Simulating Civil Security Activities in Stability Operations

Susan Aros, Mary McDonald
Naval Postgraduate School
Monterey, CA
skaros@nps.edu, mlmcdona@nps.edu

ABSTRACT

Civil security activities are an essential aspect of stability operations that address threats and reduce violence to establish a safe environment in which civilians can freely move about and carry out normal activities. When addressing threats to civil security, a military force must balance the use of nonlethal and lethal force in a way that best meets mission objectives. However, civilians may respond in unexpected ways to varying uses of force, making this balance difficult to achieve. Researchers at the Naval Postgraduate School's Center for Modeling Human Behavior have developed the agent-based simulation Workbench for Refining Rules of Engagement against Crowd Hostiles (WRENCH) to address this challenge. This simulation models a security force carrying out civil security activities, addressing any emerging threats via the use of both nonlethal and lethal weapons. WRENCH also models realistic complexities of civilian individuals and groups that drive both crowd behaviors and responses to the use of force. In the simulation, the user can specify many characteristics of the security force, including the set of available weapons and tactical rules of engagement to govern weapons use. Many aspects of the civilian population can also be specified by the user. In this paper, we first provide a brief overview of WRENCH's design and capabilities in the context of securing an urban area and compound and describe key output metrics of interest. We then discuss an experiment that explores the effects of different weapon types under different sample rules of engagement across several different civilian population scenarios. Finally, we present insights from both the analysis of experiment results and the challenges associated with model design and validation efforts.

ABOUT THE AUTHORS

Susan K. Aros is a Research Assistant Professor in the Naval Postgraduate School's Department of Defense Management and Director of NPS's Center for Modeling Human Behavior. She has a PhD in Information, Risk, and Operations Management, an MEng in Operations Research and Industrial Engineering, an MA in Spiritual Formation and Soul Care, and a BA in Psychology. Her published research spans the areas of human behavior modeling, peace support operations, disaster response operations, inter-organizational communication, production planning and scheduling, and RFID use in remanufacturing, and uses various methodologies including agent-based simulation, discrete-event simulation, and mixed integer programming. Her email address is skaros@nps.edu.

Mary L. McDonald is a Faculty Associate for Research in the Naval Postgraduate School's Simulation Experiments & Efficient Designs Center for Data Farming. She received her BA in Mathematics from Northwestern University and MS in Applied Mathematics from the Naval Postgraduate School. She has taught courses in probability and statistics and simulation analysis and has lectured for courses in combat modeling. She has over 25 years of experience in applying simulation modeling, design of experiments, evolutionary algorithms, and high-dimensional data analysis to Department of Defense problems. Her email address is mlmcdona@nps.edu.

Simulating Civil Security Activities in Stability Operations

Susan Aros, Mary McDonald
Naval Postgraduate School
Monterey, CA
skaros@nps.edu, mlmcdona@nps.edu

INTRODUCTION

When conducting stability operations, security is an immediate concern. In establishing civil security and addressing emerging threats, the planning and execution of tasks "must consider the joint principle of restraint and the ethical application of force." (ADP 3-07 S.2-2) An important part of this is balancing the use of nonlethal and lethal force in a way that best aligns with the threat levels and meets mission objectives. "Finding ways to accomplish a mission with an appropriate mix of lethal and nonlethal force remains an important consideration for every commander. Commanders analyze situations carefully to balance lethal and nonlethal actions." (ADP 3-07 S.3-64) To do this, commanders need to "consider the results of lethal or nonlethal actions on an area of operations and their effect on the intended end state." (ADP 3-07 S.2-4) However, civilians may respond in unexpected ways to varying uses of force, making this balance of force difficult to achieve.

One approach to better understanding the potential effects of using various levels of nonlethal and lethal force is to employ simulation modeling. A simulation approach, when used properly, produces a model of a real environment and situation that can be explored at a speed much faster than live exercises, tested under a broad array of parameters and choices not possible to fully explore in either the real world or historical case studies (Hughes, 1997). When modeling complexities of human social behavior and individual decision-making, the agent-based simulation (ABS) modeling methodology is particularly useful (Macal & North, 2005). ABS facilitates the creation of autonomous virtual people that can interact in a virtual environment over time. Each virtual person will iteratively and independently perceive aspects of their surroundings, nearby people, and any events that occur, react internally to this information, decide how to respond, and then act out the chosen behavior(s), providing a realistic representation of humans.

To address the particular challenges associated with civil security stability operations, researchers at the Naval Postgraduate School's Center for Modeling Human Behavior have developed the ABS Workbench for Refining Rules of Engagement against Crowd Hostiles (WRENCH). WRENCH simulates a security force carrying out civil security activities, addressing any emerging threats via the use of both nonlethal and lethal weapons. It also models realistic complexities of civilian individuals and groups that drive crowd behaviors and responses to the use of force, going beyond the usual aggregation of individuals or the crowd as a united group. In the simulation, the user can specify many aspects of the security force, most notably which intermediate force capabilities (IFC) are available and the tactical rules of engagement (ROE) that govern their use. Many aspects of the civilian population are also specified by the user. In this paper, we first provide a brief overview of WRENCH's design and capabilities in the context of securing an urban area and compound and describe key output metrics of interest. We then present an experiment that explores the effects of the use of different types of weapons under different sample rules of engagement across different civilian population scenarios. Finally, we discuss insights from both analysis of experiment results and the challenges associated with model design and validation efforts.

