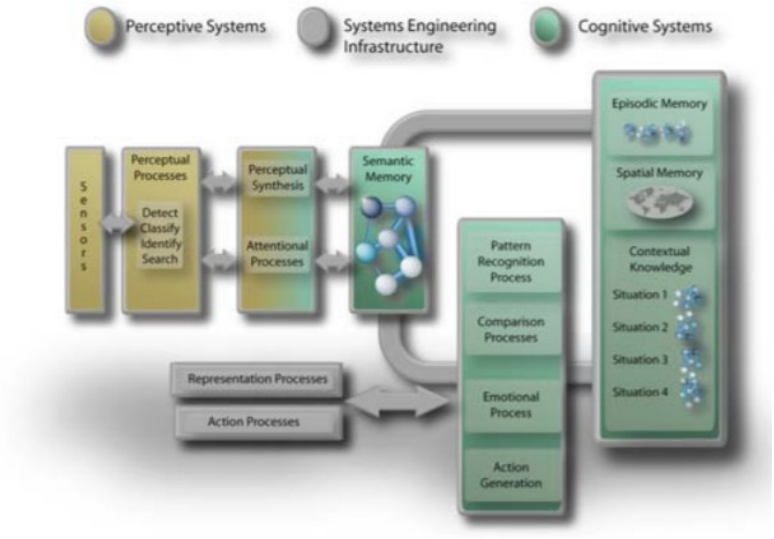


Figure 2. A high level view of the SNL cognitive modeling architecture

cognitive elements with activation-levels updated according to dynamics distinguishes it from more common production-rule-based approaches. Some other uses of SHERCA was to use it in cultural training, it can adapt various cultures to change nuance of different cultural imprinted agents. “in SHERCA, an impression of culture can be generated by varying a simulated human’s emotional response to particular perceptions. Cultures also exhibit variations within their high-level and intermediate goals. As a result their intended and actual behaviors will show cultural uniqueness. The result is a complex set of behaviors that have certain emergent properties common to a particular group. [also see Memetic Engineering below] So how is SHERCA put together? The research team offers this overview:

> SHERCA allows for multiple cues, cognitive perceptions, goals, action intentions, etc., to concurrently have some degree of activation. In SHERCA’s model of decision-making, once a cognitive perception—an element of perceptual/ situational awareness—has been activated by cues in the environment, it may trigger activation of specific, intermediate goals that are consistent with higher-level goals and other active cognitive perceptions. For example, one high-level goal might be to protect family, and another, to protect oneself. Intermediate goals help support the higher-level goals by breaking down the goals into discrete tasks. The overall emotional state mediates activation of action intentions from the intermediate goals. As a consequence, the intended actions are a product of both the intermediate goals and the current emotional state of the simulated human. This emotional state may change dynamically, for example from very low to very high levels of anger, if the perceptions change. Action intentions that are contradictory with respect to goals can become concurrently highly activated due to the influence of emotion. At the same time, cognitive perception is influenced by a hierarchy of higher-level goals/directives or moral states, as well as state within a behavior (e.g., current step in a procedure). (Djordjevich, 2008)

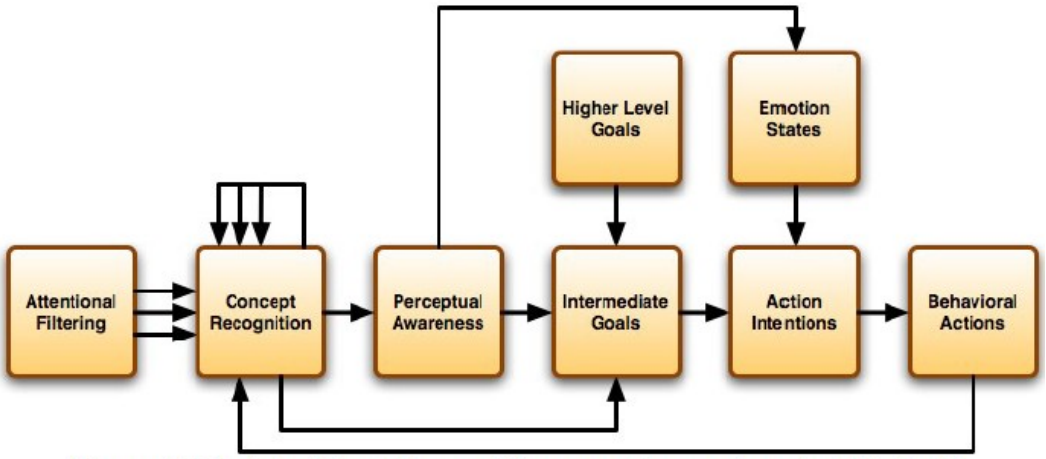


Figure 5 Model of decision-making to select actions in SHERCA.

The main elements in SHERCA are Concepts, Contexts, Cues (which are common to many cognitive models), also there are Intermediate Goals, Higher Level Goals, Perceptions, Action Intentions and Behavioral Actions. Currently we need to take a deeper look at Concepts and Contexts.

Concepts:

- Fundamental element of SNL Cognitive framework a representation of known regularity arising from external or internal

Contexts:

- meaningful perceptual representation stimuli
- Theta update rate (4-7Hz)

data sources
- High Alpha (10-13Hz update rate)
- each concept is represented by a neural assembly and the neural assembly associated with a concept that is 'currently used' in cognition will rise above baseline in terms of both frequency and amplitude.
- A concept, is a cue, when it plays a role in the recognition.
- Sorted b Instances: associate objects with concepts. Concept Instances map the slots of a concept to entities (objects)
- Context Instances
- Context Recognizer pattern recognition with:
  A. cue vector context patterns
  B. XQ context patterns

A bit of an explainer on a couple terms for Contexts. Cue Vector Context Patterns which are templates for context instances, instances of different concepts. This is represented as:

Q = $\{Q_i\}$ of triple $Q_i = P_i, w_i, m_i\}$ where P = concept w/ a name; w = weight; m=integer vector

Implementation example:

> Suppose a cognitive model has the unary (i.e., arity one) concepts *dominant* and *submissive* and a binary concept *attacks*. Then we can define a binary context *Dominated-by* that has the pattern *Dominated-by 3 submissive 0.4 0 attacks 0.6 1 0 dominant 0.5 1*
>
> whose vector of cues
>
> *Q*(Dominated-by) = { *(submissive, 0.4, {0}), (attacks, 0.6, {1, 0}), (dominant, 0.5, {1})* }
>
> In our notation, index numbering for elements of a vector begins with 0. The *{0}* in (*submissive, 0.4, {0})* maps slot 0 of *submissive* to slot 0 of *Dominated-by*. The *{1, 0}* in (*attacks, 0.6, {1, 0})* maps slot 0 of *attacks* to slot 1 of *Dominated-by* and slot 1 of *attacks* to slot 0 of *Dominated-by*. Finally, the *{1}* in *(dominant, 0.5, {1})* maps slot 0 of *dominant* to slot 1 of *Dominated-by*.

From this datastructure we can get a picture of how this system works in terms of dominance and submission, in a binary confrontation, a fuller definition of concepts and contexts are below:

> A ***concept*** is the fundamental semantic element in SCREAM. For convenience, concepts have names, but SCREAM associates no meaning with those names. To function in environments with multiple entities (e.g., things, creatures, features, etc.) of a given type requires a mechanism to associate concept activations with specific entities. SCREAM takes the simple approach of endowing concepts with slots. A ***concept instance*** associates slots with entities. Concept instances are created as needed. Each has its own activation state and is uniquely identified by concept and a vector of entity identifiers, which are merely labels to enable convenient interaction with people and other non-SCREAM system components. For example, *chases {22, 31}* identifies an instance of the concept *chases* (i.e., entity #22 chases entity #31). Thus, a concept is similar to a fuzzy predicate, but we do not claim that SCREAM implements any logic. (Djordjevich, 2008)

> ***Contexts*** can be defined as meaningful perceptual representations that are based on recognizable patterns of stimuli, as well as, consistent with situation models, schema and theme-based representations of events. Context activation is governed by pattern recognition applied to the activation states of concepts that are the cues for/against that context. For example, the concepts *bicycle, clown, elephant* and *popcorn* might be cues for the context *Circus*. A ***context instance*** is related to a ***context*** similar to the way that a *concept instance* is related to a *concept*. In SCREAM, a concept whose raw (input) activation is driven by the contextual pattern recognition process is also called a context. (Djordjevich, 2008) A context is used to describe the 'mental' processes that occur when we try to make sense of, and interact with, our environment, these processes can be broken down into perceptional context or 'perception' and 'goal' states (Bernard et al, 2006, 35).

A further element of the SCREAM system is that of using emotions. Although there is not many emotions included in SCREAM in early implementations there is only fear and anger mapped to the system, which are the prime motivators in quick/snap decisions.

> A basic capability for modeling emotional processes in cognition [3], [24] has been implemented in SCREAM. SCREAM updates the level of activation of each emotion based on concept activation levels and parameters that specify how the concepts influence emotional state. Each concept can be associated with a level of activation and a weight coefficient for each emotion. For example, in an emotion parameters file,
> *cee clown 2 fear 0.6 0.7 anger 1.5 0.9*
> specifies that concept *clown* influences two emotions. For *fear* it has a weight coefficient of 0.6 and a target activation of 0.7, and for *anger* it has a weight coefficient of 1.5 and a target activation of 0.9. [amygdala hacking with fear and anger] (Djordjevich, 2008)

As an example of the entire SCREAM and SHERCA systems working together for training of emergency response personnel, the developers came up with the game 'Ground Truth'

To see how the program works lets take a brief dive into it's class hierarchy and how different agent classes interact with each other. Each cognitive agent is an instance of the ScreamAgent Class which extends other base classes at higher levels which allows it to carry out the cognitively-selected behavior. When a cog agent loads it inputs the cognitive model definition files such as the concepts, contexts and context patterns, context-instance to behavior/action conversion patterns, spreading activation (priming) and emotional association parameters. The only difference between cognitive definition files is limited to emotional association parameters and levels of activation of high-level goals, reflecting differences in personality, culture and values. Setting activation of other specific concepts appropriately allows us to customize the generic model for each specific type of GroundTruth NPC. As seen earlier in the Serious Games chapter there are Game managers for different aspect of any game. The AgentManager interacts with other game Managers, updating the NPCs is the main role of the AgentManager, which updates the states of the agents in an update cycle: physical states, based on their current states and game state external to the agents. It has the agents update their perceptual states, with the help from the Perception Manager and drives agent decisions.

