Thesis Proposal
AI-Enabled Social Cyber Maneuver Detection and Creation

Matthew Hicks

April 23, 2024

School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213

Thesis Committee:
Kathleen M. Carley, Chair, Carnegie Mellon University
Patrick Park, Carnegie Mellon University
Mohamed Farag, Carnegie Mellon University
David M. Beskow, United States Military Academy

Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Abstract

As social media platforms emerge as primary conduits for information dissemination, influence operations, and narrative shaping, discerning their significance within the broader information environment becomes vitally important. The BEND framework provides scaffolding for accomplishing this by assessing influence and manipulation online through the detection of social-cyber maneuvers. Previous research has focused on BEND maneuver detection primarily through message/text-maneuver matching. While this detects executed BEND maneuvers, it does not evaluate their effectiveness.

The purpose of this thesis is to transform BEND into an operationally relevant framework for evaluating social cybersecurity. I accomplish this by first nesting BEND within current US military doctrine and providing recommendations for incorporation in future doctrine. This allows leaders to properly place social media campaigns within the context of the broader information domain.

Next, I use effects-based metrics to identify BEND maneuvers within networks and narratives. I connect these maneuvers with broader campaigns and validate this new technique using both a real-world study and a simulations study. For the real-world study, I look at four different datasets. These datasets include Twitter/X, Telegram, and Reddit data, and concern the 2022 Ukraine-Russia conflict, the COVID-19 vaccine, the 2020 US elections, and the 2020 Nice terrorist attack. For the simulations study, I conduct an initial exploratory simulation using NetLogo before advancing to more comprehensive simulation in Construct. Through these simulations, I demonstrate the pair-wise effectiveness of maneuvers and investigate multi-maneuver interactions to draw conclusions about using maneuvers as countermeasures.

Finally, I leverage Artificial Intelligence to develop synthetic training data that closely mirrors real-word data. My AI-Enabled Scenario Orchestration and Planning (AESOP) tool allows training planners to create scenarios and output templates describing events, social media accounts, and narratives. AESOP also constructs news sites, articles, and URLs - at custom levels of reliability and stance. I then use these outputs to generate social media traffic - at scale - using Large-Language Models.

These contributions will improve our ability to detect BEND maneuvers, increase our understanding of how narrative campaigns execute these maneuvers, suggest countermeasures to the maneuvers, and support creation of realistic training to practice implementing the countermeasures.

Contents

Chapter 1

Introduction

1.1 Overarching Thesis Goals

Social media platforms have immense impact on information dissemination, influence operations, and narrative shaping. This provides an avenue for interested actors to influence public opinion, manipulate people, and even conduct war by other means.[11] Placing the actions of these actors within the context of broader information operations and characterizing the maneuvers they conduct is critically important to countering or negating them.

The purpose of this thesis is to operationalize a social cyber security framework to add such context to adversary maneuvers, to characterize those maneuvers, and then use that context and characterization to formulate counters. This is done with the US military in mind, so I first look at current US military doctrine and its intersection with social media campaigns and the information environment. Second, I develop effects-based methods for identifying the maneuvers that actors take on social media - this enables clear characterization of the impacts of maneuvers as well as a method for assessing their effectiveness. Finally, I provide tools for constructing social media datasets that enable training against these maneuvers by applying other maneuvers.

1.2 Social Cyber Security

Looking at cyberspace through the lens of warfare is not new. Interactions between adversaries within cyberspace have often been referred to in military terms of attack and defense [20]. Cyberspace simulations have been used to model these conflicts, often closely emulating current physical military doctrine [30]. However, these simulations focus primarily on the cyber-terrain itself - accurately deducing that terrain has a large impact on the outcome of conflict [30]. However, just as the physical domain of warfare stretches into the digital space, so too does the information environment. This is social cyber security [17] [16]. Social cyber security welds the methodologies of the social sciences with the need to identify, assess, and counter the impact of information maneuvers.[16]

This field is often claimed/mishandled by multiple interested parties. In the US Department of Defense (DoD), there is not only Joint (all-service) doctrine[35] addressing it but there is also Service doctrine[23] and within services there is Branch doctrine[26] - all sticking a finger into the mess that is the social cyber security component of the much broader Information Operations.

While newly identified as an academic field and still nascent within the DoD, the need for a framework to scaffold understanding of these issues is not new and has led to the rise of various contenders. These include Ben Nimmo's 4 D's - dismiss, distort, distract, and dismay - focusing on Russian propaganda techniques.[40] Also, the ABC framework developed by Camille Francois[31], which looks at the Actors, Behaviors, and Content of a disinformation campaign and its successors ABCD[9] and ABCDE.[42] Finally, the SCOTCH framework - focusing on Sources, Channels, Objectives, Targets, Composition, and Hooks - was brought forward by Blazek in 2021.[15]

Amidst this crowded field lies the BEND framework. BEND provides a framework for discussing social-cyber interactions using narrative and network structures, but borrows the idea of informational maneuver from maneuver warfare [12]. BEND is shorthand for the social-cyber maneuvers: back, build, bridge, boost, engage, explain, excite, enhance, negate, neutralize, narrow, neglect, dismiss, distort, dismay, and distract. These maneuvers and their definitions are taken from Beskow and Carley's 2019 work Social cybersecurity: an emerging national security requirement [12] as refined and validated by Blane et al. in 2022 [14] and later in Blane's thesis work.[13] BEND arguably stands above other frameworks in that its approach is systematic and high level while still analysing key actors and behaviors. It is also not limited to disinformation - in contrast with Nimmo's 4Ds - and can be applied at the influence campaign level - similar to SCOTCH. It is also specifically aimed at enabling recommendations in the same way as ABCDE.[13]

Currently, the CUE+ method as outlined by Blane [13] is the cutting edge in BEND maneuver detection. This method has seen several iterative improvements - first in Uyheng et al. in 2020[46], then by Blane et al. in 2022[14], then Alieva, et al. in 2022[10], before Blane outlined the current method in her thesis work.

In this method, linguistic cues are extracted from message text using NetMapper software.[18] These NetMapper cues are a proprietary blend of concepts that represent a message's sentiment and the author's emotional state.[16] These particular cues - now referred to as CUES - are then loaded into ORA-PRO[18] - a network visualization and anlysis tool - where they are mapped to the original message. ORA-PRO is able to use these CUES, along with supplemental network information about the message sender, to provide a report that identifies BEND maneuvers associated with messages and actors - who is doing what to whom.

This is necessary for identifying BEND maneuvers within messages. However, the BEND maneuvers themselves - even in the extended definitions provided by Blane are impact-based. Indeed, the BEND maneuver descriptions and illustrative impacts provided by Blane are diagrams or illustrations of the effects the BEND maneuver will have. For instance, the diagram for the Boost maneuver is shown in Fig. 1.1

There is no mention of the content required within the message that caused the boost - the implication is that the message is defined by the impact it had. Blane goes on to derive message content requirements based upon CUES. Again, this is necessary to identify which messages are attempting to conduct which maneuvers. However, assessing the effects of the messages - detecting BEND not within the messages but in the effects these messages had on their targets - is missing.

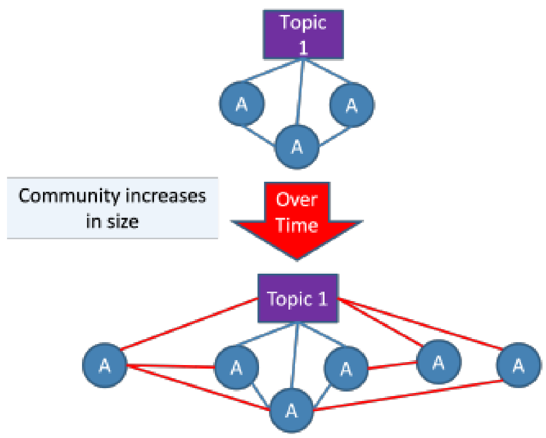

Figure 1.1: Excerpt from Blane's thesis work "Social-Cyber Maneuvers for Analyzing Online Influence Operations".[13]

Chapter 2

Data and Tools

2.1 Data

Previous research has examined several large-scale datasets for BEND maneuvers using existing techniques.[14] [13] I will be enriching these datasets by supplementing their predominantly Twitter/X collection efforts with analogous Reddit and Telegram data. I will then be reapplying BEND maneuver detection, both through the current methodology[13] as well as the extended effects-based BEND maneuver detection discussed later in this thesis.

Corpus Topic	Twitter/X	Telegram	Reddit
Nice, France Terrorist Attack 2020			
COVID-19/Vaccines			
US Elections 2020			
Ukraine-Russia Conflict			

Figure 2.1: Summary of Datasets used in this thesis.

French Attack in Nice 2020 An Islamic extremist attacked and fatally wounded three individuals inside a Roman Catholic church in Nice, France, on October 29, 2020. A research team from Singapore gathered two weeks' worth of Twitter data spanning from October 28 to November 4, 2020, covering the event. The Twitter data will be matched with corresponding Telegram (pending) and Reddit (pending) data on the same incident from the same period.

COVID-19 Vaccine This dataset is comprised of tweets discussing COVID-19 and the Pfizer vaccine. Theses tweets were collected using pandemic-related keywords and further refined to focus on vaccine discourse. They were collected from three distinct timeframes surrounding the introduction of the Pfizer/BioNTech vaccine: December 1-7, 2020 (preceding the rollout), December 8-10, 2020 (coinciding with the vaccine's deployment in the US and UK), and January 25-31, 2021 (six weeks post-rollout). These are augmented by additional segments collated by Janice Blaine[13], specifically addressing conspiracy theories and vaccine-related discussions. In total, this dataset captures a year-long narrative of the pandemic's impact. The Twitter data will be matched with corresponding Telegram (complete) and Reddit (complete) data on similar subjects from the same time period.

US Election 2020 Twitter data related to the 2020 US Election was collected using relevant key terms and accounts. Spanning from October 31, 2020, to November 7, 2020, this timeframe includes the days leading up to the election (October 31 to November 2, 2020), election day itself (November 3, 2020), and the ensuing period post-election (November 4 to November 7, 2020). During this period, discussions regarding a contested election and allegations of voter fraud surged due to delays in result announcements. The dataset offers insight into narratives surrounding the US elections process, spanning from the primaries through the aftermath of voting. Notably, the Democratic party emerged victorious, with Joe Biden declared the 46th President of the United States. The Twitter data will be matched with corresponding Telegram (pending) and Reddit (complete) data on similar subjects from the same time period.

Ukraine-Russia 2022 This dataset consists of Twitter data sourced sourced from November 2021 through November 2022. It encompasses a wide array of key terms in Russian, Ukrainian, and English, focusing on political figures, geographical locations, and other pertinent topics related to the conflict. The Twitter data will be matched with corresponding Telegram (complete) and Reddit (pending) data on similar subjects from the same time period.

2.2 Tools Used

This section describes the tools used throughout this thesis.

ORA-PRO

Formerly, the Organization Risk Analyzer (ORA) - Professional version, now known simply as ORA-PRO. ORA-PRO is a dynamic network analysis and visualization tool. ORA-PRO can import Twitter, Telegram, and Reddit data for detailed analysis.[18]. This thesis takes advantage of the built-in stance detection, Leiden grouping, and social metrics as well as the BEND maneuver detection through CUES.