WRENCH OVERVIEW

WRENCH is an ABS model, coded in NetLogo, developed to simulate a security force (SF) conducting a civil security mission and addressing any emerging threats using both IFCs and lethal force. It models a relatively small geographic area at a tactical-level using one-second timesteps. The simulation's design draws on literature from multiple disciplines to model complex drivers of individual and group behavior. Simulated civilian individual people move about an urban environment specified by real-world geographic information system (GIS) data and interact with other civilians and security force members. If civilians are perceived by members of the SF to present a threat, the SF will

address them according to the ROE and could cause injury or death. Some civilian individuals arrive with others in a connected social identity group (SIG), while others arrive by themselves; people can also choose to join, form, or leave a group dynamically over time. These groups are informal gatherings that tend to stay together; members of a SIG stay in proximity to the SIG leader. Each SIG member influences, and is influenced by, their SIG. Individuals and SIGs are modeled as having state variables representing emotions (fear, anger), cognitive elements (belief about the legitimacy of the forces, hostile intent, memory), needs (being, belonging, becoming) and an overarching objective such as protest, attack, or stay-safe. Each state variable is updated over time as a function of that agent's experience, perceptions, and other attributes. For example, emotional responses of individuals are recalculated every timestep as a function of immediate experience and perceptions while higher-level cognitive functions, needs, and objective are recalculated on a less frequent periodic basis as a function of emotions in addition to current and recent experiences and perceptions. The attributes of each SIG change more slowly relative to the parallel attributes of its members. A person's behavior is typically driven by their objective but can also be driven by a sudden jump in fear from weapons impacts received or witnessed. Additionally, movement is adjusted to avoid collisions, consider preference for personal space, and keep the individual within proximity of their leader if in a group.

In the compound defense scenario, the SF consists of a squad that guards the compound, with one guard stationed at each gate, as well as several patrolling squads. Patrolling squads start in patrol vehicles, but once the compound goes into defense mode the patrolling squad members will dismount and move to the gates to help defend the compound. The security force can have a user-designated set of one or more types of IFCs issued to them and must operate within user-specified ROE. IFC details and immediate effects on individuals are modeled according to available specifications, literature, and a study of historical video evidence.

WRENCH can be run interactively where the effects of changes to certain variables and SF choices can be observed through animation, or can be run constructively on a high- performance computing system to perform a large number of runs in parallel to facilitate large-scale experimentation across a wide variety of combinations of input variables. For each experimental run, the simulation outputs a large amount of data including state variable information about each identity and person, as well as metrics pertaining to intruders, security force activities, social identity activity, and the population as a whole, each timestep. For a more detailed overview of WRENCH see Aros et al. (2021).

EXPERIMENT

WRENCH v5.4 has many parameters as well as security force and population input variables that can be specified by the user. In order to explore how the use of various IFCs under different ROEs could affect a civil security mission, we selected several factors to test that represent choices made by the SF, and also tested these SF choices on different population types. Here the term *factor* refers to any input or variable that will be experimented with using different values (Sanchez et al., 2020). The factors selected for this experiment are described in more detail below.

A primary choice pertaining to the SF is the ROE. Within the simulation the ROE is represented as a set of detailed tactical rules listed in order of priority. Each rule within an ROE ruleset details a range of threat levels to be addressed within a given distance range from the compound gate and the type of threat to look for: hostile individual, hostile cluster of individuals, or high-density cluster of individuals. The SF members assess threat levels according to their perception of each person's behavior. In the simulation, the user can design a custom ROE ruleset or use one of the sample notional ROEs provided; for this experiment we tested three sample ROEs: QuellAggression (QA), MixedCritRingsByThreat (MC), and RingsNearEmph (RN). With the QA ROE the SF only engages with aggressive, destructive, or violent individuals which allows low-to-moderately hostile people to remain in the area. Under the RN ROE the SF first addresses all nearby individuals of any threat level, then individuals in the moderate distance range who are at least moderately hostile, and finally anyone in the farthest range who is highly hostile. This RN ROE allows people with lower hostility levels to remain in the middle area, and lower- and moderately-hostile people to remain in the far area. Under the MC ROE, the SF focuses first on the highly hostile threats, addressing any nearby individual threats first, then any hostile cluster threats at a moderate distance, followed by any high-density cluster threats farther away. This is then repeated for moderate-hostility threats, and then low-hostility 'threats' (even if compliant), thereby encouraging all people to leave the entire area.

The weapons database developed for the simulation includes a variety of IFCs where each IFC has a specified range for use, impact zone size and shape, and affects either a single person (SP) or multiple people (MP) with a single

firing. IFCs are also categorized by severity and type of effect: three levels of pain IFCs cause physical effects varying from pain to immobilization, and three levels of psych IFCs cause varying levels of intimidation and possible discomfort but have no lasting physical effects. For this experiment, we focused on five types of IFCs: a long-range acoustical device (LRAD: psych 1 or 2, MP), a laser dazzler device (LDD: psych 3, SP), a flash-bang grenade (FBG: pain 1, MP), a blunt trauma device (BTD: pain 2, SP), and an active denial device (ADD: pain 3, SP). For the IFC factor we tested ten options, five consisting of the SF having only one type of IFC and the other five allowing the SF to use combinations of IFCs: LRADandBTD , LDDandFBG, FBGandBTD, ADDandBTD, and AllFive. The simulation also models the level of Escalation of Force (EOF) that is allowed when an SF member does not have an IFC that exactly matches the threat level. The EOF can be set to none, low, moderate, or high. Each additional step up in the EOF setting allows an additional step up in severity of weapon to be used if a more suitable weapon is not available. The final SF characteristic varied in the experiment is their stance toward the population. While this may not be a conscious choice, it is an important one nonetheless. The SF stance affects how the force members perceive the hostility level of the people. By observing a person's behavior an SF member can deduce the range of hostility levels they perceive the person to have. Under a Nurturing stance, they 'assume the best' and act based on the low end of the perceived hostility range for that person; under the Repressive stance, they 'assume the worst' and act based on the high end of the range; and under the Cautious stance, they assume a level of hostility in the middle of the range.