> States in the state machine of a cognitive agent can access its emotional state for use in modeling affect within a behavior. For example, dialogue output takes into account emotional state. Generally, determination of low-level behavior, such as path planning, also makes use of separate algorithms that are called from within states. (Djordjevich, 2008)

A deeper look at SCREAM will show us the mechanics of Serious Games and Military Simulations as handled by the UKUSA defense industry. Sandia Labs has a fuller explanation of their cognitive computational systems in Bernard et al 2006 *'Simulating Human Behavior for National Security Human Interactions'.* The developers remark regarding the purpose of SCREAM: "to allow cognitively modeled simulated humans or 'cognitive characters' to interact with each other, their environment and with actual humans in a behaviorally realistic and psychologically plausible manner." (Bernard et al, 2006, 9)

SCREAM:

1. it produces a psychologically and sociologically reasonable computational framework of human behavior.

2. cognitive models are customisable individually

3. cognitive characters operate autonomously

4. designed to plug-n-play with other MIL apps

The main purpose behind SCREAM is to provide realistic NPCs in military simulations, in disaster training simulations, etc. Some of these simulations run in virtual reality and through normative video game interfaces mixing human players with NPCs, some of whom can be aides/team members and some enemies as is normative in contemporary video games. The platform was created to address shortcomings in commercial video games, specifically dealing with realistic NPCs in video games. The NPCs known as cognitive characters in SCREAM can behave in amazingly realistic ways for humans, and possibly even pass Turing tests. For instance they can respond with fear at a threat and hide, they can manifest aggression depending on their role within the platform and other complex emotional responses usually shown by real humans. An example from the framework even includes that when a Cog Character is nervous it seeks a person to talk to:

WIFE OF VENDER ONE - INTERMEDATE GOAL STATES
SG3 I seek to look innocent/friendly
SG4 I seek to sell my merchandise, but I am nervous
SG5 I seek to converse with a nearby person

(Bernard et al, 2006, 51, Appendix C)

And in this way through their scripted responses they seem very human but may not pass the Turing test based on the observer's knowledge and expertise.

HI Framewok Modules:

1. Semantic Knowledge – associative network with nodes, representing each critical concept or schema in each cognitive character

2. Pattern Recognition & Comparator- a. evaluating the evidence provided by cures favoring or conflicting with each situation. b. assessing the validity of the current situation. c. determination of a valid situation when current situation invalid, d. implementation of top-down activation-levels

3. Action Generation

Some important aspects to realize about SCREAM are that it is not rule-based, as in some older systems from the mid-90s such as the British Mannequin system for monitoring the northern Irish,

rather uses models of human decision making with levels of activation for perceptions and states. It also allow for multiple perceptions, goal states and action-intentions which can concurrently have some degree of activation. Perceptions are activated by cues which activate Intermediate Goals, the Intermediate Goals will trigger an action-intention state, which are mediated by current emotion states (i.e. fear/anger) that are affected by what they perceive in the environment. The decision to act or call on a behavior is in the cognitive subsystem which serves in the model as the conjunction of diverse emotions, stressors, memories which are all integrated into a decision for action which should lead to the goals.

> When constructing a model, each concept may be attributed to one or more emotional component that is associated with specific levels of activation. For example, a disliked individual may be represented as a concept for which there is an association with a high level frustration-anger. Activation of a specific concept or situation contributes to the weighted averages that determine the overall activations of associated emotional components. The specific emotion activation levels are converted to fuzzy set (e.g., *high-fear*) representations that are then feed into context recognition patterns. In the near future, emotions will have a reciprocal effect on cognition, causing an increase in concept or situation activation that triggers the emotion and active inhibition of other concepts and situations. In the future, as with the current interactions, emotion and cognition will be consistent with neuropsychological findings and, thus, will allow certain neuropsychological phenomenon to be demonstrated by the framework. (Bernard et al, 2006, 13)

3 High Level Components:

a. pre-cognitive (attention)

b. cognitive (perception, states, goals)

c. action-generation states (motion control) [movement is managed by Boston Dynamics DI Guy, it is also important to note that not all human movements, such as micro-expressions are not included in movement, such as the forehead not being mapped by facial landmark detection libraries such as d-lib, while the simulation of human behavior and emotion is effective they are still lacking absolute simulation of human elements]

Semantic Memory System:

Knowledge and the relatedness of concepts is represented within a semantic memory system. Relatedness refers to the awareness that two concepts are associated with each other by virtue of being members of the same category and operating together. When concepts are extracted via automated knowledge capture techniques, the relationship between concepts is based on a representation of each concept as a vector in a high-dimensional space. The cosine similarity of vectors for a given pair of concepts provides the basis for the strength of the relationship between concepts. The relatedness of concepts derived in this manner then provides a basis for simulating the priming that occurs for a concept in response to the prior presentation of a related concept (Bernard et al, 2006, 14)

Certain concepts occur with with perceptual contexts. These contextual percepts play a part in the formation of expectations, seeing a hostile context influences recognizing further elaborations in that context, priming an expectation in hostility for say ‘fighting’. If the existing pattern of concepts is on the

verge of triggering an activation then only a minor amount of evidence that hostile people are in the area is needed to prime the expectation. It also might be that very strong evidence from only a small number of concepts is enough to activate a context. Once a context is activated, the associated concepts are primed so that corresponding sensory inputs produce accelerated and supplemented activations. [Expectation data poison attacks: consciously interfering with the systems expectations to defuse a situation on your way to an attack vector]

In what is similar to typical video game experiences dealing with horror games, the creepy cues that can be accompanied in the game experience of the player are mimicked in SCREAM. For instance, a certain context can prime humans to continuously attend to specific stimuli in the environment, so long as that context is activated. Being in a dangerous area will prime a person to hear footsteps, or see shadows. In this framework certain contexts may heighten the activation of specific cues in the environment. This occurs when contexts with salient cues that indicate danger to the cognitive character. The prime factor in fear response is the amygdala. In fact the designers of this system do target the amygdala complex, neurophysiology, with their framework:

> Research has shown that early in stimulus processing, the fear/surprise emotional centers of the brain (amygdala) receive direct input about the potential significance of a stimulus. The "direct route" conveys a fast, rough impression of the situation because it uses a sub-cortical pathway in which no high-level cognition is involved. At the same time, stimulus information is processed via another information pathway to allow for a deeper, cognitive assessment of this information. This "indirect route" allows for more deliberate assessments of the situation. The multiple pathways enable both an initial, fast response as well as the integration of emotion with higher-level recognition and understanding. The memory process output is appraised for its emotional meaning through which a behavioral assessment is made. (Bernard et al, 2006, 22)

> As with humans, cognitive models should incorporate the ability to react quickly to certain stimuli without the need first to deliberate the degree of threat. In the HI framework, this psychological phenomenon was modeled such that the cognitive characters exhibit defensive reactions of fear in response to perceived stimuli that represent potential dangers and/or anger towards certain perceived events or actions. Specifically, the cognitive characters have responses representing the direct route, in that it essentially represents the reflexive thalamo-amygdala pathway that bypasses the cortex. This pathway exhibits fear and defensive reactions, but not other emotions such as happiness. Other emotions and aspects of the indirect route will be modeled in the very near future. (Bernard et al, 2006, 23)

As noted before on Soviet Reflexive Control the amygdala is the pathway of Fast Reflexion, which is to say unconscious control. It is also noted that political conservatives as they are called, have a variance in their right amygdala, which is much larger then moderate and liberals, hence messing with their amygdalas creates unforeseen complexities, luckily this is just a virtual representation of human interactions and does not antagonize a real humans amygdala responses which could create a large problem if ever attempted, although the phenomena of Targeted Individuals suggests that these systems have been applied to non-virtual real human agents.

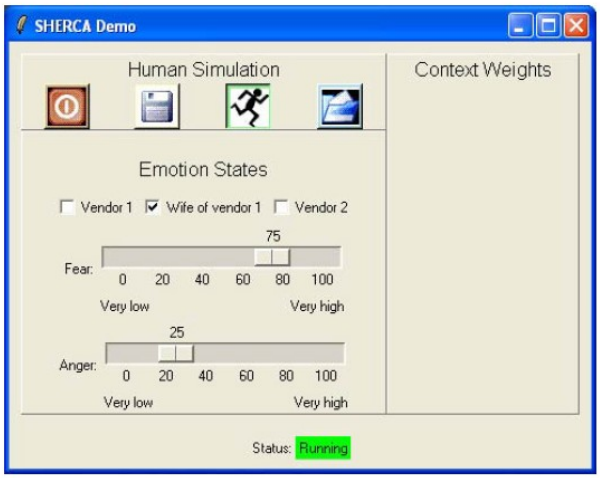


Figure 4. The Emotion Regulator GUI

(Bernard et al, 2006, 25)

Example from emotional parameter data file for emotions:

> cee clown 2 fear 0.700 .7 anger 0.100 .9
> //It tends to push the activation level of *fear* to 0.700 with a weight of 0.7 times the activation level of *clown* and the level of *anger* to 0.100 with a weight of 0.9 times the activation level of *clown*.