NetMapper

NetMapper processes text to identify concepts and their network relationships. It uses dictionaries and custom parameters to enrich text before extracting concepts, which it then links together to create either semantic or conventional meta-networks.[18] This thesis relies on NetMapper for extracting CUES from social media corpora before importing the CUES into ORA-PRO.

NetLogo

NetLogo is a multi-agent modeling environment primarily used for simulating natural and social phenomena. Developed at Northwestern University, NetLogo provides an easy-to-use interface and a simple programming language.[45] This thesis used NetLogo to run agent-based simulations of two competing pole groups on the same stance in order to evaluate the head-to-head effectiveness of BEND maneuvers.

Construct

Construct is a simulation framework for implementing agent-based modeling in C++20[29]. Construct can parse DynetML files from ORA or CSV. It allows for the custom creation of models - including those for information diffusion. For this thesis, I propose to implement a BEND effects-based simulation using Construct.

OpenAI Models and API

This thesis required large amounts of LLM interaction and the OpenAI API was user friendly and easy to integrate with Python.[7] I used a GPT3.5-Turbo model (gpt-3.5-turbo-0125) for text and DALL-E model (dall-e-2) for images.

Local Large Language Model

Unfortunately, OpenAI models refused to respond properly to some requests for negative BEND maneuvers or on some negative topics due to guardrails. In those cases I ran a large language model locally. For this thesis, I used mixtral-8x7b based on its effective responses, lack of guardrails, and small size.[6]

Chapter 3

Research Plan

This thesis is organized into three chapters. The first chapter is concerned with understanding how social-cyber security fits within the broader Information Environment. The second chapter addresses enhancing our ability to detect and respond to social-cyber maneuvers. The third chapter provides solutions for developing social media datasets to help in training detection and response to social-cyber maneuvers.

3.1 US Military Doctrine

3.1.1 Research Questions

This chapter fits the BEND framework into current US military doctrine and provides recommendations for BEND incorporation in future doctrine. The focus is on producing doctrinally conformant, coherent social media staff inputs that use lexicon consistent across capabilities and functions. The output is a set of BEND products that mirror current doctrine and suggestions for a more comprehensive incorporation if BEND into doctrine.

Research questions addressed in this chapter are:

- How does the BEND framework fit into current military doctrine?
- How can it enhance current information environment analysis?

3.1.2 Doctrinal Synthesis

Current military doctrine on the social media aspect of the Information Environment (IE) is scattered between dozens of manuals and instructions and is encumbered by issues of both authority and ability. Current doctrinal examples that address the IE include:

- JP 3-13 Information Operations[35]
- NWP 3-13 Navy Information Operations[28]
- ADP 3-13 Information[27]
- AFDP 3-13 Information in Air Operation[23]
- CJCSI 3210.01C Joint Information Operations Proponent[19]

- DODI 3600.01 Information Operations[22]
- ADP 5-0 The Operations Process[25]
- MCWP 3-32 Marine Air-Ground Task Force Information Operations[32]
- JP 2-01.3 Joint Intelligence Preparation of the Operational Environment[34]
- ATP 2-01.3 Intelligence Preparation of the Battlefield[26]
- JP 3-60 Joint Targeting[33]
- JP 3-61 Public Affairs[36]

This is neither a comprehensive list of all applicable doctrine nor does it include those manuals which retain a either a Secret classification or Controlled but Unclassified Information (CUI) identifier. Throughout these doctrinal examples the IE is defined as "The aggregate of individuals, organizations, and systems that collect, process, disseminate, or act on information."[35] Unfortunately, because of the importance of the IE it is often discussed as the "Information Domain" - something not explicitly found in US doctrine. In none of these manuals is "domain" officially defined[4]; however, JP 1 Doctrine of the Armed Forces of the United States does discuss the "the physical domains (air, land, maritime, and space); the information environment (which includes cyberspace); and political, military, economic, social, information, and infrastructure (PMESII) systems and subsystems."[37] The fact that information is both an environment akin to the physical domains and a separately listed system should speak to its importance. The implication is that there exists separate regions marked by distinct physical characteristics (land, air, sea, etc.) and a mostly intangible information domain. This is borne out in discussions of domains -they often include the information domain despite its lack of doctrinal pedigree[8].

Doctrine elsewhere includes the information domain as an instrument of national power[37]. These instruments of national power are laid out as a part of the DIME framework which defines diplomatic, informational, military, and economic instruments - more recently expanded to include finance, intelligence, and law enforcement (DIME-FIL)[43]. Indeed the integration of the information "domain" or environment was the driving force behind the Department of Defense's Joint All Domain Command and Control (JADC2) initiative[5]. Whether a domain or an environment, information holds relevance equal to any of the physical domains.

As social media becomes increasingly important within the Information Environment[8], the BEND framework provides a solution for proper analysis and lexicon across warfighting functions. There are a wide variety of actions and actors within the DoD concerned with information operations. These include Public Affairs Operations (PAO) - "provide accurate and timely information" to the public[36] - Military Information Support Operations (MISO) - "influence the attitudes, opinions, and behavior of foreign target audiences"[36] - and Military Deception (MILDEC) - "deliberately mislead adversary... decision makers"[36]. All of these benefit from the common lexicon of tactics and maneuvers provided by BEND.

Unfortunately, US Department of Defense (DoD) actions in the IE are also hampered by issues of authority and capability. These issues are highlighted most clearly when considering social media. The authority for the DoD to conduct "operations" within social media is limited by a wide array of both law and policy. Limiting factors include the Posse Comitatus Act[3], the Fourth Amendment[1], the Privacy Act of 1974[2], and numerous DoD directives and regulations. These factors severely limit the the collection of information on US citizens and the conduct of narrative campaigns (beyond informational) that target US citizens. This thesis does not address solutions for these issues and assumes the DoD has established the property authorities to conduct analysis and response within the social media subsystem of the information environment.

While authorities remains a legal question, the capability issue is something this thesis can address. We know that decision makers need mappings of the information environment to know how to analyze it and respond within it. In other domains, they rely upon the Modified Combined Obstacle Overlay (MCOO)[34]. The land domain MCOO is perhaps the easiest to conceptualize. Obstacles, avenues of approach, key terrain, observation and fields of fire, and cover and concealment are graphically depicted atop a topographic map of the battlefield. This provides decision makers with a clear understanding of how the pieces of the land battle interact. The MCOO is further enriched by adding the enemy situation template (ENY SITEMP). The enemy situation template shows the disposition of known enemy positions overlaid on the MCOO, as well as the enemy's most dangerous or most likely course of action based on the enemy's doctrine. The lifting of this concept and application of it to the IE is not a new idea and others have posited the creation of a Combined Information Overlay [21]. This is encouraged by the description of the Consolidated Systems Overlay in JP 2-01.3 Joint Intelligence Preparation of the Operational Environment [34], see Fig. 3.1, which is unfortunately never explicitly applied to the IE.

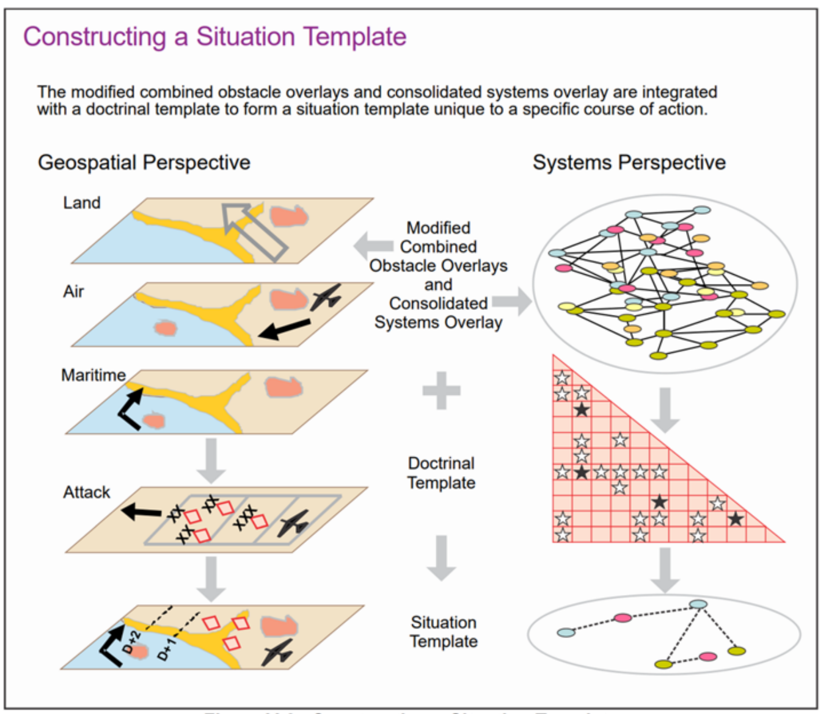

Figure V-2. Constructing a Situation Template

Figure 3.1: MCOO and CSO comparison from JP 2-01.3 Joint Intelligence Preparation of the Operational Environment

3.1.3 Proposed Work

The Social Media CSO/MCOO This thesis proposes to match the methodological rigor of the Modified Combined Obstacle Overlay with the systems perspective of the Consolidated Systems Overlay in order to graphically depict the current state of the social media component of the information environment for decisions makers. This will result in an IE MCOO for the social media subsystem and feed the social media enemy situation template, the social media running estimate, the friendly social media situation, aid in course of action development, and provide meaning to an IE impact assessment.

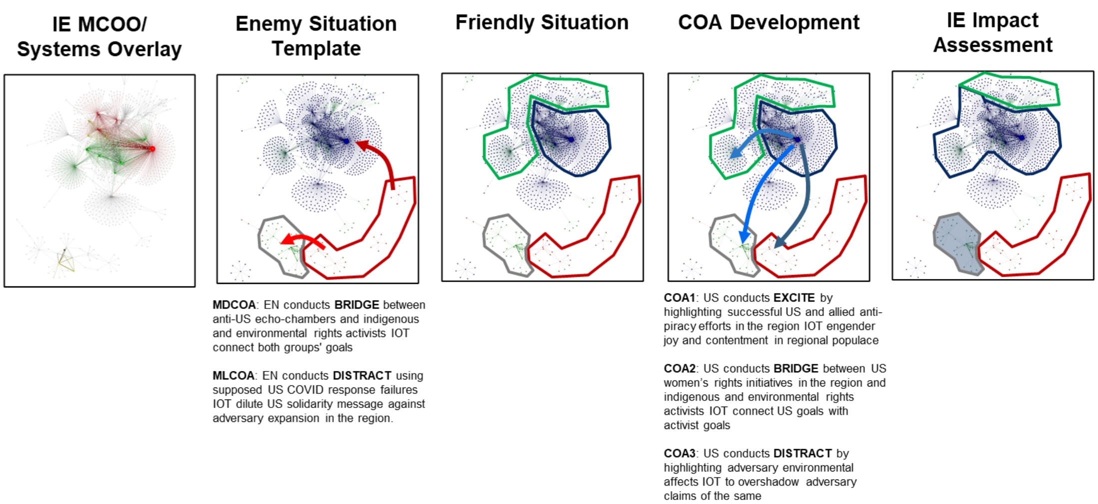

Figure 3.2: Information Environment products influence by BEND as a part of this thesis.