We hypothesized that the effects of the SF choices might depend somewhat on the population the SF was facing, so we also designed and tested three notional populations: Protest_SIG, Protest_indiv, and Market. In both Protest populations, most adults arrive with the objective to protest with a few having the objective to attack, and few children are present. In the Market population, the majority of adults arrive with no particular objective other than to wander through the area to shop, while some arrive with a protest objective, and a large minority of the population are children. In the Protest_indiv population, every adult arrives individually (not in a group), whereas both in the Protest_SIG and the Market populations, some adults arrive in family groups, some in social groups, and a small minority of adults arrive individually. The Protest_SIG population has a somewhat larger minority of adults arriving individually relative to the Market population. All other population settings were kept the same across these three populations.

Design of Experiment

When developing an experiment, a design must be chosen that will then be used to specify the combinations of inputs to test, termed design points (DPs). A full-factorial design tests every possible combination of the inputs. With highly-dimensional ABS models, the number of design points required for a full-factorial experiment that explores just a fraction of available inputs can become impossibly large, so an efficient experimental design that can retain near-orthogonality of the factors is often selected to reduce the number of design points to a manageable level (e.g. Cioppa & Lucas, 2007; Sanchez et al., 2020; Vieira et al., 2013). However, there is a tradeoff to consider here. When an efficient, non-gridded design is used and the results are then subdivided into many smaller sets during analysis, creating highly constrained subspaces, the experimental factors are more likely to be correlated with each other within each subset (Nisbet, 2017). For this experiment, we wanted the flexibility to dive into any subspace and retain orthogonality of the factors and the ability to make "apples to apples" comparisons, so we chose to use a full-factorial design made feasible by testing a relatively limited number of factors, and levels per factor, as shown in Table 1. Our experiment used a total of 1,080 design points (DPs). Since WRENCH is a stochastic model (makes use of pseudo-random numbers to account for intrinsic variability) we ran each DP for 30 replications yielding a total of 32,400 simulation runs. The simulation was run for a simulated time of 30 minutes to capture the potential for a situation to immediately escalate.

Table 1: Summary of Experiment Factors and Levels

Factor (variable)	Levels (values = options)
Tactical Rules of Engagement ruleset (ROE)	{QA, MC, RN}
Intermediate Force Capabilities available (IFCs)	{LRAD, LDD, FBG, BTD, ADD, LRADandBTD, LDDandFBG, FBGandBTD, ADDandBTD, AllFive}
Escalation of Force allowed (EOF)	{none, low, moderate, high}
Stance of SF toward population	{Nurturing, Cautious, Repressive}
Population type	{Protest_SIG, Protest_indiv, Market}

Metrics of Interest

In order to keep within the required page limits while exploring the results we carefully selected just a few of the many interesting metrics available from the simulation, as described below. For some metrics that were not monotonically increasing we captured both the escalation of the metric (the increase from the starting value to the maximum during the run) and the change in the metric (the difference between the ending and the starting values) due to the fact that using the change measure alone, while important, would mask an equally important large escalation that was then mitigated. For metrics pertaining to the SF's management of the crowd we chose to focus on: the total number of intruders at the end of the run, as a measure of success in securing the compound; the escalation and change in number of attackers, as measures of immediate and overall escalation or de-escalation of the situation; and the change in total number of people in groups, as a measure of emergent crowd organization. For metrics pertaining to the effects of the SF activities on the population we focused on: the total number of times psych-level IFCs were fired, as a measure of the level of psychological coercive force deployed using IFCs; the total number of times pain-level IFCs were fired, as a measure of the level of painful non-lethal coercive force deployed using IFCs; the total level of injuries sustained by the people from pain weapon hits (averaged over people), as a measure of the risk to the civilian population (pain 1 IFCs only cause sustained injuries to children); the change in the mean hostility level of people (averaged over people), as measure of immediate and overall escalation and for implications for future interactions; and, finally, the change in the mean legitimacy belief of the people (averaged over people) about the security force, for implications for future interactions.

EXPERIMENT RESULTS AND ANALYSIS

To "consider the results of lethal or nonlethal actions on an area of operations and their effect on the intended end state" (ADP 3-07 S.2-4) we conducted several types of analyses. We first explored relationships between metrics by calculating the pairwise correlations between several metrics of interest. We then used partition trees and bootstrap forests to gain insights into what combinations of SF choices may lead to a more desirable outcome on several of the metrics. The insights from this analysis then inspired further analyses. For all analysis, the metrics were averaged over the replications within each design point, yielding one data point for each of the 1,080 design points.