SCREAM continuously updates the level of activation of each emotion based on a concept activation levels. Concept is directly activated by perception as a result of semantic processing, a concept may be specified to influences emotional state, for instance using narrative networks or memetic learning, popular memes. The activation level of the emotion anger can be represented by membership levels in the three fuzzy sets lowAnger, medAnger, highAnger that cover the range of anger activation levels, these are used as cues in context (pattern) recognition. SCREAM's emotional state computational capability to model the direct route emotions of fear and anger.

How to Instantiate and Initialize Cognitive Characters in SCREAM:

To create a Cognitive Character initialize the stimuli and contextual associations (perceptions, activations and states)-- by first cataloging a series of low level environment cues the Cog. Chr. might perceive in a given scenario. Concepts prime concepts, priming specified by series of cue-to-cue relationships with a certain degree of priming. Cue1 primes Cue2 with weight=.8.As well perception-to-cue-priming operates similarly. Activated Perception primes potential intermediate Goals which become activated if environment situation matches. Intermediate Goal state is associated with environment cues and its level of cue each is associated with a hierarchy of increasingly abstract-level goals. When subject is activated it activates on action-intention state, which bind emotional affect with behaviors. SCREAM uses Active State Machine activated by action scripts consisting of a coded sequence of behaviors. After the behaviors are played out the character will reassess [hive huddling] the environment and follow new behaviors based on new perceptions. Situations not recognized activate additional contexts based on past experiences induced from earlier cues, rather than actual evidence in the current situation, ignores it's current reality.

As we can see from the previous SCREAM and SHERCA provide a robust and almost real experience with modeling agents after humans. Some other areas that Sandia is focused on for these human like simulations is that of economics:

> In addition, the LDRD project, *Cognitive Modeling of Human Behaviors within Socio- Economic Systems* (06-1102), is working to scale the number of cognitive characters to at least 10,000 entities, along with modeling economic, cultural, and stress-induced behaviors by FY2008. The goal of this project is to develop a science-based cognitive modeling framework of the individual-level economic decision-making that is critical to national economic security. Specifically, this project is developing a defensible neuroeconomic and cognitive science-based model of economic decision making before, during, and after "extreme events" such as acts of terrorism or natural disasters. By expanding the current state-of-the-art in modeling and simulating them in large-scale computing clusters SNL is working to produce high-fidelity, internally consistent analysis of these types of events on the economy and public confidence. (Bernard et al, 2006, 44)

So we have an understanding how to cognitively model agents in simulations, next we shall cover the automation of troops, which may also include using the cognitive software from above.

Automated Troops

As mentioned before SCREAM was based on SOAR, in the creation of automated troops or forces SOAR has also been used, which leads to the conclusion what SOAR can do so to can SCREAM/SHERCA. The need to make up for loss manpower has prompted the creation of such things as automated forces. In the following we review the work of Whetzel et al (2010) in regards to the programming of automated forces, which of course could be coupled to actual troops using a non-invasive BCI.

> **Semi-Automated Forces (SAF)** SAFs address the need for reduced-manpower simulation. SAF tools such as JSAF, OTB, and OOS allow entity behavior to be specified ahead of time. At a computational level, SAF behavior specifications typically amount to some form of finite state machine (FSM). An FSM can be thought of as (1) a set of states, each of which corresponds to some behavioral state (e.g., patrol along a given path); and (2) a set of transitions between states (e.g., when an intruder is detected, move to confront). Advantages of FSM-style SAF behaviors include clarity and predictability. FSMs are particularly good for implementing well-defined doctrinal behavior of limited complexity. However, SAFs have a limited capability to respond dynamically to changing circumstances. As allies, SAFs have little capability for coordinating or communicating with students. As enemies, SAFs are often little more than target drones. They do not model an adaptive, thinking enemy. Scenarios with such scripted behaviors have a very short useful life because students quickly learn to anticipate scenario events. (Whetzel, 2010)
>
> **Intelligent Automated Forces** Intelligent automated forces go beyond conventional SAFs with the ability to generate behavior dynamically in response to simulation events. Examples are TacAir-Soar (Coulter et al.) and ACT-R agents (Best, Scarpinatto, & Lebiere) for military operations in urban terrain (MOUT). The cognitive architectures underlying these capabilities are informed by a large body of accumulated psychological research. When used properly, these approaches can realistically mimic many aspects of cognition, particularly with respect to resource constraints such as reaction time and attention. (Whetzel, 2010)
>
> **Trainable Automated Forces (TAF)** Depicted in Figure 1, our vision is for SMEs, such as instructors, to train synthetic forces directly by demonstration, as in training human students.

Technical experts (e.g., computer programmers) must initially implement TAF for each type of role player required. Subsequently, however, SMEs can directly interact with TAF to enhance the domain expertise of TAF over time, without further support from a technical expert. TAF then relieves the SME of role playing so that the expertise of a single SME can be shared with any number of students. In our vision, the sharp distinction between the construction and operational phases (as required for traditional expert systems) is blurred. If an instructor recognizes a skills gap during one exercise, the instructor should be able to alter the behavior of automated forces in the next exercise to address the gap. Ideally, the long and expensive development pipeline can be virtually eliminated. This is the TAF approach. TAF training is an ongoing interaction between the instructor and the role-playing agent. The interaction is based on demonstrations of correct behavior by the instructor, and demonstrations by the system of its current understanding. Our goal is for the instructor to be able to interrupt and correct TAF when its actions diverge from the instructor's intent. TAF learns from such corrections and will not repeat the same mistake. Because this approach is data driven, objective behavior validation is more feasible compared with other approaches for automated forces. (Whetzel, 2010)

TAF is based on behavioral cloning, which is an established technique for building agent behaviors. Widrow and Smith (1964) first studied the technique for the pole-balancing (or inverted pendulum) task. The term —cloning [mimicry] implies that the agent simply replays previously recorded behavior, but most applications of behavioral cloning have some capability to generalize to new situations. Nevertheless, the performance of a clone will degrade as it encounters situations that are dramatically different from any encountered during training. Behavioral cloning has been successfully applied in simulations of tasks such as piloting an airplane, operating a crane, and riding a bicycle. Similar techniques are popular within robotics and are known as learning by observation or learning by imitation [memetic learning]. (Whetzel, 2010)

TAF cannot learn correct behavior if the behavior depends on missing information.

TAF Technical Description Implementing TAF consists of creating a role and then populating the role with example behavior. The role consists of information (e.g., from a nine-line brief) such as the location of a target and the time on target. Any information that cannot be gleaned from the inputs specified for the role cannot influence the behavior of TAF; thus incomplete information may lead to incorrect behavior. However, extraneous information may also degrade performance by confusing TAF. If training data are relatively sparse (i.e., the number of inputs is large relative to the amount of training data provided), spurious patterns are likely to be found in the training data, and learning these patterns will lead to unpredictable TAF behavior. (Whetzel, 2010)

Perception The world model presents ground-truth information from an allocentric (or global) perspective. Each TAF actor must be provided with a perception model to filter and adapt this information for its own needs. (Whetzel, 2010)

Perception then transforms the allocentric information provided by the world model to egocentric inputs for the learning or inference algorithm. (Whetzel, 2010)

Inputs As shown in Figure 2, information from the world model and setting is filtered and transformed by perception to form the *input* for TAF's *learning and inference algorithms*. The completeness (or incompleteness) of the input is a constraint on the realism of TAF behavior. Several factors prevent TAF from receiving a complete set of inputs, leading to degraded behavior. (Whetzel, 2010)

> TAF receives ground-truth information that may be unavailable to human combatants, granting TAF (in effect) a 360-degree field of view through walls, mountains, water, jammers, and darkness. In practice, software agents may use this unfair advantage to counter their own inherent limitations in perception and intelligence, with uneven results. More realistic perception models can also be implemented if necessary. (Whetzel, 2010)

> The role template also specifies a learning algorithm. The current implementation of TAF uses the open source Sandia Cognitive Foundry (http://foundry.sandia.gov), which supplies a wide range of learning algorithms, e.g., linear regression, nearest neighbor with locally weighted linear regression, ID4 rule induction, support vector machines, and backpropagation for neural networks. The role template also specifies the type of knowledge base for each algorithm (e.g., learned link weights to parameterize a neural network, or the rule set created by ID4). (Whetzel, 2010)

> Also, manual intervention in TAF control (e.g., moving an entity from the wrong trajectory) interferes with the simulator's internal dynamics model (position and speed) and thus may impart enormous momentum on the simulation entity. (Whetzel, 2010)

As we have seen in work with automated forces it is possible to create a set of agents, and these agents use copying to learn in trainable automated forces. Another area that uses copying in warfare is that of Memetic warfare.