While the idea of mapping BEND to US military doctrine is my own, Project OMEN has been a major factor in developing examples. Project OMEN is a training scenario designed to educate players on social media analytics.[39] As a part of Project OMEN, Dr. Jonathan Morgan and I have worked to provide examples of products that match each of these areas from the BEND reports and outputs available. Beyond this, I will develop a social media running estimate template for information operations staffers that includes these examples and accompanying narratives.

I will synthesize these products within current US military doctrine. Specifically articulating how they align with the Joint Targeting Cycle (JTC)[33], Joint Intelligence Preparation of the Operational Environment (JIPOE)[34], and the Joint Operation Planning Process (JOPP)[38].

It is important to note that while social media is a critical component of the information environment, it is still only a subsystem. These products help scaffold the leaders and their staffs in understanding the role social media plays; however, social media is not a domain or an environment level consideration. The products outlined here reflect only a subset of those that would feed larger information environment planning processes and cycles.

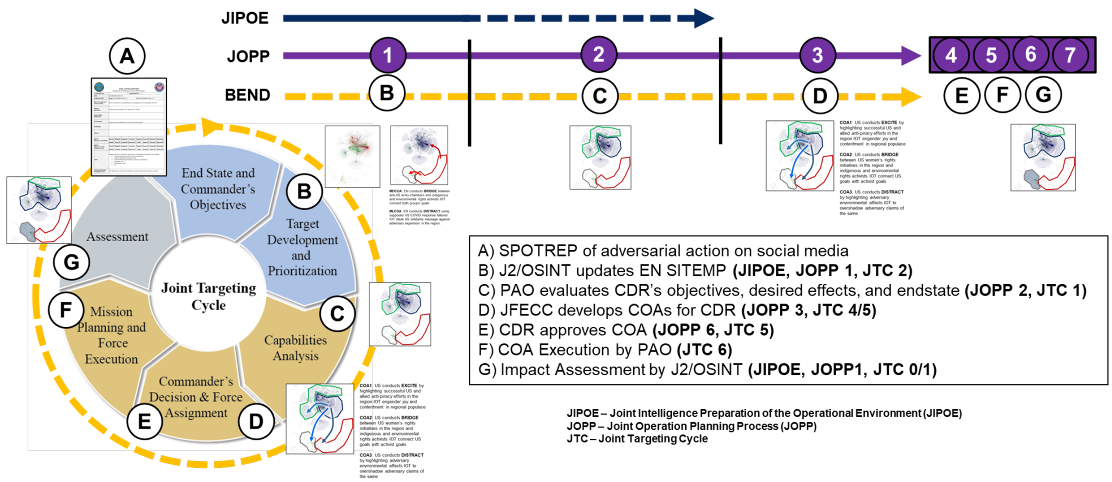

Figure 3.3: BEND integration with the Joint Targeting Cycle, JIPOE, JOPP.

3.2 BEND: Effects-based Detection

3.2.1 Research Questions

ORA and Netmapper combine to provide a BEND report that automatically detects BEND maneuvers. This methodology is based on work by Blane, who laid out a framework for analysis that uses a complex method for weighting CUES to identify maneuvers.[13] CUES here refers to the linguistic cues extracted from the message text through the NetMapper software. These linguistic cues are what ORA-PRO uses to identify BEND maneuvers per message text.

In this chapter, I propose to move beyond identifying BEND maneuvers within specific messages based upon derived intent. Instead of taking a source and a message and extrapolating an intended effect, I want to identify effects experienced by a target in order to determine the BEND maneuver experienced. To use an analogy from strategic bombing in World War II, rather than looking at a B-17 and its payload and determining that this will be a firebombing mission on Dresden, I want to look at the burned out city to assess not only the action taken but also the effectiveness of that action. I want to be able to look at Schweinfurt, see that the ball-bearing factory is destroyed, and point to a bombing group action that occurred on a particular night. I could not tell you which bomb destroyed the factory or even which bomber, but I can definitely point to a specific raid. In the same way, I will not be able to point to a specific message or actor but will be able to identify a narrative campaign associated with an effect induced in the target.

This will require bounded, over-time comparisons of groups in order to detect changes in the metrics. I will also need to account for more than one effect occurring at a time. Additionally, target identification - especially group target identification is an outstanding issue. Even more importantly, this thesis will require network and narrative metrics tied to the effects of the BEND maneuvers.

The key research questions for this chapter are:

- How can we detect the presence of BEND maneuvers through their effects?
- Can we match these maneuvers to narrative campaigns?

3.2.2 Effects-Based Metrics

With input from Jacob Shaha, I have developed a draft set of network and narrative metrics to begin using for this problem. Generally these metrics involve changes over time above the baseline corpus - requiring a computation of the metric against both the corpus and the target/target group. Additionally, many require heterogeneous graphs - involving actors, topics, and stance. The definitions below are derived from Blane[13] and paired with their corresponding effects-based metric.

Back maneuvers have discussion or actions that increase the actual, or the appearance of, an actor's importance or effectiveness relative to a community or topic. In order to detect if an actor has been the target of a back maneuver, I will be looking for a positive change over time - above the baseline corpus - in the centrality of an actor within the actor interaction network.

Build maneuvers are marked by discussion or actions that create a group, or the appearance of a group, where there was none before. Therefore, new group emergence is required for this maneuver - agent interactions within the group should change positively over time more than in the baseline corpus.

Bridge maneuvers involve discussion or actions that build a connection between two or more groups or create the appearance of such a connection. To detect this, I will look at both centrality and betweenness of the edge nodes of two groups. A positive change over time above the baseline corpus indicates a bridge maneuver.

Boost maneuvers require discussion or actions that increase the size of a group and/or the connections among group members, or the appearance of such. I will look at group size and graph density for positive changes over time above the baseline corpus.

Excite maneuvers include discussion or actions related to a community or topic that cause the reader to experience a positive emotion such as joy, happiness, liking, or excitement. For excite maneuvers, I will look for target output message emotional valence to be higher in happiness and surprise over time above that of the baseline corpus.

Explain maneuvers will exhibit discussion or actions that clarify a topic to the targeted community or actor often by providing details on, or elaborations on, the topic. For effects-based detection, I will look at topic specialization – with additional jargon and a net shift towards similar stance over time above the baseline being indicative of an explain maneuver.

Engage maneuvers involve discussion or actions that increase the relevance of the topic to the reader often by providing anecdotes or enabling direct participation and so suggesting that the reader can impact the topic or will be impacted by it. In order to detect the effects of engage maneuvers, I will look for a positive change over time in the proportional representation of the topic with the target group above the baseline corpus.

Enhance maneuvers show discussion or actions that provide material that expands the scope of the topic for the targeted community or actor often by making the topic the master topic to which other topics are linked. Effects-based metrics will be increased linkages (density) and centrality or betweenness changing positively over time above the baseline corpus.

Negate maneuvers include discussion or actions that decrease the actual, or the appearance of, an actor's importance or effectiveness relative to a community or topic. Therefore, I will be looking for a negative change in a target actor's centrality over time of a magnitude greater than that of the baseline corpus.

Neutralize maneuvers have discussion or actions that cause a group to be, or appear to be, no longer of relevance, or the group is dismantled. I will be looking for target group nodes that have more in common (connections) with other groups than themselves (group disappears) over the time-frame observed.

Narrow maneuvers exhibit discussion or actions that lead a group to be, or appear to be, more specialized, and possibly to fission, or appear to fission, into two or more distinct groups. Effects-based metrics will be the appearance of multiple groups where only one was present before within the actor network or the disappearance of links on the bipartite network from meta-agent group node to topic/stance nodes.

Neglect maneuvers show discussion or actions that decrease the size of a group and/or the connections among group members, or the appearance of such. For effects-based detection of the maneuver, I will look for density and/or size of a target group changing negatively over time above the baseline corpus.

Dismay maneuvers involve discussion or actions related to a community or topic that cause the reader to experience a negative emotion such as worry, sadness, disliking, anger, despair, or fear. As the inverse of excite, I will look for target message emotional valence increasing in anger, sadness, fear over time more than the baseline corpus.

Distort maneuvers include discussion or actions that obscure a topic to the targeted community or actor often by supporting a particular point of view or calling details into question. This should induce in the target increased topic specialization – additional jargon in message and net shift towards opposite stance over time more than the baseline corpus.

Dismiss maneuvers are marked by discussion or actions that decrease the relevance of the topic to the reader often by providing stories or information that suggest that the reader cannot impact a topic or be impacted by it. I will be looking for a negative change over time in the topics proportional representation - greater in magnitude than the in baseline corpus.

Distract maneuvers require discussion or actions that redirect the targeted community or actor to a different topic often by bring up unrelated topics, and making the original topic just one of many. For this, I will look for decreased linkage/density and decreased centrality and betweenness over time - in greater magnitudes than experienced by the baseline corpus.

A summary of the maneuvers - their definitions and effects-based metrics can be found in Fig. 3.4.

3.2.3 Proposed Work

Study 1: Real-world

This thesis proposes to measure the effects of BEND maneuvers in four different datasets: the 2022 Ukraine-Russia conflict, the COVID-19 vaccine, the 2020 US elections, and the 2020 Nice terrorist attack. The Twitter/X portion of these datasets were previously used for BEND detection by Blane.[13] This will allow me to compare the number of detected BEND maneuvers executed - based on CUE+ detection within messages using ORA-PRO - against the new effects-based metrics.