Correlations Between Metrics

As a first step in the exploration of the results we calculated the pairwise Pearson's correlation coefficients among several metrics of interest including total psych fired, total pain fired, change in mean injury level, escalation of attackers, change in total intruders, and total people in SIGs. A Pearson's correlation coefficient is a measure of the strength of a linear relationship between two variables which falls into the range [-1, 1], where 0 indicates no relationship, positive values indicate that the two variables rise and fall together, and negative values indicate that as one variable rises, the other falls (Devore, 2014). Figure 1 depicts a color map of the Pearson's correlation coefficients. The correlations revealed several interesting findings. We found a strong positive correlation (0.7388) between use of psych IFCs and the escalation in number of attackers, which seems counter-intuitive; one would think that a greater

use of pain-level IFCs would be more correlated with an increase in attackers than a greater use of psych-level IFCs. In addition, there was a positive correlation (0.3898) between use of psych IFCs and total intruders, which corresponds with the prior result because attackers are the ones who attempt to enter the compound. The results also show a strong positive correlation (0.6505) between the total intruders and the change in number of individuals in groups. In other words, the level of spontaneous organization within the crowd was highly correlated with the number of intruders. Another interesting result is a small correlation where one would expect a larger one: between the total amount of pain IFCs fired and the number of intruders (-0.0741). It would seem reasonable that the use of pain IFCs would significantly reduce the number of intruders.
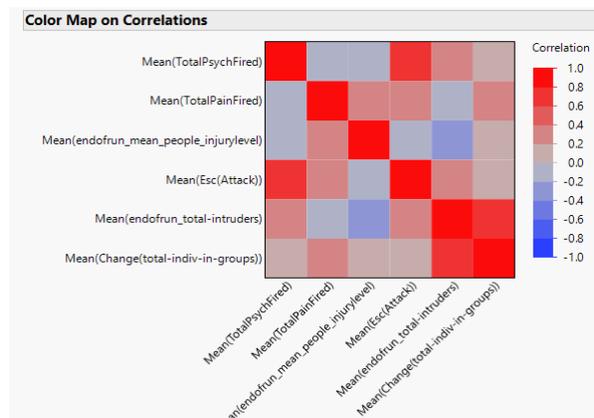

Figure 1. Correlations (all Populations).

To further investigate the unexpectedly small correlation between the total amount of pain IFCs fired and the number of intruders we looked at only the results for the two Protest populations, since other analysis showed that many Market population runs resulted in no intruders and no use of force. For this dat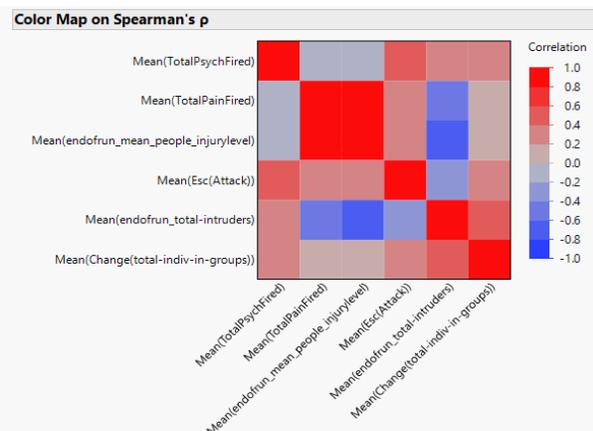a from the protest populations the Spearman's rank correlation coefficients, which measure the strength of a potentially nonlinear but monotonic relationship between two variables, was used. This nonparametric method revealed stronger relationships than the corresponding parametric version for this data. Figure 2 displays the Spearman's correlations on the Protest populations subset, and we now observe a much higher negative correlation (-.5789) between total amount of pain IFCs fired and the number of intruders. A takeaway is that a pairwise correlation analysis does not consider the influence of additional factors and does not reflect complex relationships. This type of analysis should be complemented by other methods to explore relationships between more than two variables. In addition, one should consider if a relationship might only hold true for a specific subspace, consistent with a reasonable hypothesis.

Figure 2. Correlations (Protest populations).

Effects of SF Choices on Metrics of Interest

To directly address the research question, "What combinations of choices available to the forces may lead to better outcomes?", we generated a partition tree for each metric allowing only factors representing SF choices to enter the tree. A partition tree, also known as a classification or regression tree, is a nonparametric approach that recursively splits input space in a manner that best explains the response, or metric (Hastie et al., 2008). We generated partition trees for three metrics pertaining to the SF's management of the crowd, with what is considered the 'best' outcome indicated in parenthesis: total intruders (least), change in attackers (most reduction), and change in people in groups (most reduction). On the best branches of all three of these trees the QA ROE and a Repressive stance appeared, and on two of these trees the RN ROE, Cautious stance, and moderate or high EOF all appeared on the best branch. Interestingly, an IFC split (separating subgroups of IFC options) only appeared on the best branch of one tree, and one of the six IFC options on the best branch also appeared on that tree's worst branch! We also generated partition trees for three of the metrics pertaining to the effects of the SF actions on the people, with 'best' outcome indicated in parenthesis: accumulated injuries (lowest), change in hostility (most reduction), and change in legitimacy beliefs (most increase). The QA ROE and a Nurturing stance each appeared on two of these trees, along with certain IFC options; but, again, for one of the metrics some of the IFC options on the best branch also appeared on that same tree's worst branch! The RN and MC ROEs each appeared on one of these metric's best branch.

Examining these partition tree results, we see that the only SF option among all four SF choices that seems to perform well across most of these six metrics is the ROE choice of the QA option, contributing to fewer intruders, most reduction in attackers, most reduction in crown organization, lowest civilian injuries, and most increase legitimacy beliefs about the SF. A likely main driver is that the QA ROE has SF members only engage with people who are more hostile, whereas the RN and MC ROEs have the SF also engaging with less hostile people potentially causing unintended escalation. With regard to stance there appears to be a trade-off; the Repressive and Cautious stances contribute to better crowd control, but the Nurturing stance contributes to better outcomes for the people. However, there aren't any clear insights from this analysis regarding IFC options. Only three out of six trees had an IFC split on the best branch, appearing as the last split on the branch for all three, and on two out of those three trees some of the best-branch IFCs also appeared on the worst branch.