Cyborg Soldiers

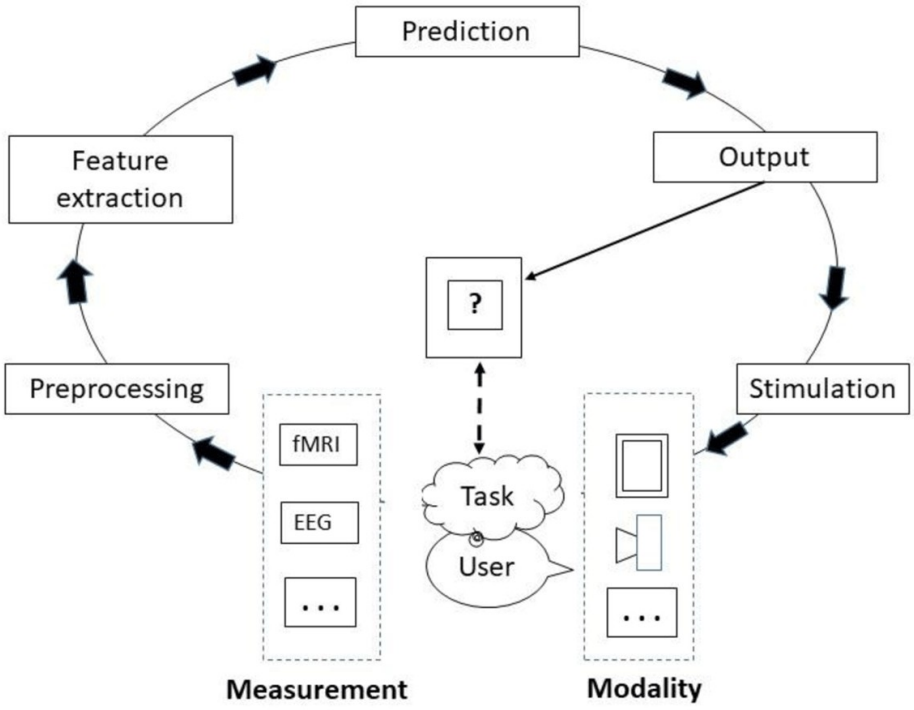
In recent years the discussion of the transition to human-machine hybrid's has become popular among military theorists and planners. In the US, the discussion of transitioning to cyborg soldiers by 2050 has been discussed (Emmanuel et al, 2019). In the literature the discussion makes a distinction between human-machine interactions that are not cyborged and those that are Nørgaard & Linden-Vørnle (2021). They delineate a distinction between cyborgs and 'centaur warfighters':

> it is important not to confuse the notion of the cyborg warrior with the concept of the 'centaur warfighter', which is often used as a metaphor for human-machine teaming. The two concepts are closely related, but not synonymous. This distinction can be expressed as the difference between integration and automation of machine intelligence, perception, and reasoning. Whereas centaur human-machine teaming consists of humans plus machines, with machines performing clearly demarcated automated functions, the cyborg warrior functions as a neurally enhanced and integrated system architecture, merging human and machine cognition. Centaur human-machine teaming does not necessarily imply cognitive or sensory enhancement of the human operator. Human and machine cognition is not neurally integrated. Instead, humans and machines perform different role-specific tasks that are largely based on predetermined decision models where the machine's role is conditioned by one or more rule sets.

Emmanuel et al (2019) studied the creation of Cyborg soldiers in their DoD sponsored study 'Cyborg Soldier 2050: Human/Machine Fusion and the Implications for the Future of the DOD' in which they identified several key areas that will have added value in combat:

- ocular enhancements to imaging, sight, and situational awareness;
- restoration and programmed muscular control through an optogenetic bodysuit;
- auditory enhancement for communication and protection; and
- direct neural enhancement of the human brain for two-way data transfer.

As can be seen this is a total re-imagining of human existence, where humans become little more then software to AI programming. We can also see how automated troops using RC spoken of previously would be easy to deploy to a Battalion of Cyborgs. It should be pointed out they discuss these enhancements as being derived from implants into the human body, we know from this research work that this is probably obfuscation as it is possible to do all these things remotely using Ahronov-Bohm based technology, for instance optogenetics bodysuit can be replaced by LED or other laser based photon emissions.



Flow chart on cyborg soldier loop of processing brain waves, AI processing of data, then re-importation for the inverse function or 'stimulation' to the soldier after reading brain state. (Nørgaard & Linden-Vørnle 2021)

Memetic Warfare: Memes, imaginal encoding of suggestion

Memes to most people are those quirky and funny images with captions usually based on a common image set that get passed around social networks, like the Hitler video phenomenon. It may seem odd to include a section on Memes in a book on Neurowarfare or Information Warfare. However, it is a common research area in the Military Industrial complex, NASA, and other government scientific research centers. The military interest in Memes is attested to by UMD Professor Finkelstein who had a contract with DARPA to develop a curriculum for the DoD on Memes (Finkelstein, 2008, 12), typically meme research is cited in counter terrorism literature, later we shall study some of these approaches of using memes to redirect or steer terrorists away from acts of violence, whereas terrorists use memes to create more mimicked terrorist attacks, where the propagandists never put themselves on the line, letting the memes do their work through others for their purposes. Finkelstein writes regarding the purpose of the research for the military:

> The purpose of the following overview is to provide an indication of the prospective value of memetics to the U.S. military for conventional and asymmetric operations, including counter-terrorism. The attempt to establish a scientific basis for memetics is critically important. For example, within a suitable memetics framework could be the means to prevent irrational conflict and promote rational solutions to endemic national and international problems. Of

course, without safeguards memetics can become a double-edged sword. (Finkelstein, 2008, 12)

He calls for a quantization of memetic engineering so it will be of a greater military value:

> If memetics can be established as a scientific discipline, its potential military worth includes applications involving information operations to counter adversarial memes and reduce the number of prospective adversaries while reducing antagonism in the adversary‘s military and civilian culture, i.e., it could have the ability to reduce the probability of war or defeat while increasing the probability of peace or victory. (Finkelstein, 2008, 17)

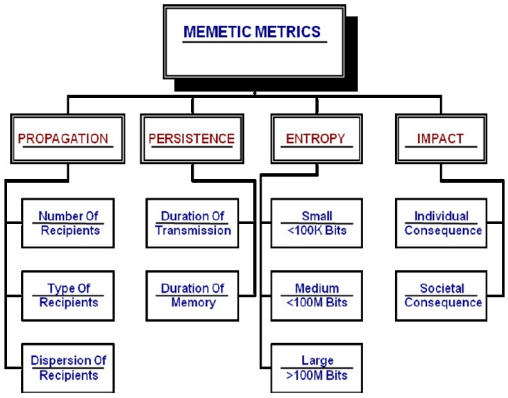


Figure 1: Memetic Metrics and Submetrics

The purpose of the research matches up with typical Reflexive Control in the sense of convincing an enemy to do what you want using deception as the primary tactic, which is the purpose of memes. This is directly related to overall information warfare, specifically in such areas as counter-intelligence and influence. One can imagine the need for being able to communicate desired outcomes through memes in such a contemporary environment where memes become viral on national information networks, media, and social nets. Finkelstein writes how this can affect the battlespace of mind:

> Potentially, memetics can have a major effect on psychological operations, military deception, and public affairs. Psychological operations (PSYOP) are intended to induce or reinforce foreign attitudes and behavior favorable to the originator‘s objectives and to convey selected information and indicators to foreign audiences to influence their emotions, motives, objective reasoning, and ultimately the behavior of foreign governments, organizations, groups, and individuals [Reflexive Control]. PSYOP focuses on the cognitive domain of the battlespace and targets the mind of the adversary. It seeks to induce, influence, or reinforce the perceptions, attitudes, reasoning, and behavior of foreign leaders, groups, and organizations in a manner favorable to friendly national and military objectives. It exploits the psychological vulnerabilities of hostile forces to create fear, confusion, and paralysis, thus undermining their morale and fighting spirit. There are strategic, operational, and tactical PSYOP, as described in Joint Publication 3.53, *Doctrine for Joint Psychological Operations* (5 September 2003). Strategic PSYOP consists of international activities conducted by US Government agencies primarily outside the military arena but which may use DOD assets. Operational PSYOP is conducted across the range of military operations, including during peacetime, in a defined operational area to promote the effectiveness of the joint force commander‘s campaigns and strategies. Tactical PSYOP is conducted in the area assigned to a tactical commander, for a range of military operations, to support the tactical mission. Psychological operations may occur across the spectrum of peace to conflict to war, integral to diplomacy, economic warfare, and military action, from negotiations and humanitarian assistance to counterterrorism.
> (Finkelstein, 2008, 17-8)

Memetics is not limited to just military actions, it also is applicable to economic warfare and diplomacy. Of course we are all very aware of the commercial as one of the most commonly experienced meme samples. Counter-Propaganda is a major goal of using memes:

> [information warfare] involves actions executed to deliberately mislead the adversary‘s military decision makers about friendly military capabilities, intentions, and operations, thereby causing the adversary to take specific actions (or inactions) that will contribute to the accomplishment of the friendly force‘s mission. According to *Information Operations Roadmap*, DOD (30 Oct. 03), military deception should be one of the five core capabilities of IO and the value of military deception is intuitive. Counter-propaganda includes activities to identify and counter adversary propaganda and expose adversary attempts to influence friendly populations and military forces situational understanding. It focuses on efforts to negate, neutralize, diminish the effects of, or gain an advantage from foreign psychological operations or propaganda efforts. (Finkelstein, 2008, 18-9)

In contemporary times the main propaganda channel for some extremist groups is that of social networks from Facebook, Telegraph, Twitter, etc. communications on these channels is often of a condensed form, neither prosaic nor of large bit size. Human propaganda communication mainly is conducted in audio, visual and tactile means. In a previous topic we discussed the work of Dr. John Norseen, the concept of ‘Thought Injection’ is a form of altering images, in the case of a terrorist suspect it is to change the ideation toward negativity through violent outlashes and replace that thought or semiotic with another non-violent image or ideal. In network communications, like a Social Network, which can easily also be a Terrorist Cell, there is the ideal of learning by simulating what others do, this is known as memetic learning, also used in video games, such as Shadow AI of *Killer Instinct*, which we see on a daily basis on Facebook or Twitter or any other number of social networks, though this goes on all the time in all minds, it is when a meme is transferred from one mind by any means, that the memetic learning occurs, or imitation. Norseen wrote of the similarity of Though Injection with Memes:

> Semiotic binding seems to work best when the brain is entrained between 7.83 Hz and 14 Hz, with special tunneling and neurochemical surges in the 9 to 10 Hz regime. The meditative Theta and modified Alpha-Theta states would appear to be the quiet zones where the ability to attend to internal mental representations can best be captured for reconstruction back through the efferent central nervous system pathways to show the world what floats in the mind. This delivery mechanism which brings forth creation back into the world in any number of newly reconfigured states (eolithic capacity) could be the semiotic description for the concepts in Richard Dawkins 1970's notion of 'The Building Blocks of Comprehension,' the Thought Memes. (Norseen, 1996)

Here Norseen, who is not alone is seeking to engineer memes for counter-terrorism (see below), specifically calls out the ideals of Richard Dawkins and his notion of ‘memes’

> Memetics has been a recent subject of interest as a new method for information exchange. In communities, memes have been studied to understand and enhance group learning. Richard Dawkins first defined memes as a unit of cultural transmission. Essentially, memes are ideas that evolve according to the same principles as biological evolution. Memetic learning works by transmitting units of cultural ideas or symbols from one mind to another. All ideas that exist within an individual's mind are examples of memes. Memes that are good at replicating leave more copies of themselves in minds. Examples of memes are catch phrases, musical themes,

> scientific ideas and sayings. In robotics, examples of memes are algorithms, observations, and instructions. (Truszkowski, 2014)

In another chapter on Reflexive Control we discussed the concept, simulacrum (information packets), used by the Russians or Soviets to influence an adversary. A meme is a simulacrum or as the terminology used in the West is that of a 'semiotic' an image or ideal, a simulacrum. In a later section we will learn how memes are used in robotic systems to learn through copying each other. The history of the ideal of memes is a good starting point in understanding how memes could be both important to humans (animals) and machines.