A global within-corpus comparison of the CUE+ detection against the effects-based detection will give a rudimentary understanding of the effectiveness of BEND maneuvers within the cor-

Name	Definition	Effects-Based Detection
Back	Discussion or actions that increase the actual, or the appearance of, an actor's importance or effectiveness relative to a community or topic	Centrality in interaction network, importance to group/ changes positively over the time more than the baseline corpus
Build	Discussion or actions that create a group, or the appearance of a group, where there was none before	New group – agent interactions / changes positively over the time more than the baseline corpus
Bridge	Discussion or actions that build a connection between two or more groups or create the appearance of such a connection	Centrality, betweenness of the edge nodes of two groups / changes positively over the time more than the baseline corpus
Boost	Discussion or actions that increase the size of a group and/or the connections among group members, or the appearance of such	Size of group, graph density / changes positively over the time more than the baseline corpus
Excite	Discussion or actions related to a community or topic that cause the reader to experience a positive emotion such as joy, happiness, liking, or excitement	Target packet emotional valence will be higher in happiness and surprise / changes positively over the time more than the baseline corpus
Explain	Discussion or actions that clarify a topic to the targeted community or actor often by providing details on, or elaborations on, the topic	Topic specialization – additional jargon and net shift towards same stance / changes over the time more than the baseline corpus
Engage	Discussion or actions that increase the relevance of the topic to the reader often by providing anecdotes or enabling direct participation and so suggesting that the reader can impact the topic or will be impacted by it	Topics proportional representation / changes positively over the time more than the baseline corpus
Enhance	Discussion or actions that provide material that expands the scope of the topic for the targeted community or actor often by making the topic the master topic to which other topics are linked	Increased linkage and centrality, betweenness / changes positively over the time more than the baseline corpus
Negate	Discussion or actions that decrease the actual, or the appearance of, an actor's importance or effectiveness relative to a community or topic	Centrality of node / changes negatively over the time more than the baseline corpus
Neutralize	Discussion or actions that cause a group to be, or appear to be, no longer of relevance, e.g., because it was dismantled	Group nodes have more in common with other groups than themselves (group disappears)
Narrow	Discussion or actions that lead a group to be, or appear to be, more specialized, and possibly to fission, or appear to fission, into two or more distinct groups	Multiple groups where only one was present before, fewer links on bipartite network from meta-agent group node to topic/stance nodes
Neglect	Discussion or actions that decrease the size of a group and/or the connections among group members, or the appearance of such	Density and/or size / changes negatively over the time above the baseline corpus
Dismay	Discussion or actions related to a community or topic that cause the reader to experience a negative emotion such as worry, sadness, disliking, anger, despair, or fear	Target packet emotional valence will be higher in anger, sadness, fear / changes over the time more than the baseline corpus
Distort	Discussion or actions that obscure a topic to the targeted community or actor often by supporting a particular point of view or calling details into question	Topic specialization – additional jargon and net shift towards opposite stance / changes over the time more than the baseline corpus
Dismiss	Discussion or actions that decrease the relevance of the topic to the reader often by providing stories or information that suggest that the reader cannot impact a topic or be impacted by it	Topics proportional representation / changes negatively over the time more than the baseline corpus
Distract	Discussion or actions that redirect the targeted community or actor to a different topic often by bring up unrelated topics, and making the original topic just one of many	Decreased linkage and centrality, betweenness / changes over the time more than the baseline corpus

Figure 3.4: BEND definitions to effects mapping.

pus. A between-corpus comparison will hopefully provide insight into the overall effectiveness of the BEND maneuvers.

Importantly, I will then use these global measurements as a baseline measurement to compare specific target groups against. I will conduct both the CUE+ based maneuver identification and effects-based detection again but now against these specific target groups. By normalizing these results against the corpus I hope to eliminate broader crossfire effects and retaining only those effects and messages specifically targeted at the group. This should provide more granular measurements of the effectiveness of BEND maneuvers. Additionally, this will enable me to link BEND effects to a general set of BEND maneuvers. Although, this will require the results of time-lag experiments - essentially providing the flash-to-bang average of BEND maneuver execution to BEND maneuver effect - that are being conducted elsewhere by Jacob Shaha.

Study 2: Simulation

This thesis proposes two simulations in order to increase our knowledge of BEND maneuver effects and their interactions. The first simulation represents work previously completed as a part of BEND Battle: An Agent Based Simulation of Social-Cyber Maneuvers - a currently unpublished paper. The second simulation is a complete reworking away from NetLogo and into Construct.

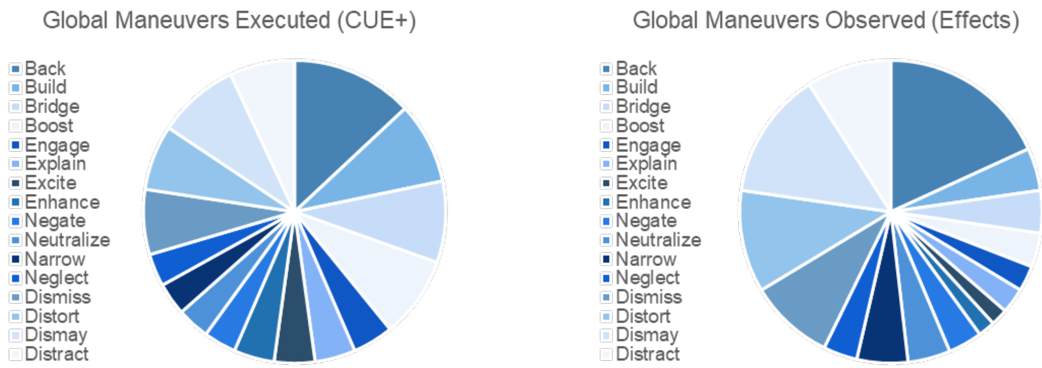

Figure 3.5: Global BEND maneuvers identified compared against the global BEND maneuver effects detected.

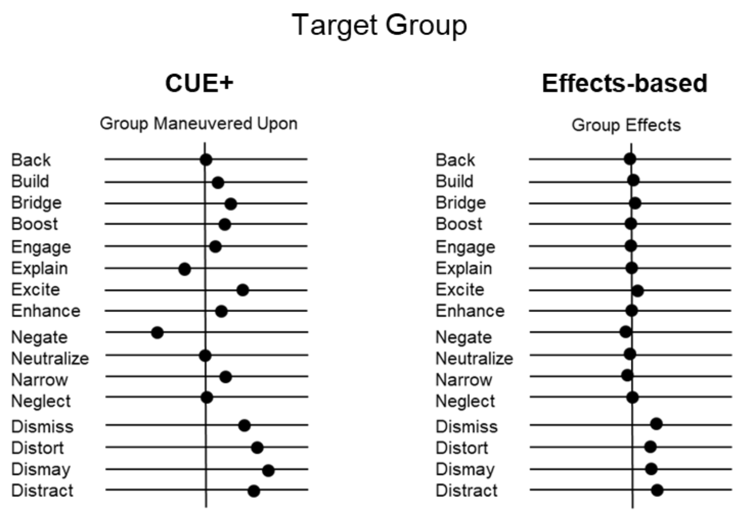

Figure 3.6: The total number of BEND maneuvers targeting a group (above baseline) compared to the level of BEND maneuver effects detected (above baseline)

BEND Battle: An Agent Based Simulation of Social-Cyber Maneuvers

BEND maneuver comparisons and interactions suffer from many of the same problems as tasks in other military domains. Attempts to compare the interaction of military tactical tasks fails due to the overwhelming influence of all the other factors involved. In the US Army, this is called METT-TC. From Army Field Manual 3-21.8, METT-TC stands for Mission, Enemy, Terrain/Weather, Troops available, Time available, and Civilian considerations [24]. When planning a mission, all these factors influence the success or failure of the planned task - often more than the form of the task selected.

For an even simpler example, while an attacker may generally consider a 3:1 advantage in combat power sufficient to conduct an attack, if the defender occupies especially favorable terrain, then they might require a 5:1 or even a 10:1 advantage. Other factors almost always matter more than the maneuver.

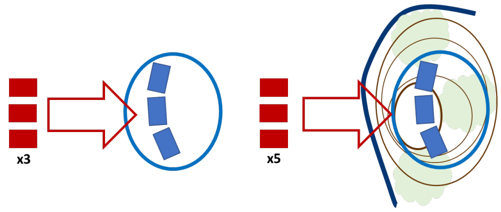

Figure 3.7: Generally an attacker requires a 3:1 advantage to succeed (left), but given additional factors - high ground, a water obstacle, dense forests, etc. - the ratio may increase (right).

However, this is not to suggest that any attempt at comparison is fruitless. It is important to know that, generally speaking, attackers require a 3:1 advantage – just as it would be important to know that Dismay is more effective than Engage (or vice versa). In order to make such comparisons the confounding factors need to be equalized.

BEND Battle is an agent-based model that simulates social media users conducting BEND social-cyber maneuvers in order to influence each other on a single topic. The focus of BEND Battle is on providing a visualization of two sides simultaneously conducting maneuvers against each other and the effects of those maneuvers. The winner is determined by evaluating which side used BEND maneuvers to effectively retain agents of their own and pull opposing agents over to their side.

BEND Battle is built on the NetLogo simulation platform [45]. The game space within NetLogo for BEND Battle is comprised of two opposite stances on a single topic – denoted by red and blue squares at the far left and right of the model visualization. Between the two stance squares, users - depicted by circles - compete to push and pull each other towards one stance or the other through the application of BEND maneuvers. The environment is not Twitter, Reddit, or YouTube; rather, the environment is a broadcast information transmission medium through which users enact BEND effects upon other users.

The entities being modeled are individual users of social media. In order to facilitate the visual understanding of BEND maneuver effects, BEND Battle uses four primary attributes for all users (shown in Fig. 3.8.): strength, topicality, affiliation and leadership.

Strength is a measure of a user's stance on the topic. It is visualized on the x-axis of BEND Battle, ranging from -100 (fully anti-stance) to 100 (fully pro-stance).

Topicality is a measure of the exclusivity of a user's engagement with the topic. Because BEND Battle is meant to be a simplified, normalized encounter on a single topic, all engagement with other topics is amalgamated into a single attribute per user - topicality. This can be thought of as the overall focus of a user on this particular topic vice other topics.

Together strength and topicality - variation of stance engagement and variation of topic engagement - determine the overall engagement of a user. The stronger they feel about their stance and the more focused they are on the topic - the more likely they are to act each tick.

Leadership is a measure of the influence a user has over the BEND battlefield. It is visualized by the overall size of the user ranging from 0 to 100.

Affiliation is a measure of how influenceable an agent is (visualized by color saturation). Affiliation is used here for disambiguation and to address a number of related concepts. Generally, BEND maneuvers that would typically target network structure - Bridge, Narrow, etc. - and in-

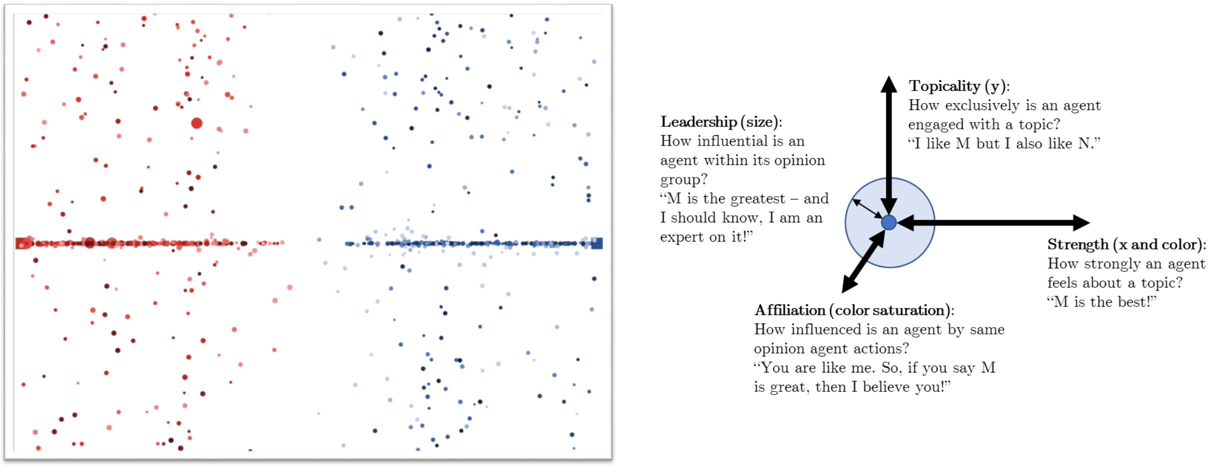

Figure 3.8: BEND Battle visualization of agents is based on four attributes.

fluence homophily, cohesion, or related measures, instead affect the affiliation attribute of users within BEND Battle.