A partition tree is constructed using a 'greedy' algorithm to identify the single best tree, so it is possible that some factors with true explanatory power were not allowed to enter that single tree. A bootstrap forest algorithm, on the other hand, fits many hundreds or thousands of individual partition trees, systematically limiting the variable options and number of data points considered for each tree, providing a better overall picture of the relative impact of SF choices on a given metric (Berk, 2006). Developing a bootstrap forest for each of these six metrics revealed that, for four of the metrics, the IFC choice provided the highest column contributions, meaning it had the highest explanatory

power, and was second highest for another. Thus, despite the muddled results from the partition trees regarding which IFC options led to the best results, and its lack of appearance on some trees, the bootstrap forests make it clear that the choice of IFCs does matter. Taking the partition tree and bootstrap forest results together, it appears that the best IFC option may depend on multiple interactions and differ across metrics.

Factor Influences and Interactions, Broken out by Population and Force Stance

In the first two analyses the results for all population types were mostly analyzed together. In this section, we explore the effects of different SF choices on results for each population type separately, focusing in on three of the metrics (total number of intruders, accumulated injuries to civilians, and the change in the number of people in groups). Also, since the previous analysis indicated that IFC stance was associated with a trade-off across metrics, we also split the results data by force stance yielding nine data population-stance subsets. To explore the nuances of relative IFC effectiveness, we created bar charts with each bar representing the average outcome for the indicated metric under one of the IFC options, and overlayed the 95% confidence interval. Here we discuss the key takeaways for these three metrics seen on numerous charts and include a few in figures as examples. We note that the analysis of bar charts can be followed up with a deeper exploration of shape and spread of the distribution of outcomes and/or use of a statistical test particularly where differences are small, but for our purpose here we are mainly interested in identifying overall trends and discovering where larger differences appear.

Compound Intruders
Figure 3 shows the set of bar charts that illustrate effects of IFC, force stance, and population type on the total number of intruders. The left set of bar charts shows bars for each single-type-of-IFC option, for each population-stance subset, and the right set similarly shows the bars for each multiple-type-of-IFCs option; each chart set shows population types across columns (left to right: Market, Protest_SIG, Protest_indiv) and SF stances across rows (top to bottom: Nurturing, Cautious, Repressive). Looking at the overall differences across population type reveals that a Market population produced the least number of intruders, the Protest_SIG population many intruders, and the Protest_indiv the most. For both Protest populations, a Nurturing stance allowed the most intruders, the Cautious stance a moderate number of intruders, and the Repressive stance marginally fewer than the Cautious stance, on average. The differences across stances were minimal for the Market population.
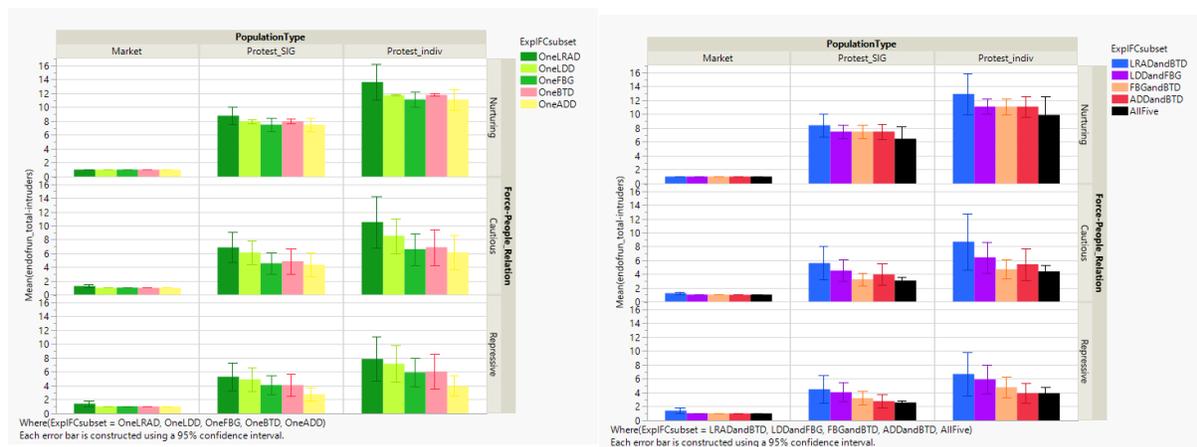

Figure 3. Effect of IFC Choice on Intruders

Focusing in on each IFC option bar for the Market population shows that the choice of IFCs also made minimal difference for this population, with a very small increase in intruders when the IFC option was LRAD or LRADandBTD, but only under a Cautious or Repressive stance. For both Protest populations, the IFC options that included multiple types of IFCs generally resulted in somewhat fewer intruders than the single-IFC-type options, and the higher severity level IFC options caused somewhat fewer intruders than the lower severity options (but under a Nurturing stance this difference was only marginal). The observations about how many types of IFCs were available and their severity can be seen more clearly in Figure 4. In the left set of charts in Figure 4 each bar shows the results of a different number of types of IFCs available to the SF (one, two, and five types, respectively) showing that providing the SF with more types of IFCs contributed to fewer intruders. The right set of charts shows results by which