In terms of understanding the relationship of memes to security, specifically counter-terrorism, is the work of Pech & Slade, they have studied the similar concept but without any kind of physical change to consciousness, of re-wiring terrorists thoughts or memes. In Pech & Slade 2005 they bring forth the original context of Dawkin's meme then apply it to security:

> Dawkins (1998), the original memeticist, saw memes as replicable and transmittable units of information that travel from mind to mind. The meme became ''the unit of cultural inheritance'' and was first coined in 1976. Wilkins (1998) in turn saw that the meme was a simple self-replicating packet of information and defined its two essential characteristics. First, the size of the meme was not fixed but could vary, and second the importance of the fidelity, or resistance to change, of the meme was critical to its survival. Dawkins (1998) viewed memes as self-selecting, and that ''Individuals who are predisposed [psyop softening]... toward imitation are on a fast track that may have taken others a long time to build up'' (Dawkins). The notion here is that memes not only transfer themselves from mind to mind, but also encourage replication through the imitation of behaviours individuals see as desirable, particularly in role models. (Pech & Slade 2005)

The meme is a viral element in communications theory, it's first criteria to be a meme is that it is immune to large alteration, it has high formal fidelity, but from this fidelity comes an effect where the meme takes on a life of it's own, becoming like a virus, toxic memes, in a computer, replicating and promoting further damage:

> Pech's (2003, p. 61) statement ''the attachment of such labels confers a communicable mantle upon the act of violence and takes the event in question to a level where copying behaviors are triggered'' indicates that if such memes are allowed to exist and replicate then further acts of violence can be expected. (Pech & Slade, 2005, 48)

One could see the security threat this could pose to security personnel with access to media, particularly social media. The emergent or viral nature of memes is discussed:

> Dawkins (1998, p. 306) has argued that ''a mind can become prepared by certain memes to be receptive to particular other memes'' [suggestion cascade]. Clark (1997) describes the concept of emergence, which has relevance to the clustering effect of like memes and the existence of the memeplex, or complex of memes, as coined by Dawkins (1998, p. 306). Clark (1997, p. 74) divided the phenomenon of emergence into two distinct categories. The first, direct emergence, occurs where the properties and relations between individual elements primarily take on a life of their own and environmental elements have only background influence. Clark's notion of indirect emergence describes the predominant influence of the environment in triggering behaviors both individual and collective. (Pech & Slade, 2005, 48-9)

Memes can group up and form connections, this is referred to as a Memeplex. A pool of options or adaptations of the original meme, but appealing or having resonance with an audience. The viralness or emergence has two modes: direct (non-environment, hypnotic), indirect emergence. Out of the

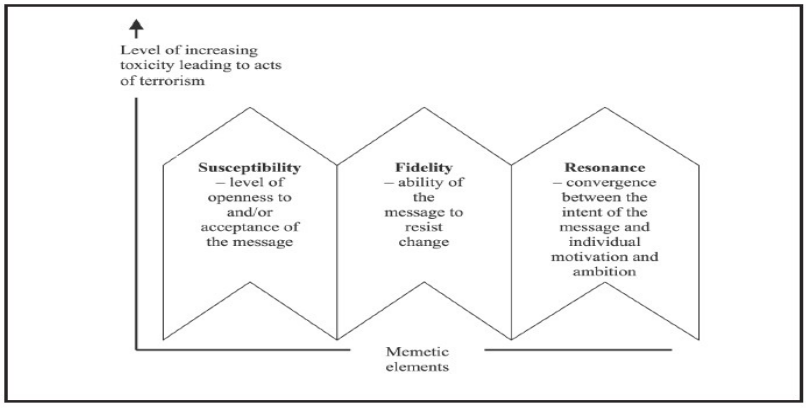
memeplex, like in nature, a virus or toxic adaptation may develop, these memes that cause people to act destructively are known as toxic memes:

> A toxic meme can be described as a self-replicating packet of influential information communicating a message of empathy that generally runs counter to the values and norms of society or which is in conflict with the needs and expectations of that society. Such memes, it may be argued, have the potential to threaten the very social fabric itself. (Pech & Slade, 2005, 50)

The lifecycle of a meme has three distinct stages, Pech & Slade 2005 a target site (a mind), fidelity (self-replication structural consistency), resonance within the target to act:

> Meme replication requires three significant elements. First, a person must be susceptible to the message within the meme. They must be open to the intent of the message or open to the form in which the message is communicated.
>
> Second, the meme must possess fidelity. It must be immune to change that may be affected upon it by the different cultures, education, and socialization that makes up the world views of potential hosts. Without such fidelity the message held within the meme will change or dissipate as it replicates from mind to mind; this concept has been provided by Dawkins (1976) and Wilkins (1998).
>
> Third, the meme must resonate with some intrinsic, emotion or value already possessed by, or appealing to, the host. A meme, such as a terror meme, can only transmit its existence through the behaviors of an empathetic host. In such a host, the message in the meme may not necessarily be consciously supported at the outset. However it may resonate through a variety of emotions and empathetic reactions until it is consciously accepted and its resultant behavior demonstrated in acts replicating the violence buried within the message. This concept has been partially articulated by Dawkins (1998). These fundamental concepts driving the nature of the meme have been modeled by Pech and Slade (2004) within the context of re-engineering organizational behavior, and have been adapted to counter mimicked acts of terrorism in Figure 1. (Pech & Slade, 2005, 50)

Figure 1 Elements for mapping memetic catalysts and determining which of the three elements are most prone to re-engineering



(the 3 stages of a Meme, Pech 2003)

The argument behind being able to target a receptive resonant chamber of a human mind involves a process of scripting, not unlike coding algorithms in silicon based hardware, researchers actually refer to it as a meme algorithm.

> According to Huesmann (1986) script theory proposes that violence observed in the mass media, provides aggressive scripts that vicariously define situations and guide future behavior

> for receptive individuals. The risk for both society and the individual, under script theory, is that an individual may select a script populated by toxic memes, analogize this to a given situation, and adopt the role provided. (Pech & Slade, 2005, 54)

With the equivalence in human community is that of a Trojan Horse attack:

> Donald explains the impact that a culturally-derived memory field, stored externally from the individual's consciousness and which describes knowledge, values, and information, can have on the mind. He refers to this as a cerebral Trojan Horse, which he argues can, to an extraordinary degree, make the human mind externally programmable. These external programmes can create deliberately engineered experiences [which we study in Reflexive Control and Perception Management]. The commercial world has manufactured vicarious experiences to such an extent that some people may have difficulty differentiating between what is real and what is not. Cultural influence can have similar effects and provide similar scripts. (Pech & Slade, 2005, 56)

Below we will see the direct security vulnerabilities and viral growth of memes implementation by terrorists and counter-terrorist activities to shape warfare to one's advantage. Primarily aimed at destabilizing an adversary, preceding an invasion.

This is based in studies in the US from the 1950s, Festinger 1957, who studied cognitive dissonance which is a prime research topic in western influence operations.

> The phenomenon of indirect emergence from such an externally programmed and scripted environment sees the act of terrorism adopted as an appropriate means of managing the terrorist's state of cognitive dissonance. Festinger's (1957) version of cognitive dissonance describes a state where more time is spent rationalizing misbehaviors rather than actually engaging in rational action. The resulting terror meme rationalizes and justifies illegal, irrational, and dangerous behavior to help maintain a sense of cognitive consistency for those who are raised according to the conditions described by de Mause (2002), those who are dissatisfied with their current government/authority structure, those who are easily led, and those who are susceptible, for a variety of other reasons, to the terror meme's message. In addition a number of cultures are becoming increasingly susceptible to fundamentalism as a means of protecting minority power bases, and as a means of preventing or rejecting external ''interference'' from more moderate states. (Pech & Slade, 2005, 56)

Now that we understand the basics of memetics the question then becomes how does one manage memetics. Given that automated troll farms can spew out an amazing amount of memes through automation, it is necessary for a military to also develop automated processes in counter-adversarial operations, memetics is a very red versus blue approach in cybersecurity. To get to full automation of both attack (red) and defense (blue) we must have a systems representation of memes. This is where memetic engineering steps in and the creation of artificial agents that can learn from memes, which is to say they learn by imitation of other agents, in addition to learning they can also generate new memes and contribute to memeplexes, developing robot agents is the first step to automation of memetics. But what is a meme to a robot?

