

Mr. Larry Willis
Information Awareness Office (IAO)
Predicting Human Behavior

Good morning. I'd like to take this time to talk about predicting human behavior.

Much of our lives involve predicting, anticipating, and preparing for the future. Business journals predict the course of economic developments, almanacs predict the weather, and newspapers predict the outcome of sporting events. The ability to predict future events is the goal of many sciences, both physical and behavioral. While the predictive capability in the physical sciences is well documented, this capability in the behavioral science is still evolving. That is, even though behavioral prediction theory has been successfully applied to select narrowly constrained areas within clinical and personnel psychology, it remains relatively unexplored for more complex and dynamic environments such as military analysis and planning. In fact a common viewpoint is that the seemingly irrational nature of human behavior renders prediction impossible in a complex environment. The Wargaming the Asymmetric Environment team respectfully disagrees.

The mission of the WAE program is to develop and demonstrate models and tools "tuned" to specific adversaries, thus enabling analysts and decision-makers to better anticipate, predict, and act against those who threaten U.S. and Allied interests. Asymmetric adversaries, for WAE's purpose, span the continuum from guerrilla warfare as experienced in the Balkans to the present terrorist threat.

The research questions for WAE are numerous and include at a high level: do signatures exist that provide relevant predictions of our adversaries future behavior? If predictive signatures exist, what predictive fidelity do they possess? Can these signatures predict the timeframe, target, direction of interest, or tactical characteristics of future attacks? What is the shelf-life of these signatures? Can these signatures provide any insight into how to influence these adversaries? So let me quickly provide a disclaimer, WAE's technology does not possess the ability to predict the specific day, time, target address, and method of an attack. However, WAE does contend that prediction of an adversary's behavior is possible at a detail level that, at a minimum, dramatically increases the specificity of the indication and warning space. The metrics for assessing WAE's success are straightforward, and the predictive technology is frequently validated against both historical information and simulated real-time information.

WAE's conceptualization of the role of prediction within the analytical process traces back to the FBI's Behavioral Sciences Unit, which coined the term "signature" to characterize behavior suggesting psychological intent. This term was developed as a way to distinguish intent-based behaviors from capability-based behaviors required to logistically execute a criminal plan. The Behavioral Sciences Unit postulated that both of these types of behavior are required to derive accurate insights into an adversary behavior to enable an understanding of who, what, where, when and why.

As WAE views the current analytical process through the lens of the Behavioral Sciences Unit's definitions of intent- versus capability-based analysis, it is apparent the current process and associated technology is capability-based. This process is characterized by the near continuous examination of detailed information regarding group members, their movements and communications as well as the movements of money, logistics, etc. Although emerging technology is increasingly being used to automatically search the information space to discover, extract, and structure data into a graphical representation of a group's capability, the inference process itself remains complex and manual. In fact, even with substantial information and relative certainty about the available information, the sheer volume makes the inference process so complex that reasonable analysts can differ greatly in their inferences about likely future actions.

WAE's approach, while complementary to the current analytical process, differs from it in a number of significant ways. First, WAE's focus is on select behaviors, such as attack behaviors. The rationale for this is WAE is not attempting to establish an overall assessment of a group's capability but rather to derive the predictive signatures associated with the decision to use that capability. Second, WAE's focus is on deriving predictive patterns from more high-level information associated with the political, cultural, and ideological environment surrounding the group. The rationale for this is the covert nature of group behavior, which by

definition attempts to disguise or vary behavior, and dilutes any potential predictive patterns at detailed levels of behavior. Third, WAE's focus is on deriving signatures that can directly address the question of how the U.S. can potentially influence the adversary's behavior. The rationale for this is, that by analyzing this broader behavioral environment, the predictive signatures will include behaviors by the U.S., NATO, and others, which can be correlated against the adversary's behavior and thus provide further insights into potential interactions. Fourth and finally, WAE's focus is on developing computer-based reasoning techniques to automatically quantify the predictive contribution of each signature preceding and following the target behavior. If successful, the result of this approach will be an automated prediction tool to augment the analyst's inference process and support hypothesis testing of potential intervention strategies.