severity category of IFCs were available to the SF (psych only, pain only, or both, respectively) showing that not providing the SF with any pain weapons contributed to more intruders.
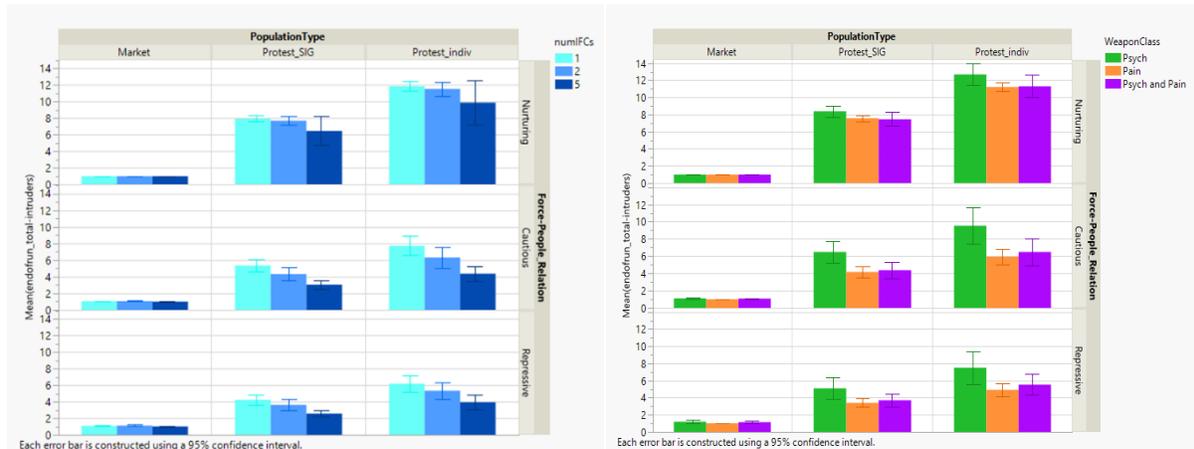

Figure 4. Effects of Number of IFCs Available and IFC Severity on Intruders

Injuries to Civilians

The bar charts for the injuries to civilians resulting from use of different IFCs (analogous to Figure 3) show that, for the Market population, there were no accumulated injuries with a Nurturing SF stance, low injuries with a Cautious stance, and (for most IFC options) more injuries with a Repressive stance. For the Protest populations the level of injuries under the Nurturing stance were low, but with a Cautious stance they were much higher (with some IFCs they were even tripled relative to injuries under the Nurturing stance); under a Repressive stance the injuries were somewhat higher for most IFC options relative to injuries under a Cautious stance, but the FBG and LDDandFBG options actually produced fewer injuries under a Repressive stance than under a Cautious stance. Looking into more detailed differences between IFCs also reveals interesting results: for the pain-level single-IFC-type options, the more severe SP IFC option (ADD) did result in somewhat higher injury levels than the less severe SP IFC option (BTD) across populations and stances (omitting the no-injuries data points from nurturing a Market population); however, the relative performance of the MP pain IFC option (FBG) varied. For both Protest populations under a Nurturing stance, and for the Market population under a Cautious stance, the MP FBG option produced more injuries than the SP BTD option, and nearly as many as the SP ADD option. The bar charts also show that having more IFC types available can increase overall injuries with one notable exception: adding the LDD IFC type when the FBG was available did not noticeably affect the overall level of injury. These charts illustrate the benefits of examining the data using different types of summary measures.

Crowd Organization

The metric that captured the level of spontaneous organization within a crowd is the change in the number of people in groups. With some populations arriving in groups, and another population not, it is particularly important to break out population types for this metric. The Protest_indiv population, where no one arrived in a group, showed by far the greatest amount of spontaneous crowd organization, as expected, but the amount of organization was highest under the Nurturing stance, somewhat less under the Cautious stance, and even less under the Repressive stance. The Protest_SIG population saw a low amount of crowd organization under the Nurturing stance, marginal under the Cautious stance, and effectively none under the Repressive stance. The Market population showed effectively no change in crowd organization under the Nurturing stance, a small decrease under the Cautious stance, and a moderate decrease under the Repressive stance. It is important to think about the implications of these changes; while an increase in crowd organization of a potentially hostile crowd is clearly not desirable, a negative change in crowd organization may also not be desirable, particularly in a Market population, since many of the initial groups are family groups. When adults in a family group part ways, that frees up one or more adults from the constraint of keeping track of children, perhaps allowing them to choose more freely to aggressively protest or attack in response to a repressive force. This may be a contributing factor to a marked increase in attackers seen in the Market population under a Repressive SF stance.

Tradeoffs and Balancing Multiple Objectives

In analyses up to this point we have noted that there are tradeoffs, primarily between metrics that indicate effective crowd control and metrics that highlight effects on the population, and that results can differ across populations. For example, Figure 5 shows the clear tradeoff between a key crowd control metric (total intruders, in blue) and a top population effects metric (injury levels, in red). As the SF is given the leeway to use greater escalation of force the number of intruders decreases while the injury levels to the people increases. In light of this the question then arises, "How can we best balance tradeoffs?"
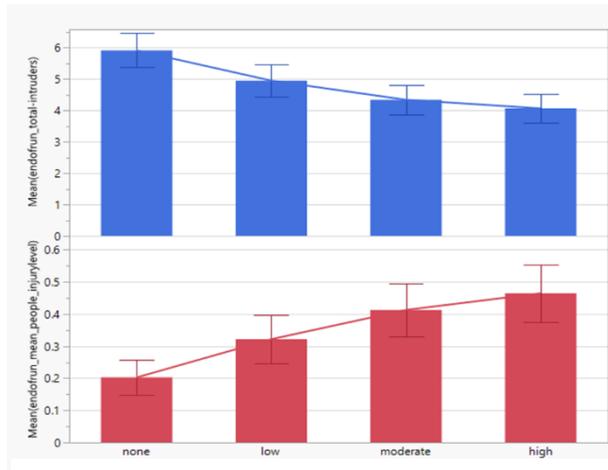

Figure 5. Tradeoff between Intruders and Injuries across EOF choice.