Each BEND maneuver has an effect on one or more of the user attributes - strength, topicality, affiliation, and leadership (STAL). Because one of the primary motivations behind BEND Battle is to normalize the effects of maneuvers with respect to one another, each of the maneuvers is given three points to allot towards changing these attributes. The English language definition of each BEND maneuver has been translated into effects on STAL with a gross magnitude of three. A full layout of all maneuver effects is provided in Table 3.1.

BEND Battle uses a simple three step algorithm for resolving the interplay of user initiated BEND maneuvers. These three steps are check-action, take-action, and resolve-action. All three steps are run each tick (unit of time/iteration within NetLogo).

check-action: All users determine if they are going to act this tick based upon the distance from the user to their stance box.
take-action:
A) Acting users determine which BEND maneuver they will execute this tick. The maneuvers are chosen based upon the probabilities given for each of the BEND maneuvers.
B) Acting users determine which other users will be the target of their BEND maneuver.
C) All acting users conduct their BEND maneuvers in a random order.
resolve-action: All users that were affected by BEND maneuvers appropriately change their position, color, and size - they also recalculate their chances for acting next turn.

In order to visually represent the variations across the runs, for each set of variables I subtracted the number of runs that red won from the number of runs that blue won. Thus, a 100 represents that blue won every iteration for that combination of variables, a -100 means that red won every iteration, and a 0 indicates that they either tied every iteration or the number of red wins equaled the number of blue wins. In general, positive values indicated that blue is winning more often, whereas negative values indicated that red is winning more often. These results are mapped in Fig. 3.9 and Fig. 3.10.

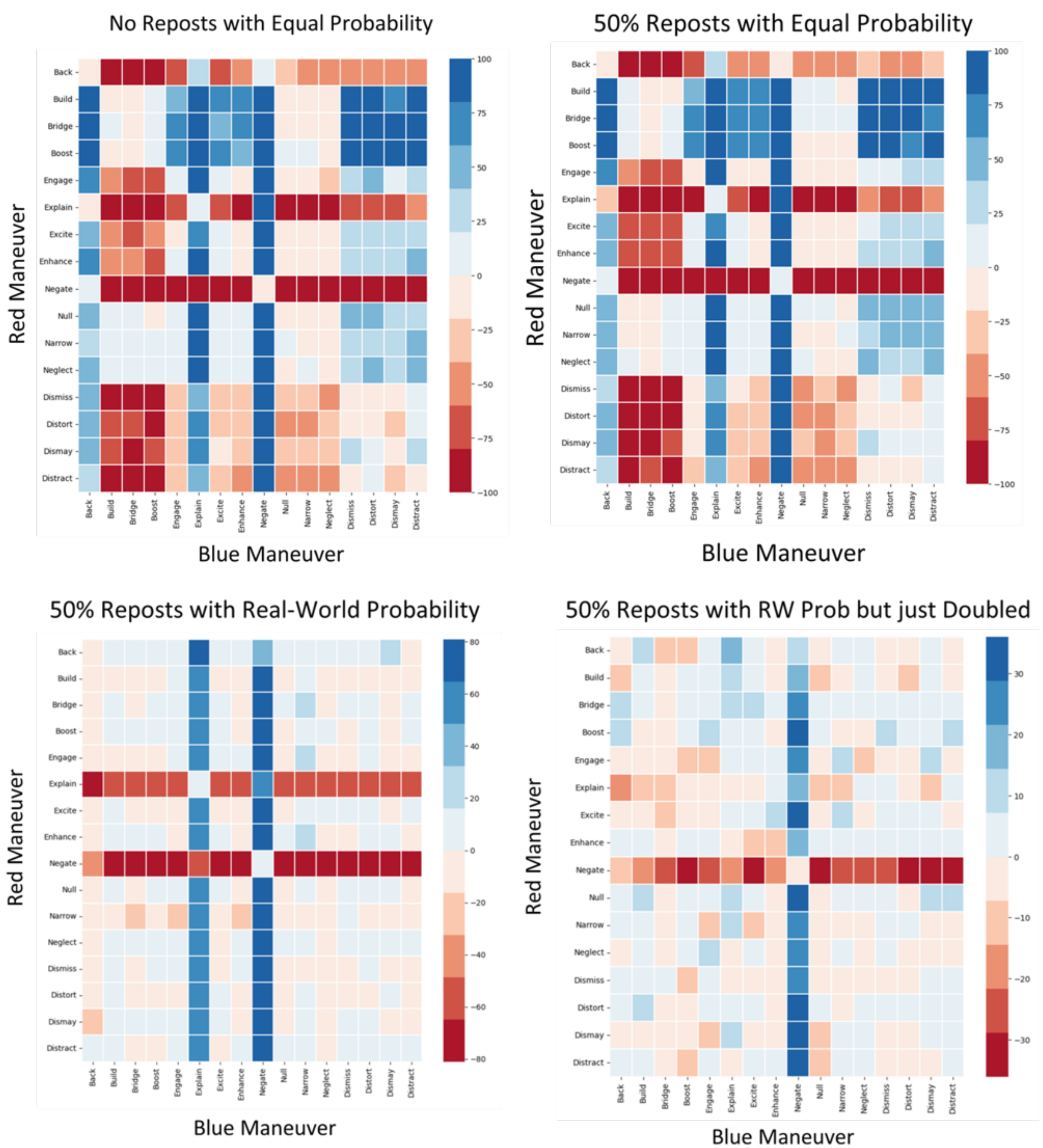

Figure 3.9: Intial results of BEND Battle head-to-head maneuver match-ups. Real-world probabilities were taken from Alieva, et al. 2022.[10]

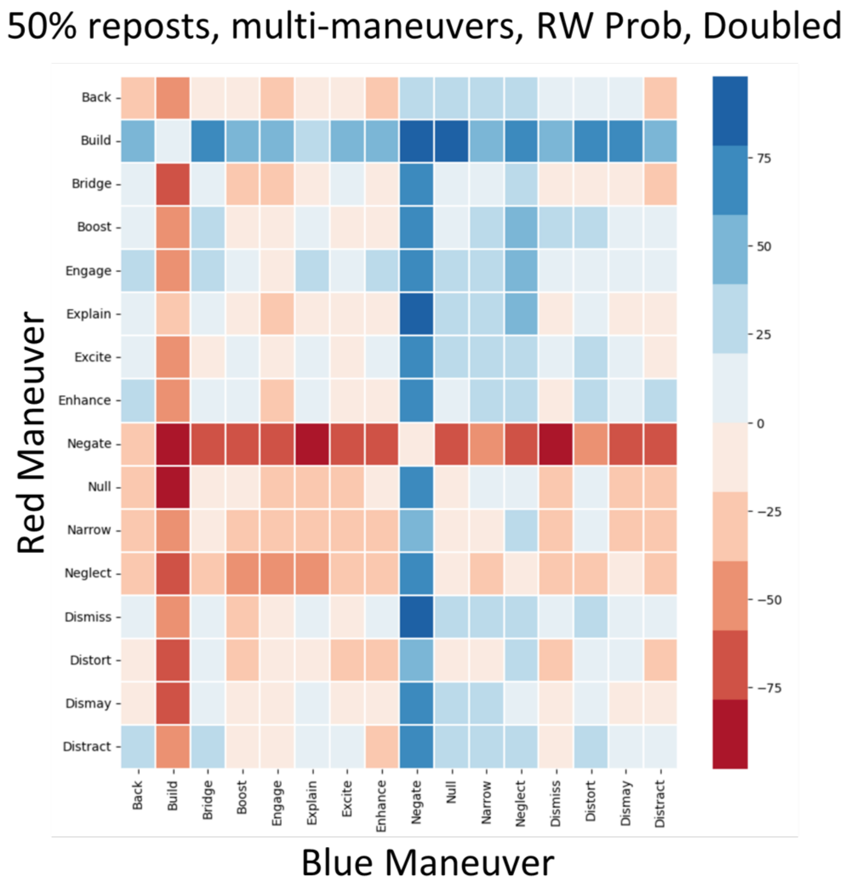

Figure 3.10: The most comprehensive BEND Battle head-to-head maneuver match-up results. These iterations used real-world maneuver probabilities[10], 50% reposts, allowed for multiple maneuvers per post, and the selected maneuvers for red and blue had their real-world probabilities only doubled.

Table 3.1: BEND Maneuver Effects

	Friendly				Enemy			
	Strength	Topicality	Affiliation	Leadership	Strength	Topicality	Affiliation	Leadership
Engage	++	+			- -	-		
Explain	+	+		+*	-	-		
Excite	+	+	+		-	-	-	
Enhance	+++				- - -			
Dismiss	+++				- - -			
Distort	+	+	+		-	-	-	
Dismay	++		+		- -		-	
Distract		+++				- - -		
Back				+++				
Build	+		++		-		- -	
Bridge	+		++		-		- -	
Boost	+		++		-		- -	
Negate								- - -
Null	+		++		-		- -	
Narrow	+		++		-		- -	
Neglect	+		++		-		- -	

This thesis will compare the results from this simulation to the results for the real-world study discussed previously.

Construct: BEND Effects

The intentional reduction of the network dimension to a single variable within BEND Battle limits its effectiveness in evaluating BEND maneuvers. In order to move beyond the Euclidean world posited in BEND Battle, this thesis proposes a Construct simulation with similar goals. Construct will allow for a more complete agent-based simulation by also simulating the network.

3.3 AI-Enabled BEND Training Scenarios

3.3.1 Research Questions

Understanding the BEND framework within existing doctrine, identifying BEND maneuvers and their effects, and even conceptually understanding counter-maneuvers falls short of fully operationalizing BEND. What is required is a realistic training on a realistic corpus where a training audience can apply these concepts.

The key research questions for this chapter are:

- How can we develop exercise training scenarios for the BEND framework?
- Can we extract a training scenario from real data without resorting to hand-crafting messages?
- How can we leverage AI/LLMs to enhance training scenarios and generate multi-modal BEND maneuvers?

Table 3.2: 16x16 Example Virtual Experiment

Independent Variables	# Test Cases	Values Used
Build	2	.7/.9
Bridge	2	.2/.4
Boost	2	.35/.55
Back	2	.25/.45
Null	2	.16/.36
Negate	2	.15/.35
Neglect	2	.05/.25
Narrow	2	.1/.3
Distract	2	.2/.4
Distort	2	.15/.35
Dismiss	2	.26/.46
Dismay	2	.25/.45
Explain	2	.36/.56
Excite	2	.3/.5
Enhance	2	.3/.5
Engage	2	.4/.6
Control Variables	**# Test Cases**	**Values Used**
Time Periods	1	500 ticks
Red/Blue Starting Entities	1	500
In/Out Percent / Time Period	1	0.01
Dependent Variables		**Values Expected**
# of agents on the red team		0-500
# of agents on the blue team		0-500

16x16	256 cells
Replications per cell	100
25,600 total runs (12,800,000 data points)	

The idea of realistic training for social-cyber security is not new. Project OMEN has been steadily increasing the complexity of training scenarios for the DoD since 2021[39], culminating in February 2023 with a blend of hand-altered real-world data and synthetically generated data. This mix of data is a result of increasing demands for reconfigurability, scalability, and realism in networks and narrative.