> Dawkins coined the term 'meme' to describe a unit of cultural transmission, and we use this terminology here. We propose a definition of a robot meme as follows: a contiguous sequence or package of behaviours copied from one robot to another, by imitation. In the artificial culture lab we 'seed' each Copybot with initial behaviors which, in this paper, are self-contained movement sequences. (Winfield & Erbas, 2011)

A team of NASA and Lockheed-Martin developers write regarding these robots:

> For robots, memes have been defined as sets of instructions that can be followed to evolve behavior. Instructions can be encoded as written text and visible or vocal action. To allow for memetic learning, memes should also include observations of the environment. A robot that is able to observe and intelligently imitate the behavior of others is able to participate in memetic learning. In order to perform intelligent imitation, a robot needs to be able to process memetic information. This process involves evaluating models, examples, and patterns which the robot observes. In addition, the robot is expected to analytically compare its current knowledge to the new information it is observing.
>
> A robot has modified its individual knowledge base when it learns a new meme. Each robot is expected to evaluate active memes in the community knowledge base for strengths and weaknesses when deciding whether to learn them. It can be expected that each individual robot will benefit from the aggregation of other robots which are also participating in memetic learning. When the community knowledge base and size expands, there is a larger selection of memes which can be evaluated and learned. With a larger community knowledge base, a robot has a larger selection of memes to modify to develop novel memes. All individuals capable of participating in memetic learning are able to generate new memes. Also, individuals who are able to broadcast observations are capable of generating new memes.
>
> The memes that are in the knowledge base of a robot are in constant competition with all other memes in the meme pool. The meme pool is the collection of all existing memes that are accessible to the other individuals in the community. An individual robot may develop new memes that become candidates for imitation in the community meme pool. The community meme pool increases with the addition of novel memes generated by individual robots in the community. The connection between an individual knowledge base and the community meme pool is similar to the structure of a distributed cloud network. A distributed cloud network is structured so that each individual is connected to the cloud where they can access the knowledge bases of others in the community pool. Individuals will be able to quickly access, process, and analyze the collective knowledge within the cloud (Truszkowski, 2014).

So we have learned that in memetic learning in robot groups that there is a common knowledge base, the memes add to the knowledge base, there is selection for adaptation by the robotic agents and this occurs in a cloud network, non-local storage/action. Looking ahead in 2014 the team envisions developing meme agents, where the memes themselves are intelligent.

> An area for future investigation is to modify the meme's structure so that the memes themselves are intelligent. Memes could be encapsulated with intelligent software, similar to a mobile agent, that can make the meme an active instead of a passive entity. Active memes could monitor a robot’s state, listen to communications between robots or look at other memes that are being passed between robots to determine if it may be needed. If so, the meme could then push itself to a robot that needs it or insert itself into the active reasoning being done by the host robot. Memes that are not needed after a period of time could decay and destroy themselves if they are outdated or no longer needed. This could help limit the proliferation of memes in a system. Active memes could also automatically update themselves based on changes in the environment and observed learning in the robot host (e.g., seeking protection when a sandstorm is forecasted). They could also seek out similar memes and combine with them to form better memes, or even use techniques like genetic programming to improve or transform themselves into something new. This could speed up the evolution of new memes.
>
> An intelligent and mobile meme would have to be lightweight so it could easily move between robots without having a large communications overhead. A large number of heavy weight memes being sent between robots over a limited bandwidth network, such as might be on mars, could overload the network. Memes could also use swarming or other behaviors to increase their individual impact and to quickly react to new situations a robot may encounter where there is no one meme that has all of the information. Swarming could also help to keep

individual memes small since the swarm would provide all of the needed knowledge. Since each situation could be different, new swarms would be spawned based on the situation at hand. An additional area which requires further investigation is the implementation of such a communication and analytical model to communities (Wang, 2008).

There are unknowns regarding future requirements and specifications for a group of robots to be fitted with the intelligence to perform memetic learning. The communication structure of modern systems may be challenged and need to be modified to allow future systems to be capable of implementing this technology.(Truszkowski, 2014)

Memetic Learning is a sub-discipline within Genetic Programming or Algorithms, which are developed in Defense applications, and is used in automated systems that also perform self-evaluation and modification of their own programs based on fitness measurements similar to organic genes hence the name, genetic programming. Hougen explains the derivation:

> Memetic learning algorithms are related to genetic algorithms in the form that solutions may take. As with standard genetic algorithms, one defines possible solutions as sequences of discrete values, typically binary strings. For genetic algorithms, each entry in the sequence is considered a gene, whereas with memetic learning algorithms, each entry in the sequence is a meme. The possible values for that entry are the alleles. Learning in memetic learning algorithms proceeds as follows: A population of individuals with random alleles for each meme is constructed and tested on the task. Individuals observe their own overall fitness values and those of the other individuals in the population. Further they observe the partial fitness values of the partial candidate solutions that they are able to identify, both for themselves and for the others. They then replace their memes for those portions of the solution for which they have low fitness values, using imitation. (Hougen et al, 2003, 4-5)

The robotic agents using fitness measures are able to tell if their solution to a problem is of a higher or lower value thereby they can then imitate the higher fitness valued solutions of other agents. In Learning there are two modes; learning by direct experience and learning by imitation, the key variable here in determing the type of learning is interaction with the environment, direct experience is environmental interaction learning but imitation leaves out the environment for direct copying:

.

> **3.1.3 Memetic Learning Algorithms.** For memetic learning algorithms, we use a simple gene-splicing method. As with the GA, each policy table is encoded as a one-dimensional chromosome, where each allele at each locus is a left or right action decision. The population, again, is a collection of policy tables and the population size is set again at 50. However, rather than using a selection mechanism to generate new individuals for successive generations, we retain all individuals—generations are marked by changes *learned* by individuals. Each individual in the population is given a single trial. Based on that trial, individuals learn in up to two ways. First, all individuals learn by direct experience. Second, if an individual fails, it learns by imitation.
>
> **Learning by direct experience.** As with reinforcement learning, each policy table entry has associated with it a score that reflects our confidence in that action and each score has associated with it an eligibility value. The score and eligibility values are updated during and after each trial using Equations 1, 2, and 3.
> **Learning by imitation.** If an individual fails, it also learns by imitation by considering each of its loci separately. For each locus, the probability that the current allele is retained is equal to $max(s; 0)$. For alleles that are not retained, a replacement allele is chosen for that locus using proportional probability selection based on the scores of all of the alleles for all individuals for that locus. (Hougen et al, 2003, 8-9)

The framework for developing automated memes is presented by University of Arizona, working for the Office of Naval Research (award #N00014-18-1-2761) , researchers in their project “memeBot”, which is able to automatically read tweets and generate memes from those tweets, so it would not be difficult to aggregate sentiment and produce both adverse and averse effects through those memes on social networks. Below is a flow diagram for how the automated memes are produced, either to fight terrorism or to promote it:

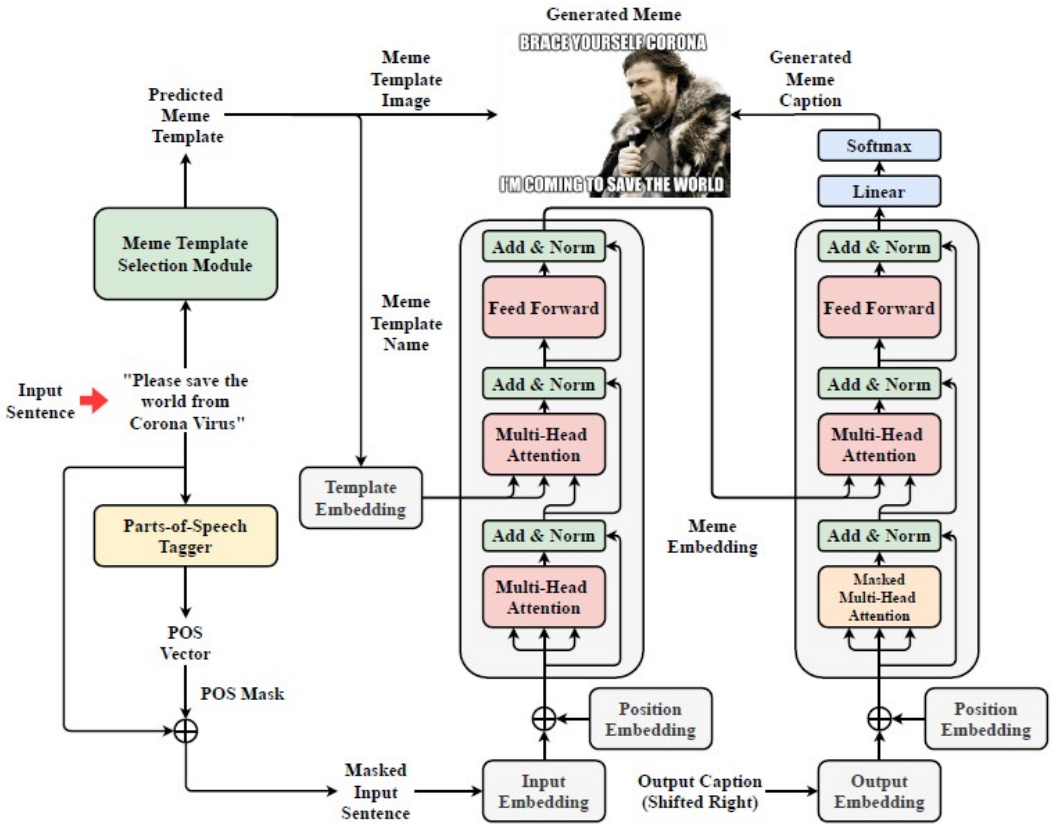


Figure 3: memeBot - model architecture. For a given input sentence, a meme is created by combining the meme image selected by the template selection module and the meme caption generated by the caption generation transformer.

(Sadasivam, 2020, 3)

Researchers explain their project and what it can accomplish:

> We have presented memeBot, an end to end architecture that can automatically generate a meme for a given sentence. memeBot is composed of two components, a module to select a meme template and an encoder-decoder to generate a meme caption. The model is trained on a meme caption dataset to maximize the likelihood of selecting a template given a caption and to maximize the likelihood of generating a meme caption given the input sentence and the meme template. Automatic evaluation on meme caption test data and human evaluation scores on Twitter data show promising performance in generating an image for sentences in online social interaction. (Sadasivam, 2020, 9)

In later sections we shall see how autonomic systems are engineered and how this is used to manage engineered systems for memetics. Now that we have seen what memes are and how they emerge in networked groups and can even be automated into cyber agents it is time to look into a case study of the use of memes to create terrorists by studying the Boogaloo Movement in the United States and their use of memes to propagate their call for armed insurrection to start a new civil war.