WAE's research approach is based largely on behavioral learning theory and uses an operant learning framework to identify the interaction between an adversary's behavior and the more subtle behavioral forces within this broader environment. Within this framework, WAE describes the adversary's behavior as the response; that is the behavior that occurs in the presence of some triggering event and is accelerated, decelerated, or maintained as a result of its consequences. A consequence is termed the reinforcing stimuli and it controls the direction and valence of the behavior of interest. A reinforcing stimulus is positive if the behavior is more likely to occur when followed by the reinforcing stimuli and conversely is negative if the behavior is less likely to occur when followed by the reinforcing stimuli. Typical examples of reinforcing stimuli within the asymmetric context are retaliation, policy changes, cessation of negotiations, publicity, and profit. Just as important within this framework is the triggering signature or discriminative stimulus, which precedes the adversary's behavior and represents the condition which was present in close temporal proximity to when the adversary encountered the reinforcing stimuli. In operant theory the discriminative stimulus now controls that behavior, not in the sense of eliciting it, but in the sense that the behavior has a higher probability of occurring when this stimulus is present than when it is absent. One can conceive of the discriminative stimulus as a signal or trigger indicating that the previously encountered reinforcing stimulus is likely to be forthcoming if the adversary were to emit the behavior. Typical examples of discriminative stimuli within the asymmetric context are presence of U.S. or NATO forces in disputed territories, public statements defaming the group or group leadership, and political or economic pressure on the group to enter negotiations. Behavior learned within this operant framework is a function of the discriminative stimuli (signatures) and reinforcing stimuli (consequences), and thus if one seeks to predict or influence the response (behavior) one must identify and control either or both these stimuli.

An example of WAE's initial results is represented by a recent test of a predictive model of a specific European guerrilla organization. The model was derived from a training set of attacks and non-attacks that spanned the life of this organization. The model was then validated in an experimentally blind environment against an equivalent test set. The results are presented here as the percentage of correct predictions along two dimensions. First is the fidelity of the attack characteristics: attack/no attack, target, direction of interest, and tactical characteristics. The second is the predictive accuracy, true positive and true negative. True positive, for example, represents the percentage of correct predictions that the next attack would reflect the nature of each attack characteristic. Conversely, true negative represents the percentage of correct predictions that the next attack would not reflect the nature of each attack characteristic.

As you can see the results are high across the board and from a statistical perspective, the predictive accuracy for each attack characteristic is significant well beyond the traditional $p < .01$ criteria. What is clear, at least for this group, is the existence of discernable, predictive patterns to their behavior at a level specific enough to support the indication and warning process. The signatures clearly distinguish between the environmental conditions preceding an attack versus no attack, a civilian versus a military target, a privately owned versus a publicly owned target, and an attack on a U.S. versus NATO target. Furthermore, the results indicate that the fixed set of signatures is predictive over the life of this group as well as over the evolution of this group's tactical capability.

A characteristic of WAE's intent-based approach is that the predictive signatures are comprised of the behavior of our adversary, the U.S., NATO, and others. The benefit of having the signatures spread across the full range of participants within a conflict is that by correlating the U.S. and NATO behaviors with the adversary's behavior at each level of attack characteristics, we get a quantitative assessment beyond prediction to include potential influence. Specifically, we can look at U.S. and NATO behavior that precedes

or follows the adversary's attack and gain specific insights into what triggers the adversary to act in a certain way. Additionally, these correlations also provide us with insight into how to potentially influence an adversary from one behavior to another. This does not mean to imply that one can dissuade a group that is ideologically driven from attacking a U.S. interest. However, one may be able to shift the environment to one that is less conducive to an immediate attack or possibly shift the group from a more lethal tactic to a less lethal one.

The future research direction of the WAE program will focus on technology to increase the breadth and depth of the predictive capability in support of the operational areas of warning and intervention. WAE will attempt to broaden our predictive capability beyond attack characteristics to include enabling behaviors that precede an attack, indicate an impending shift in attack strategies, or sequences of behavior. WAE will attempt to obtain a deeper predictive capability by developing hybrid-reasoning technologies to automatically derive the predictive signatures and the temporal window corresponding to each prediction.

The operational benefits to the successful development of predictive technologies reside in both the warning and intervention processes. The predictive technologies when fused with automated extraction technology will support a continuous indication and warning capability that can aid analysts in providing earlier, more specific warnings for both general and group specific threats. Additionally, the influence points derived from these predictive technologies can be integrated with other planning technologies to test intervention hypotheses such as action and reaction to various U.S. or NATO initiated behavior. And finally, the automation of these predictive technologies will allow analysts to rapidly modify existing predictive models or develop predictive models of new asymmetric groups.