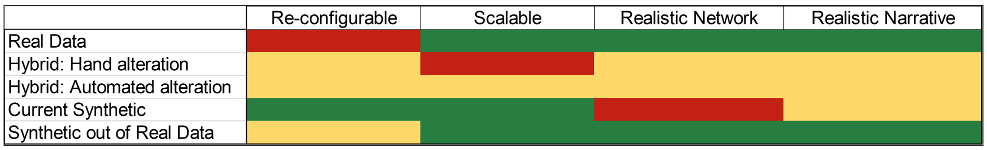

Figure 3.11: Dataset capability by category.

Real data is inflexible in its ability to adjust to changing training requirements or objectives but is scalable (just collect more) and realistic from both a network and narrative perspective. Hand altered real data is somewhat reconfigurable to meet training demands but is certainly not scalable - it takes too long to alter large amounts of data - and the altering often reduces network and narrative realism. Automated alteration retains the benefits of hand altering but adds the ability to scale. Observed current synthetic environments are imminently reconfigurable and scalable but output unrealistic networks and weak narratives. These trade-offs are displayed in Fig. 3.11.

Reconfigurable and scalable seem straightforward, but where do these evaluations of realism for networks and narrative come from? They come from MOMUS. MOMUS is a Netanomics product that takes in a synthetic corpus and compares it with amalgamated real data from the same platform across a wide range of network and semantic metrics. Momus then provides a score based on the similarity between the synthetic generated and the real data.

In order to better understand how a training data is created it is helpful to look at the overall training flow.

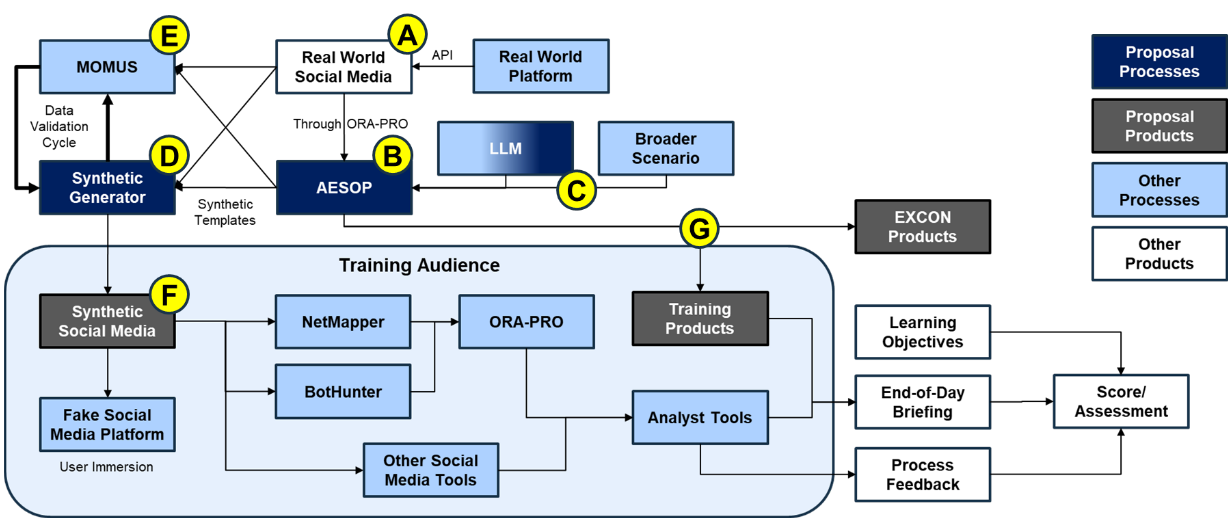

Figure 3.12: Project OMEN training flow.

In order to better understand how a training data is created it is helpful to look at the overall training flow. Fig. 3.12 depicts the Project OMEN training flow.

- A) Real-world data is pulled from a social media platform using approved APIs and without violating the terms of service.
- B) Important network and narrative characteristics are drawn out of the real-world social media data using ORA-PRO. These characteristics are passed to a scenario planning tool called AESOP (AI-Enabled Scenario Orchestration and Planning).
- C) A scenario planner uses AESOP in conjunction with their understanding of the training objectives of the training audience to construct a scenario. AESOP leverages large language models to fill in the gaps between the real-world characteristics and the desired scenario.
- D) AESOP outputs templates for actors, groups, events, narratives, and new stories that are passed to the synthetic generator. These templates represent both the ”needles” and the ”haystack” for the training training scenario. The synthetic generator uses these templates to generate social media traffic - ostensibly creating the ”needles” and hiding them in the ”haystack.”
- E) The synthetic social media is passed to MOMUS for scoring - there is a cycle of validation as the synthetic generator iterates against the MOMUS score. This is currently a manual process and not an automatic feedback loop.
- F) The synthetic social media is given to the training audience for training. The training audience uses tools to analyze the data - hopefully finding the ”needles” within the ”haystack.”
- G) AESOP also outputs two sets of documents along with the scenario templates. The two sets of documents are a list of events and hosted websites for exercise control personnel and a set of baseline documents for the training audience that act as a breadcrumbs to orient them to the ”haystack.”

Included in this thesis is work done on the scenario planning tool (AESOP), the synthetic generator, and portions of the large-language models. Outputs from work done as a part of this thesis include the synthetic social media corpus, products for the training audience, and products for the exercise controllers.

AI-Enabled Scenario Orchestration and Planning (AESOP) Tool

AESOP allows Information Environment planners to develop social-cyber exercise scenarios from scratch or develop social-cyber vignettes for integration with existing scenarios. It is a standalone GUI coded in Python with PySide6.[44]

AESOP leverages large language models (LLMs) to reduce planner load and increase realism and immersion for the training audience. Planners complete basic fields – such as date ranges and summaries – and AESOP develops an engineered prompt for a configurable LLM that is used to generate surrounding details. Planners can make additional changes to the prompt as required. Planners can also freely manipulate the details returned by the LLM. By default, AESOP reaches out to the OpenAI API[7] and uses a GPT3.5-Turbo model (gpt-3.5-turbo-0125) for text

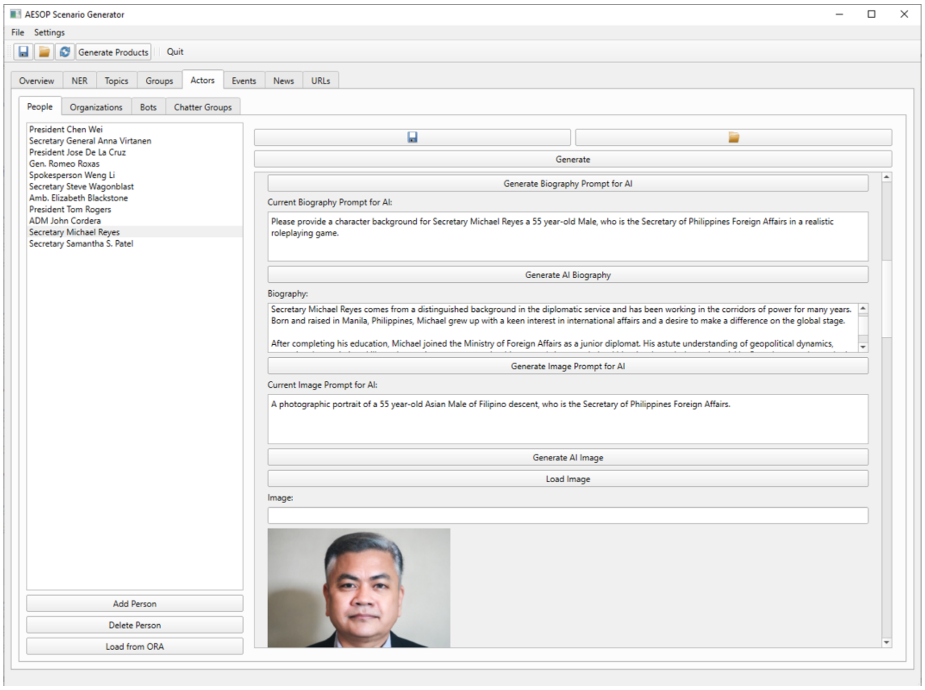

Figure 3.13: AESOP GUI Example. With an actor template being worked on.

and DALL-E model (dall-e-2) for images. However, it is configurable to run against any LLM provider that has an OpenAI compatible API - such as oogabooga/text-generation-webui.[41]

Through AESOP, planners create groups and develop group stances on topics via narratives. These are linked to events as well as news/urls. From this, planners fill out groups with actors that can be associated with multiple accounts on various social Media platforms. The

AESOP provides scaffolding for developing actor templates suitable for hand-off to a synthetic data generation entity. These templates include required information for Twitter/X and Telegram social media platforms.

Additionally, AESOP allows planners to develop US military fragmentory orders (FRAGOs), intelligence summaries (INTSUMs), and Press Releases to support stimulating the training audience. Upon scenario generation, all of these are output as Word Documents (.docx) and organized by day for exercise control distribution to distribute to the training audience. Once again, LLMs are leveraged to create these injects but planners are free to change prompts or write their own.

Finally, AESOP integrates with ORA-PRO through ORA-PRO's Node of Interest Characterization report feature. This allows for the export of characterization data for an actor from an existing corpus. Planners can then load this data into the Scenario Generator and data fields will fill appropriately.

A full list of AESOP outputs can be found in Fig. 3.15.