Boogaloo Bombs

> Boogaloo Goal: The hope of these militants is to incite violence sufficient for society to betray the American civic tradition by forcing immense violence to protect it [Nazi accelerationism, see Chapter 1 Black International]. (Goldenberg et al, 2020c)

At protests in Seattle in 2020 I first directly encountered members of the Boogaloo Bois as they refer to themselves, usually young white males from the suburbs that have an automatic weapon fetish or are exorcising some PTSD from previous US military experience leaving them with wounded heads that whether we like to admit it or not are susceptible to outside influencing (see below), making PTSD veterans easy prey for adversarial information operations such as those conducted on public social media. For the first time in my long history of attending leftist protests I witnessed strange groups of armed men wearing Hawaiian shirts parading around as 'security' for the protesters. Eventually, the open display of weapons resulted in at least one fatality, when a African American teenager was gunned down by one after joy riding in a stolen vehicle of one of the protestors in an apparent miscommunication which led to lethal force being deployed by untrained amateurs. The Boogaloo Bois were also responsible for overt terrorist attacks, where a US Military member gunned down a Federal Protective Services officer and wounded another in a drive by ambush, other terrorist related arrests of Boogaloo Bois could be mentioned here but we continue.

In a Brookings Institute report the Boogaloo is described:

> In their willingness to carry out attacks against law enforcement personnel to incite what they consider an imminent civil war, the Boogaloo movement poses a serious threat to police. The movement has its origins online, and its adherents have skillfully used memes to incite violent insurrection and terror against the government and law enforcement. Especially widespread on Facebook and Instagram, Boogaloo enthusiasts share instructions for explosives and 3-D printed firearms, distribute illegal firearms modifications, lead users into encrypted messaging systems, distribute violent propaganda, and target their recruitment efforts towards active and former military personnel. The movement is a case study in how we still do not entirely understand how radicalization occurs in the digital domain. (Goldenberg et al, 2020c)

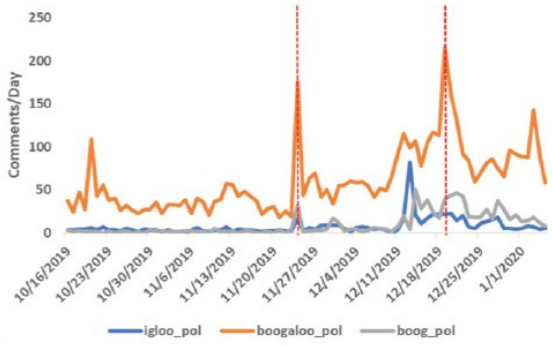


Fig 2. Trend analysis on 4chan's /pol/, a radical and trend setting Web community, shows ramping and spikes in the frequency of "Boogaloo" comments and synonyms such as "igloo" and "boog." Two of the changepoints our algorithms detected on "boogaloo" (red line) occur within 24 hours to two key events: the "Whiskey Warrior" standoff and the impeachment of Donald Trump.

The boogaloo catchphrase, or meme, is based on the 1984 movie sequel Breakin' 2: Electric Boogaloo, which critics panned as a shockingly unoriginal, near-mirror copy of the original film. As adopted by meme culture, the term is often used by libertarians, gun enthusiasts, and anarchists to describe an uprising against the government or left-wing political opponents that is a near-mirror copy, or sequel to, the American Civil War. While the reference has been around for years, recent iterations have caught on and spread quickly over the

past few months. While many still use the boogaloo meme jokingly, an increasing number of people employ the phrase to incite an apocalyptic confrontation with law enforcement and government officials or to provoke ethnic warfare. (Goldenberg & Finkelstein, 2020b)

Language on 4chan seems to associate the term to “racewar” and more coded conspiracies such as “dotr,” or day of the rope, a fantasy to instigate a civil war and murder race traitors. These acts would presumably be accomplished by “rwds,” a code for Right Wing Death Squads, such as the “atomwaffen” division, a neo-Nazi domestic terror organization. Other coded associations such as “shtf” stand for “shit hits the fan,” a slang for the end of civilization, and a term that appears near topics of doomsday preparation, “ammo” and “stockpile.” (Goldenberg & Finkelstein 2020b)

Military and Veteran Susceptibility:

Among the boogalois I have engaged with there is a strong military aspect to their organizing, which indeed if you are seeking to fight a civil war then you would be engaged in military contexts. This puts veterans at risk of recruitment as most boogalois bois have no military experience or even basic gun training, at least among those I have encountered.

> Furthermore, the meme’s emphasis on military language and culture poses a specific risk to military communities due to the similar thematic structure, fraternal organization, and reward incentives**.** (Goldenberg & Finkelstein 2020b)
>
> The military community, in particular, may merit special consideration in risk evaluation and social-climate research because seditious memes are now tailored for infection among veterans and active service members. (Goldenberg & Finkelstein, 2020b)

The main threat is to law enforcement from the boogaloo bois. This is a congruent with attacks on neutral policing by far-right groups as a means to create a privileged legal group to do whatever it feels like doing to the underprivileged legal groups. By attacking the police they seek to undermine community support for policing, mind you there are different models of policing, if they were seeking to end injustice in legal systems then why would they only attack soft-targets dissociated with any history of legal or police injustice by any of the communities they have engaged in actions in, for instance there are no attacks in Minneapolis or Louisville? Yet, with such a threat the law enforcement is little equipped to rapidly adapt:

> Memetic warfare is still very much a mystery to both policy makers and officials working within the American law enforcement community. In this ignorance, the worst actors amongst boogaloo groups possess a distinct advantage over government officials and law enforcement: They already realize that they are at war. Public servants cannot afford to remain ignorant of this subject because as sites, followers and activists grow in number, memes can reach a critical threshold and tipping point, beyond which they can suddenly saturate and mainstream across entire cultures. (Goldenberg & Finkelstein, 2020b)

Overall there is a missing link in the defense against threats such as this as noted by Ascott:

> The West is desperately lagging in its memetic capability. US Marine Corps Major Michael B. Prosser proposed that NATO open a meme warfare centre. In his 2006 thesis, he looked to Dawkins’s ideas of memes as units of cultural transmission that held the potential to ‘be used like medicine to inoculate the enemy and generate popular support’. He noted that information

> operations, psychological operations and strategic communications weren't using memes effectively. In the following decade, NATO never did open a meme warfare center, but the idea didn't go away and is now starting to gain traction again. (Ascott, 2020)

One way to counteract Memetic warfare is to setup surveillance of memes using time series analysis:

> Time series analysis can signal an extremism climate and facilitate a "weather station" for trends in extremism on a meme by meme basis. Such a station is needed to create alerts and notifications which can be adapted for the use of information vaccines, strategic counter messaging, and campaigns for better moderation. (Goldenberg & Finkelstein, 2020b)

Along with surveillance is the need for quick adaptation, not just in communications:

> ...respond adaptively, and strategize communications during sensitive domestic operations. This is especially crucial because the conspiracy is preparing a viral-social media environment to instigate an uprising in response to missteps and police violence during these operations in order to threaten security on a national scale. (Goldenberg & Finkelstein, 2020b)

The importance of early warning with memetic warfare is highlighted with the swarming nature of memetic warfare and learning in animals. Very quickly a cascade effect can emerge with swarms programmed by Memes on social networks, one example of this is the Whiskey Warrior event:

> Just as swarming insects elicit signals that can recruit entire colonies to converge on either enemy or prey rapidly, the "Whiskey Warrior" event, on November 24th in New York, demonstrates how the boogaloo meme can tactically alarm recruit followers to simultaneously deploy en mass in both cyber, and, potentially, real-world domains. When Alexander Booth posted images and videos of an ongoing standoff with police on his pro-gun Instagram handle "Whiskey Warrior 556," the former infantryman appeared in full camo and body armor with a knife clipped to his chest. Booth claimed the officers were employing red flag laws to strip him of munitions and posted memes on social media to merchandize the standoff specifically as a "boogaloo" triggering moment. These posted memes, with powerful ingroup signaling, immediately went viral on the chans and amongst several extreme boogaloo sites and right-wing militia groups on social media. From this point, followers began to obstruct police operations through targeted phone calls and online campaigns and incited armed resistance from social media, and the posts even succeeded in attracting one dedicated follower who claimed to be Facebook streaming from the scene of the standoff itself. Though Booth's Instagram account only held several thousand followers at the start of the event, it boasted over 130,000 by the time the standoff ended. (Goldenberg & Finkelstein, 2020b)

Gamification in Boogalois Bois Memes

MARTYRDOM, MEMES AND MOTIVATION: AS SHARED NARRATIVE GROWS IN THE MILITIA-SPHERE, TRACES OF AN UNDERLYING SHARED-INFORMATION NETWORK EMERGES (Goldenberg et al, 2020c, 7)

The convergence of martyr episodes, revenge attacks, terror tactics, romanticizing terrorists and terrorist movements, and group-level coordinated behaviors now appear in highly visible real-world

events. This suggests a shared identity and shared narrative in the Militia-sphere. But do these groups and individuals connect through these events in less-tangible networks on social media? (Goldenberg et al, 2020c, 15)

While still preliminary, this introduces concerns that shared narratives like martyrdom may connect distributed, militia-oriented users across mainstream and fringe networks. (Goldenberg et al, 2020c, 9)

Analysis we gathered on "Bot Sentinel," and "Hoaxy," publicly available resources for charting bot activity support this possibility. Remarkably, we find that both WWG1WGA and QAnon hashtags are often among the single the top most frequently tweeted hashtags by trollbot accounts on Twitter (appendix figure 5) and networks that promote these hashtags are high in bot-like participation (appendix figure 6). As Q conspiracy is becoming more explicitly militant (appendix figure 7) future research must seek to determine how these underlying causes differentially contribute to the popularization of seditious conspiracy (Goldenberg et al, 2020c, 11)

In a methodology that is similar to that as used by the Black Internationals policy of infiltration and intoxication the boogaloos take on a guise that is hard to pin down to far-right extremes:

> The boogaloo won't present itself as either Trump-based, right wing, or white-supremacist ideology, with any one set of predictable political grievances. These events suggest that the boogaloo seeks to co-opt several grievances, across several political and racial spectrums into a single, monolithic and anti-government mob with chilling new tactical and technological capacities. (Goldenberg et al, 2020c)

Religions ideals of matryrdom also are incorporated into the boogaloo bois context. Many see themselves as 'Last of the Mohican' types, which using reinforcement such as memes is easy to program into someone with enough repetition.