One method for exploring which choices can effectively balance tradeoffs is to examine which design points performed well on both metrics simultaneously. Figure 6 depicts a scatter plot of design points across the two key metrics of intruders (x-axis) and injuries (y-axis) (Note that the IFC options that only include psych weapons will not result in any injuries from IFCs, so the design points with OneLRAD or OneLDD are not included here.) To explore what contributes to better performance on both metrics simultaneously, and thus a good balance across both metrics, we first specify what will be used as a cut-off to distinguish between 'better' and 'not as good' performance on each metric. For this analysis we decided that the cut-off for injury level would be set to 1.0, and the cut-off for number of intruders would be a somewhat manageable 3 intruders; therefore the data points below 3 intruders (x-axis) and below 1.0 average injury level (y-axis) are interpreted as providing a good balance across
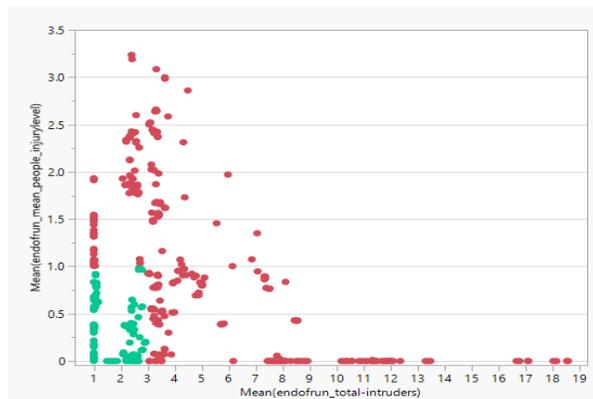
these two metrics. We then define a new binary metric BalancesObjectives that is true for the data points that do provide that better performance on both metrics, and false for those that do not. We can then construct a partition tree on this new BalancesObjectives metric. Figure 7 depicts this partition tree in a compact format that is read from left to right, where the proportion of data that does provide a good trade-off is shown in green and the proportion that does not is in red. As can be seen in Figure 7, the first split occurs on population type between the Market population and the Protest populations, and then for the Protest populations branch the next split is again on population type. This fortuitous result makes it easy to read from the decision tree which IFC options contribute to a good balance between intruders and injuries for each of the populations separately.

Figure 6. Scatterplot of design points (Intruders vs. Injuries)

For the Market population, we see that using either the QA or MC ROE contributes to a good balance regardless of stance or IFCs. If the RN ROE is in use by the SF the stance becomes important; here the Nurturing stance contributes to a good balance across these two metrics. If the RN ROE is in use and a Cautious or Repressive stance is taken toward the Market population, it becomes important to choose the IFC with care; here it appears that the FBG is helpful as long as it is not paired with the BTD.

For a Protest population (consisting almost entirely of protesters with a few attackers) where the majority of people arrive in small groups, the Cautious or Repressive stance has a better chance of leading to a good balance across these two metrics. Under either of these stances, using the QA ROE and choosing an IFC option that includes BTD is better relative to the other ROE and IFC options. If the Protest population doesn't arrive in small groups, everyone is on their own, a good, balanced outcome is harder to achieve. With this population it seems that it's important for the SF

to take a Repressive stance (i.e., assume the worst of each person based on their behavior), use the RN ROE, and choose an IFC option that includes an FBG but not a BTD.
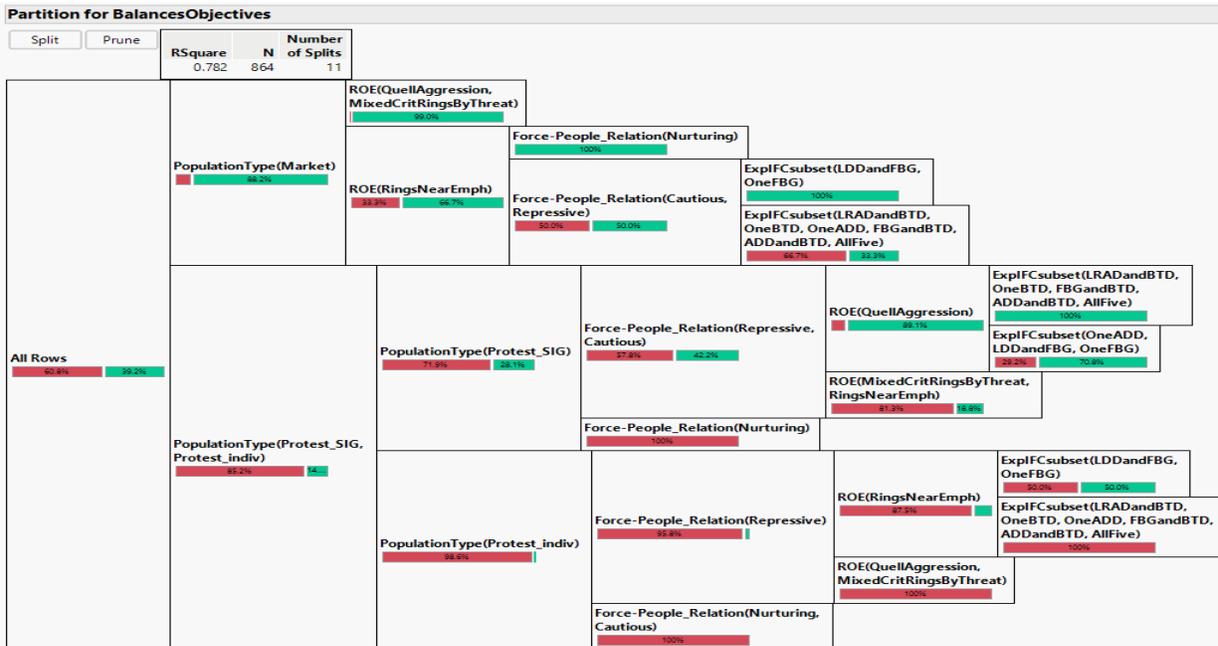

Figure 7. Partition tree on the ability to successfully balance injuries and intruders.