Synthetic Generation (Telegram)

Currently, synthetic generation only occurs for Telegram data. The Telegram generator accepts the synthetic templates from AESOP and then builds the appropriate Telegram channels and ac-

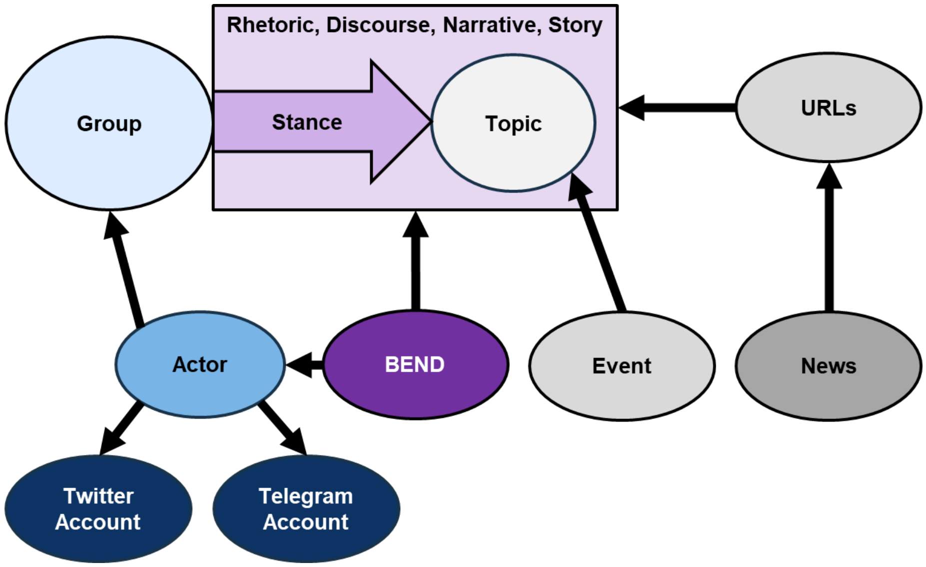

Figure 3.14: AESOP Relational Diagram. Blue is loosely defined as the WHO, purple is DOES/FEELS, and gray is the WHAT. The diagram depicts WHO DOES/FEELS to/about WHAT.

counts before populating them with messages. Message population is done using configurable calls to an external LLM - similar to AESOP the default is OpenAI but configurable to a local LLM. In general, if the topic matter or BEND maneuvers being requested are negative, then a second pass with a local LLM is required for message generation as the OpenAI models will baulk without excessive prompt engineering. The final output of the Telegram generator is JSONL and generally indistinguishable in format to data pulled using the native Telegram API.

There are two main approaches to synthetic generation - top-down and bottom-up. In the top-down approach, I create the desired heterogenous/multimodal network fabric based on real data and then fill that fabric with appropriate actors, messages, topics that match nodes and link types. Alternatively, using the bottom-up approach, I create detailed social media agents and program agent interaction from first principles. I then allow the social media agents to interact and hope for emergent networks and narratives that are realistic and relevant. Current Telegram generation focuses on the bottom-up approach. However, top-down approaches may be more appropriate for other social media.

The bottom-up flow for telegram generation of a single channel/actor node can be found in Fig. 3.16

3.3.2 Proposed Work

AESOP

- Create synthetic generation templates for noise and background messaging.
- Create additional exercise control products that provide training objectives and grading rubrics.

Synthetic Generation

- Enhance Telegram synthetic generation to more closely match real world interactions between channels.
- Extend synthetic generation to include multi-modal messaging.
- Extend synthetic generation to Twitter/X.
- Extend synthetic generation to Facebook.
- Pass MOMUS grading report card
- Provide feedback loop for training audience input into synthetic generation environment.

Feature	Archive File	Synthetic Template	EXCON	Participants
Agents				
People	JSON	JSON	DOCX	DOCX
Organizations	JSON	JSON	DOCX	DOCX
Bots	JSON	JSON	DOCX	DOCX
Telegram				
Account	JSON	JSON	JSONL	JSONL
Channel	JSON	JSON	JSONL	JSONL
Messages	JSON	JSON	JSONL	JSONL
Twitter/X				
Account	JSON	JSON	JSONL	JSONL
Messages	JSON	JSON	JSONL	JSONL
Facebook				
Account	JSON	JSON	JSONL	JSONL
Page	JSON	JSON	JSONL	JSONL
Messages	JSON	JSON	JSONL	JSONL
Groups	JSON	JSON	DOCX	
Topics	JSON	JSON	DOCX	
Events				
Event Summary	JSON	JSON	DOCX	DOCX
Fragmentary Orders	JSON		DOCX	DOCX
Press Releases	JSON		DOCX	DOCX
Intelligence Reports	JSON		DOCX	DOCX
News				
News Agency	JSON		HTML	HTML
Real News Articles	JSON		HTML	HTML
Disinformation Articles	JSON		HTML	HTML
Pink Slime Articles	JSON		HTML	HTML
URL Listings	JSON	JSON		
Scenario Overview				
MSEL			XLSX	
MOPs			XLSX	
TTP Evaluation Guidance			DOCX	
Executive Summary				
Exercise Roles				
OMEN Research Goals				
Mission	JSON		DOCX	DOCX
Commander's Guidance	JSON		DOCX	DOCX
Scenario Overview	JSON		DOCX	DOCX
Scenario Description	JSON		DOCX	DOCX

Figure 3.15: AESOP outputs. Green represents complete, yellow represents undergoing changes, and red represents not yet completed.

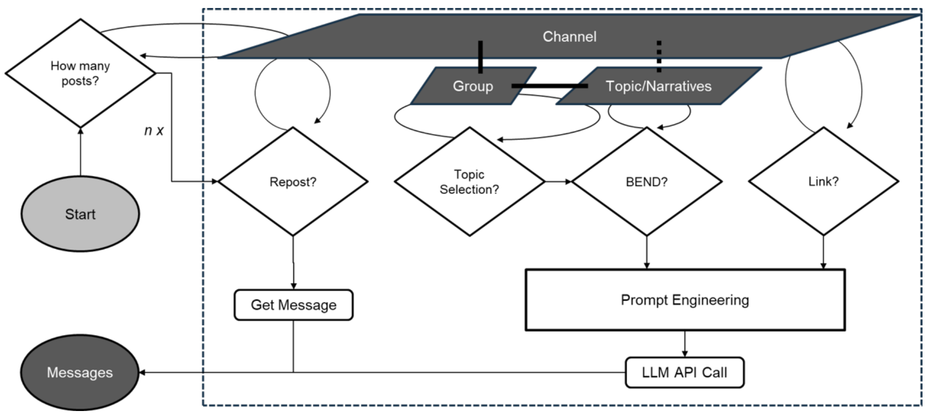

Figure 3.16: Telegram generator message creation flow. Dark gray parallelograms are synthetic templates provided by AESOP and referenced by the generator.

Chapter 4

Contributions

4.1 Theoretical Contributions

The theoretical contributions of this thesis include a synthesis of US military doctrine with the BEND maneuver framework and a re-framing of how BEND maneuvers are defined. US military Information Operations doctrine does not currently apply the same level of rigor to social media analysis as other domains do to their subsystem components. This imbalance prevents leaders from placing the proper emphasis - little as it may be - on social media. The refined Social Media MCOO/CSO outlined in this thesis rectifies this situation and provides the necessary theory to bridge the gap between current doctrine and social media.

Currently, BEND maneuvers are defined solely by the intent of their executors. While CUE+ methods can attempt to guess at the intent and thereby identify BEND maneuvers, returning BEND to be effects-based not only allows us to measure the effectiveness of maneuvers but also brings them in line with the military tactical tasks for which they are an analog.

4.2 Methodological Contributions

This thesis contributes several methodological contributions. First, it introduces methods for BEND effects-based detection. In conjunction with prior CUE+ work, this not only allows for a more comprehensive detection mechanism but also provides a way to measure the effectiveness of BEND Maneuvers.

Second, it introduces a method for the extraction of network and narrative characteristics from existing corpora. This allows for tools that provide AI-enabled scenario development based on these characteristics.

Lastly, it provides a method for passing extracted characteristics to a synthetic generator in the form of templates. These detailed templates enable Twitter/X, Facebook, and Telegram social media generation for exercise training based on extracted characteristics from existing corpora.

4.3 Limitations

US Military Doctrinal Synthesis US Information Operations doctrine is evolving and changing rapidly. Many of the referenced Joint and Service Publications are already out of date and the replacement publications are all held at a classified level or have distribution restrictions that prohibit their academic study. Additionally, Information Operations remains a complex issue with authority and titling problems that cannot be resolved in theory and require policy reforms.

Effects-Based BEND Detection There is currently no way to directly associate observed BEND effects with any single message BEND maneuver - we are not yet in the precision munitions phase of information environment maneuvers. Also, better methods for measuring BEND maneuvers above baseline are required as residual statistics will be more important than net maneuver counts. Standard ORA-PRO reports do not reflect this need.

BEND Scenario Development Without an overarching simulation, training scenario data will be static and unresponsive to training audience feedback. However, AESOP could be used to alter the scenario based upon training audience decisions and new templates could then drive additional synthetic data to get after a highly incremented simulation. Daily static training data is reasonable and appropriate since collection and attribution methods through the social media APIs do not allow for pulling all possible data instantly and continuously.

Chapter 5

Timeline

Figure 5.1 shows my proposed timeline from April 2024 through my projected thesis defense in April 2025.

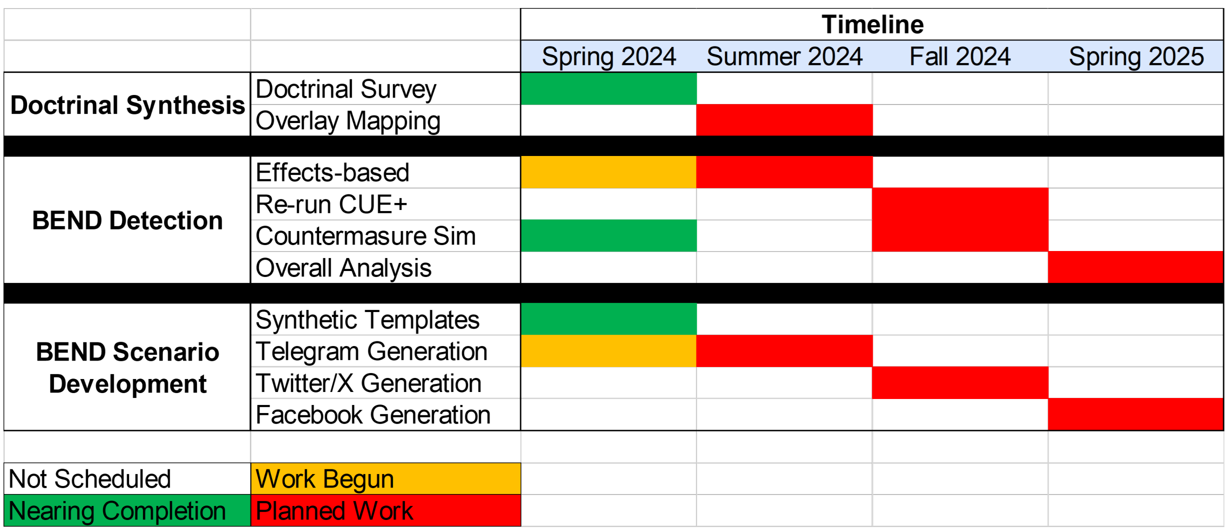

Figure 5.1: Proposed Timeline

Bibliography

[1] *U. S. Const. amend. IV*. URL https://www.senate.gov/about/origins-foundations/senate-and-constitution/constitution.htm. 3.1.2

[2] *Privacy Act of 1974, 5 U.S. Code Title 5, section 552a*. URL https://www.govinfo.gov/content/pkg/USCODE-2018-title5/pdf/USCODE-2018-title5-partI-chap5-subchapII-sec552a.pdf. 3.1.2

[3] *Posse Comitatus Act, U.S. Code Title 18, section 1305*. URL https://www.law.cornell.edu/uscode/text/18/1385. 3.1.2

[4] *USAF Doctrine Update on Domains and Organizing for Joint Operations*. URL https://www.doctrine.af.mil/Portals/61/documents/doctrine_updates/du_13_09.pdf. 3.1.2