> Aaron Swenson, a boogaloo enthusiast posted #hisnamewasduncan over a selfie featuring weapons and body armor on Facebook after Lemp's martyr episode on April 4th. On April 23rd, he live-streamed his revenge hunt on police in Texarkana. While Swenson was soon arrested after an hours-long chase, for inciting terror and possession of illegal firearms, these tactics portend a worrying evolution of real world/virtual violence in the Militia-sphere, akin to the live streaming of ISIS beheadings and other innovative uses of media by hybrid and distributed sporadic terror groups. Martyr narratives and revenge killings are likely to continue to shape a shared identity among users in these groups.
>
> In addition to martyr myths and revenge killings, more familiar methods of terror for violent organizations include the use of bombs and explosives. On May 4th, Bradley Bunn, a militia enthusiast threatening militant violence against law enforcement for their role in enforcing quarantine restrictions, was arrested in his home in Colorado in possession of two one-pound containers with gunpowder for reloading .308 caliber cartridges and four pipe bombs. Bunn admitted outright that the murder of law enforcement was his goal; these materials were set to be used as a lethal trap for entry, just like the Lemp episode. (Goldenberg, 2020c, 7-8)

Martyr myths along with such 'apocalyptic' stereotypes such as a global pandemic serve to further derange the rational hold people have that are affected by memetic warfare. The connection with the anti-government QAnon movement and anti-lockdown protests during the pandemic are noted as cross fertilization takes hold in a rich environment of grievances, real or imagined:

> In the face of COVID-19, QAnon now witnesses massive growth and appears to militarize, like the boogaloo, with revolutionary and apocalyptic themes in a more militant and global mode of inciting revolt. QAnon conspiracies, such as "the Great Awakening" for instance, refer to a

> moment in which elites will be defeated and the truth will be revealed, and are often featured at anti-quarantine rallies. Other conspiracies suggest that a “new world order” now prepares to emerge under the tyranny of Bill Gates and George Soros, the Rothschilds, and other elites, who— through their vaccination attempts—seek to establish mind control, world domination, genocide, financial gain, or some combination of these.
>
> The result is that the QAnon conspiracy now invites two modes of disinformation which converge on COVID-19 in dangerous ways. The first is anti-vaccination and anti-science disinformation about the virus itself as a weaponized plot. The second is increasing militarization in the conspiracy group which combines disinformation on COVID-19 with seditious themes that parallel the boogaloo. WWG1WGA, for instance, comprises a key QAnon slogan expanding to “where we go 1 we go all” and features with Great Awakening material and at reopen rallies both as an in-group cheer, but also as an all-at-once go signal, reminiscent of the boogaloo. Indeed, evidence of militarization in the QAnon conspiracy now abounds with references to a “Q-army” complete with military-style badges. (Goldenberg et al, 2020c, 10)

We should not be surprised that an image based sensory input serves to have such a large ‘hypnotic’ effect on people. As the visual cortex is written of repeatedly by those involved in information warfare as the seat, specifically visual processing of images, of suggestibility or control.

Bibliography:

- APDC, (2009) *Effects-based approach: is it still valid?.* In Pathfinder Air Power Centre Bulletin Issue 109, April 2009
- Ascott, T. (2020) *How memes are becoming the new frontier of information warfare.* in Australian Strategic Policy Institute online: https://www.aspistrategist.org.au/how-memes-are-becoming-the-new-frontier-of-information-warfare/ 19 Feb 2020
- Backus, G., Bernard, M., Verzi, S., Bier, A., Glickman, M. (2010) *Foundations to the Unified Psycho-Cognitive Engine.* SANDIA REPORT SAND2010-6974 Unlimited Release October 2010
- Backus, G., Glass, R. (2006) *An Agent-Based Model Component to a Framework for the Analysis of Terrorist-Group Dynamics* SANDIA REPORT SAND2006-0860P
- Behzadan, V., Nourmohammadiy, A., Gunesz, M. and Yukselx, M. (2017) *On Fighting Fire with Fire: Strategic Destabilization of Terrorist Networks.* In ASONAM '17: Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017July 2017 Pages 1120–1127 https://doi.org/10.1145/3110025.3119404
- Djordjevich, D., Xavier, P., Bernard, M., Whetzel, Glickman, M., Verzi, S. (2008) *Preparing for the Aftermath: Using Emotional Agents in GameBased Training for Disaster Response* USDOE National Nuclear Security Administration (NNSA) DOI: 10.1109/CIG.2008.5035649 ·https://www.researchgate.net/publication/221157660
- Emanuel, P.; Walper, S.; DiEuliis, D.; Klein, N.; Petro, J.; Giordano, J (2019) Cyborg Soldier 2050: Human/Machine Fusion and the Implications for the Future of the DoD online: https://community.apan.org/wg/tradoc-g2/mad-scientist/m/articles-of-interest/300458/download

- Epifanovskaya, L., Lakkaraju, K., Stites, M., Letchford, J., Reinhardt, J., Whetzel, J. (2018) *Online Games for Studying Human Behavior. In book: Social-Behavioral Modeling for Complex Systems (pp.387-406) https://www.osti.gov/servlets/purl/1470956*

- Finkelstein, R. (2008) *A Memetics Compendium*, prepared for the DARPA Task Order CA-FIN-3212-024-08 http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.731.4497

- Finkelstein, R. (2008) *Information Propogation, Impact & Persistence (InfoPip): Defining Memes, Briefing Report.* DARPA Task Order CA-FIN-3212-024-08

- Glickman, M., Whetzel, J., Basilico, J. (2010) *Trainable Automated Forces.* https://www.researchgate.net/publication/241970920_Trainable_automated_forces

- Goldenberg, A., Finkelstein, J. (2020b) *CYBER SWARMING, MEMETIC WARFARE AND VIRAL INSURGENCY: How Domestic Militants Organize on Memes to Incite Violent Insurrection and Terror Against Government and Law Enforcement.* in A CONTAGION AND IDEOLOGY REPORT. Network Contagion Research Institute https://ncri.io/wp-content/uploads/NCRI-White-Paper-Memetic-Warfare.pdf

- Goldenberg, A., Baumgartner, J., Farmer, J., Zannettou, S., Blackburn, J. (2020c) *Covid-19, Conspiracy and Contagious Sedition: A Case Study on the Milita-Sphere.* in *A CONTAGION AND IDEOLOGY REPORT. Network Contagion Research Institute https://networkcontagion.us/wp-content/uploads/NCRI-White-Paper-COVID-19-Militia-Sphere-1-June-512pm.pdf*

- Hougen, D., Carmer, J., Woehrer, M. (2003) *Memetic Learning: A Novel Learnings Method for Multi-Robot Systems.* https://www.cs.ou.edu/~hougen/mrs2003.pdf

- Margulies, P. (2016) *Surveillance By Algorithm: The NSA, Computerized Intelligence Collection, and Human Rights.* in Florida Law Review Volume 68 | Issue 4 Article 3 July 2016 https://scholarship.law.ufl.edu/cgi/viewcontent.cgi?article=1321&context=flr

- Pech, R. & Slade, B. (2005) *Imitative terrorism: A diagnostic framework for identifying catalysts and designing interventions.* in Foresight · February 2005 DOI: 10.1108/14636680510581312

- Nørgaard, K., & Linden-Vørnle, M. (2021). *Cyborgs, Neuroweapons, and Network Command*. Scandinavian Journal of Military Studies, 4(1), pp. 94–107. DOI: https://doi.org/10.31374/sjms.86

- Sadasivam, A., Gunasekar, K., Davulcu, H., Yang, Y.(2020) *memeBot: Towards Automatic Image Meme Generation* Arizona State University, Tempe AZ, United States https://arxiv.org/abs/2004.14571

- Togelius, J., Anderson, D. Stephenson, M., Salge, C., Levine, J. Renz, J. (2018) *Deceptive Games https://arxiv.org/abs/1802.00048*

- Truszkowski, W., Rouff, C., Akhavannik, M. (2014) *Memetic Engineering as a Basis for Learning in Robotic Communities* in conference presentation: AIAA 2014-1323 Session: Intelligent Learning and Decision Making Published Online:10 Jan 2014 https://doi.org/10.2514/6.2014-1323

- Whetzel, J., Abbot, R., Basilico, J., Glickman, M. (2010) *Trainable Automated Forces* Sandia National Laboratories Albuquerque, NM https://www.researchgate.net/publication/241970920_Trainable_automated_forces

- Winfield, A., Erbas, M. (2011) *On embodied memetic evolution and the emergence of behavioural traditions in Robots* in Memetic Computing · December 2011 DOI: 10.1007/s12293-011-0063-x at: https://www.researchgate.net/publication/235277651

- Xavier, P., Hart, B., Hart, D., Gayle, R., Oppel, F., Whetzel, J. (2017) *Dante Agent Architecture for Force-On-Force Wargame Simulation and Training Sandia National Labs* in AAAI Vol. 13 No. 1 (2017): Thirteenth Artificial Intelligence and Interactive Digital Entertainment Conference