[5] *Summary of the Joint All-Domain Command and Control (JADC2) Strategy*. URL https://www.doctrine.af.mil/Portals/61/documents/doctrine_updates/du_13_09.pdf. 3.1.2

[6] mistralai/mixtral-8x7b-v0.1 · hugging face. URL https://huggingface.co/mistralai/Mixtral-8x7B-v0.1. 2.2

[7] The official python library for the OpenAI API. URL https://github.com/openai/openai-python. original-date: 2020-10-25T23:23:54Z. 2.2, 3.3.1

[8] SGM Alexander E. Aguilastratt and SGM Matthew S. Updike. The information domain and social media. URL https://www.moore.army.mil/infantry/magazine/issues/2022/Spring/PDF/6_ProfessionalForum.pdf. 3.1.2

[9] Alexandre Alaphilippe. Adding a d to the ABC disinformation framework. URL https://policycommons.net/artifacts/4139892/adding-a-d-to-the-abc-disinformation-framework/4948112/. Publisher: Brookings Institution. 1.2

[10] Iuliia Alieva, J. D. Moffitt, and Kathleen M. Carley. How disinformation operations against russian opposition leader alexei navalny influence the international audience on twitter. 12 (1):80. ISSN 1869-5450, 1869-5469. doi: 10.1007/s13278-022-00908-6. URL https://link.springer.com/10.1007/s13278-022-00908-6. 1.2, 3.9, 3.10

[11] Adam Badawy, Aseel Addawood, Kristina Lerman, and Emilio Ferrara. Characterizing the 2016 russian IRA influence campaign. 9(1):31. ISSN 1869-5450, 1869-5469.

doi: 10.1007/s13278-019-0578-6. URL http://link.springer.com/10.1007/s13278-019-0578-6. 1.1

[12] David M. Beskow and Kathleen M. Carley. Social cybersecurity: an emerging national security requirement. 99(2):117–127. URL https://apps.dtic.mil/sti/citations/AD1108494. 1.2

[13] Janice Blane. Social-cyber maneuvers for analyzing online influence operations. page 13152760 Bytes. doi: 10.1184/R1/22825112.V1. URL https://kilthub.cmu.edu/articles/thesis/Social-Cyber_Maneuvers_for_Analyzing_Online_Influence_Operations/22825112/1. 1.2, 1.1, 2.1, 2.1, 3.2.1, 3.2.2, 3.2.3

[14] Janice T Blane, Daniele Bellutta, and Kathleen M Carley. Social-cyber maneuvers during the COVID-19 vaccine initial rollout: Content analysis of tweets. 24(3):e34040. ISSN 1438-8871. doi: 10.2196/34040. URL https://www.jmir.org/2022/3/e34040. 1.2, 2.1

[15] Sam Blazek. SCOTCH: A framework for rapidly assessing influence operations. URL https://www.atlanticcouncil.org/blogs/geotech-cues/scotch-a-framework-for-rapidly-assessing-influence-operations/. 1.2

[16] Kathleen M. Carley. Social cybersecurity: an emerging science. 26(4):365–381. ISSN 1572-9346. doi: 10.1007/s10588-020-09322-9. URL https://doi.org/10.1007/s10588-020-09322-9. 1.2

[17] Kathleen M. Carley, Guido Cervone, Nitin Agarwal, and Huan Liu. Social cybersecurity. In Robert Thomson, Christopher Dancy, Ayaz Hyder, and Halil Bisgin, editors, *Social, Cultural, and Behavioral Modeling*, volume 10899, pages 389–394. Springer International Publishing, . ISBN 978-3-319-93371-9 978-3-319-93372-6. doi: 10.1007/978-3-319-93372-6_42. URL http://link.springer.com/10.1007/978-3-319-93372-6_42. 1.2

[18] L Richard Carley, Jeff Reminga, and Kathleen M Carley. Ora & netmapper. In *International conference on social computing, behavioral-cultural modeling and prediction and behavior representation in modeling and simulation. Springer*, volume 3, page 7, . Issue: 3.3. 1.2, 2.2, 2.2

[19] *Joint Information Operations Proponent*. Chairmain of the Joint Chiefs of Staff, Washington, DC, USA, 2014. URL https://www.jcs.mil/Portals/36/Documents/Library/Instructions/3210_01.pdf. 3.1.2

[20] Fred Cohen. Simulating cyber attacks, defences, and consequences. 18(6):479–518. ISSN 01674048. doi: 10.1016/S0167-4048(99)80115-1. URL https://linkinghub.elsevier.com/retrieve/pii/S0167404899801151. 1.2

[21] Robert Cordray III and Marc J. Romanych. Mapping the information environment. pages 7–10. URL https://www.quantico.marines.mil/Portals/147/Docs/MCIOC/IORecruiting/MappingtheInformationEnvironmentIOSphereSummer2005.pdf. 3.1.2

[22] *Information Operations (IO)*. Department of Defense, Washington, DC, USA, 2017. URL https://www.esd.whs.mil/Portals/54/Documents/DD/issuances/dodd/360001p.pdf. 3.1.2

[23] *Information in Air Force Operations*. Department of the Air Force, Washington, DC, USA, 2018. URL https://www.doctrine.af.mil/Portals/61/documents/AFDP_3-13/3-13-AFDP-INFO-OPS.pdf. 1.2, 3.1.2

[24] *The Infantry Rifle Platoon and Squad*. Department of the Army, Washington, DC, USA, 2007. URL https://www.marines.mil/Portals/1/Publications/FM3-21.8TheInfantryRiflePlatoonandSquad_1.pdf. 3.2.3

[25] *The Operations Process*. Department of the Army, Washington, DC, USA, 2019. URL https://armypubs.army.mil/epubs/DR_pubs/DR_a/ARN18126-ADP_5-0-000-WEB-3.pdf. 3.1.2

[26] *Intelligence Preparation of the Battlefield*. Department of the Army, Washington, DC, USA, 2019. URL https://irp.fas.org/doddir/army/atp2-01-3.pdf. 1.2, 3.1.2

[27] *Army Information*. Department of the Army, Washington, DC, USA, 2023. URL https://irp.fas.org/doddir/army/adp3-13.pdf. 3.1.2

[28] *Navy Information Operations*. Department of the Navy, Washington, DC, USA, 2014. URL https://www.usna.edu/Training/_files/documents/References/3C%20MQS%20References/2015-2016%203C%20MQS%20References/NWP%203-13_Information%20Operations_FEB2014.pdf. 3.1.2

[29] Stephen Dipple and Kathleen M. Carley. *Construct User Guide*. URL https://www.cmu.edu/casos-center/publications/cmu-s3d-23-104.pdf. 2.2

[30] Geoffrey B. Dobson and Kathleen M. Carley. Cyber-FIT: An agent-based modelling approach to simulating cyber warfare. In Dongwon Lee, Yu-Ru Lin, Nathaniel Osgood, and Robert Thomson, editors, *Social, Cultural, and Behavioral Modeling*, volume 10354, pages 139–148. Springer International Publishing. ISBN 978-3-319-60239-4 978-3-319-60240-0. doi: 10.1007/978-3-319-60240-0_18. URL https://link.springer.com/10.1007/978-3-319-60240-0_18. 1.2

[31] Camille François. Actors, behaviors, content: A disinformation ABC. URL http://cdn.annenbergpublicpolicycenter.org/wp-content/uploads/2020/05/ABC_Framework_TWG_Francois_Sept_2019.pdf. 1.2

[32] *Marine Air-Ground Task Force Information Operations*. Headquarters United States Marine Corps, Department of the Navy, Washington, DC, USA, 2018. URL https://www.marines.mil/Portals/1/Publications/MCWP%203-32.pdf. 3.1.2

[33] *Joint Targeting*. Joint Chiefs of Staff, Washington, DC, USA, 2013. URL https://jfsc.ndu.edu/Portals/72/Documents/JC2IOS/Additional_Reading/1F4_jp3-60.pdf. 3.1.2, 3.1.3

[34] *Joint Intelligence Preparation of the Operational Environment*. Joint Chiefs of Staff, Wash-

ington, DC, USA, 2014. URL https://irp.fas.org/doddir/dod/jp2-01-3.pdf. 3.1.2, 3.1.3

[35] *Information Operations*. Joint Chiefs of Staff, Washington, DC, USA, 2014. URL https://irp.fas.org/doddir/dod/jp3_13.pdf. 1.2, 3.1.2

[36] *Public Affairs*. Joint Chiefs of Staff, Washington, DC, USA, 2016. URL https://irp.fas.org/doddir/dod/jp3_61.pdf. 3.1.2

[37] *Doctrine for the Armed Forces of the United States*. Joint Chiefs of Staff, Washington, DC, USA, 2017. URL https://fas.org/irp/doddir/dod/jp1.pdf. 3.1.2

[38] *Joint Operations*. Joint Chiefs of Staff, Washington, DC, USA, 2017. URL https://irp.fas.org/doddir/dod/jp3_0.pdf. 3.1.3

[39] Catherine King, Christine Sowa Lepird, and Kathleen M. Carley. Project OMEN: Designing a training game to fight misinformation on social media. 3. URL http://reports-archive.adm.cs.cmu.edu/anon/anon/usr0/ftp/home/ftp/isr2021/CMU-ISR-21-110.pdf. 3.1.3, 3.3.1

[40] Ben Nimmo. Anatomy of an info-war: How russia's propaganda machine works, and how to counter it. 15:1–16. URL https://www.stopfake.org/en/anatomy-of-an-info-war-how-russia-s-propaganda-machine-works-and-how-t 1.2

[41] oobabooga. A gradio web UI for large language models: oobabooga/text-generation-webui. URL https://github.com/oobabooga/text-generation-webui. original-date: 2022-12-21T04:17:37Z. 3.3.1

[42] James Pamment. *The EU's Role in Fighting Disinformation: Crafting A Disinformation Framework*. Carnegie Endowment for International Peace. URL https://carnegieendowment.org/files/Pamment_-_Crafting_Disinformation_1.pdf. 1.2

[43] Cesar Augusto Rodriguez, Timothy C. Walton, and Chu Hyong. Putting the "FIL" into "DIME": growing joint understanding of the instruments of power. 97(8). URL https://apps.dtic.mil/sti/citations/tr/AD1099537. 3.1.2

[44] Qt for Python Team. PySide6: Python bindings for the qt cross-platform application and UI framework. URL https://pyside.org. 3.3.1

[45] Seth Tisue and Uri Wilensky. Netlogo: A simple environment for modeling complexity. In *International conference on complex systems*, volume 21, pages 16–21. Citeseer. URL https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=a65e2af0f4a1b03db4b05357c4cb3b8a6a4d7894. 2.2, 3.2.3

[46] Joshua Uyheng, Thomas Magelinski, Ramon Villa-Cox, Christine Sowa, and Kathleen M. Carley. Interoperable pipelines for social cyber-security: assessing twitter information operations during NATO trident juncture 2018. 26(4):465–483. ISSN 1381-298X, 1572-9346. doi: 10.1007/s10588-019-09298-1. URL http://link.springer.com/10.1007/s10588-019-09298-1. 1.2