DISCUSSION

Experimentation with the WRENCH simulation model, and the different types of analyses used, provided interesting results. Regarding the analytical methods, the correlations provided a good initial look at the results and yielded some insights, but only considered pairwise linear or monotonically increasing relationships so the insights were limited, prompting further analysis. The partition trees and bootstrap forests provided some initial insights into what SF choices might lead to better results but showed a significant tradeoff with regard to the SF stance and lack of clarity on which IFCs might be best. The bar chart analysis made it evident that the outcomes of different SF choices depended significantly on the population type, and began to highlight some of the subtle distinctions in the effects between different IFC choices. A further analysis that aimed to address inherent tradeoffs on two key metrics (compound intruders and population injuries) provided insights into which SF choices led to outcomes that balanced the two. This experiment and analysis also shows that leveraging data farming analysis methods with complex simulation models can provide extensive insights into interactions between factors and facilitate consideration of multiple outcome metrics simultaneously.

Overall the results and analysis indicated that the type of population was the top factor to take into account when considering which SF options might lead to better outcomes across competing metrics. The often-unconscious choice of what stance to take toward that population also emerged as a critical factor in several of the analyses. For all population types, the choice of ROE led to clearer delineations of results than the choice of which IFCs the SF had available to use or the amount of escalation of force allowed. Another metric, the change in the number of people in social groups within the crowd, was not considered in the final balancing of the trade-offs analysis but is worthy of further exploration since our correlation analysis showed that this metric is strongly positively correlated with the number of intruders into the compound. To our knowledge, group dynamics, and particularly spontaneous crowd organization, is an area that has not been explored in crowd simulation modeling previously but has been shown to be vitally important by events such as the January 6, 2021 attack on the US Capitol. The first three analyses also showed that interactions between SF choices are prevalent and can significantly affect results. The trade-off analysis also highlights the importance of considering multiple outcome metrics simultaneously to avoid unnecessarily bad outcomes on one or more of these metrics (as shown in Figures 6 and 7).

When considering the generalizability of the insights provided here to real-world situations it is important to consider the limitations of this study. Some of these limitations include:

- The sample ROEs tested in this study were designed to show a range of possible ROEs but were not based on tactical ROEs currently in use by a specific military or security force.
- Civilians do not use any types of weapons against the SF in the current version of WRENCH, and violence on the part of civilians that results in injuries to SF members is not yet modeled.
- More work is needed to accurately model the IFCs, including more refinement on the detailed physical effects of each type of IFC hit on an individual.
- There is always room for more enhancements and refinements with a model such as WRENCH. We do strive to avoid unnecessary complexity and therefore add detail only where deemed relevant and necessary.

Any discussion of simulation results is not complete without a discussion of model verification and validation (V&V). Complex ABS models, particularly those that model rare events or human behavior, are difficult to validate because the dearth of existing input data for model construction and output data for comparison makes it impossible to employ many of the traditional V&V methods for simulation (Aros & Gibbons, 2018). There are a number of V&V methods that are still feasible though, most notably careful conceptual model development, mining source literature for model components and any available data, discussions with SMEs, code testing, and face validation; we employ all of these. In addition, data farming for V&V has proven particularly useful to catch bugs, reveal conceptual model errors, and isolate anomalous findings (Aros & McDonald, 2023). It is also important to keep in mind that, due to the difficulties in validating these types of simulation models, they should not be used to forecast outcomes for specific missions. Rather, well-designed ABS models offer insights into the range of potential outcomes and the relative importance of different factors on different metrics of interest.

NEXT STEPS

Our next steps include developing additional scenarios such as border security and humanitarian aid distribution, and incorporating several other model improvements that will focus on adding pertinent details to the representation of SF members and activities and of the civilian population and individuals. For the SF, we will focus on behaviors for forces on patrol engaging with civilians, additional nuances and restrictions regarding the use of lethal force and capturing the influence of individual identity on SF member choices. For the population, an ability to create distinct subpopulations within a single run will be implemented allowing the modeling of a primary population (e.g. Market) with one or more different subpopulation (e.g. small groups of young protesters). We will also be coding the use of weapons by civilians and explicit acts of violence resulting in injury to security force members. We are also evaluating an approach to capture the effects of social networking and social media use, thereby incorporating influences from outside the people immediately present and the potential for newly-arriving people in the middle of a run that were drawn there by social networks and social media. We are continuing V&V efforts and conducting structured discussions with force members who have trained with or used IFCs in order to support conceptual model refinement. Finally, we intend to explore the possibility of developing a civil-security wargame where the design and adjudication would be informed by WRENCH, and information gathered during game play or post-game interviews could inform further refinement of the model.

ACKNOWLEDGEMENTS

REFERENCES

Aros, S. K., Baylouny, A. M., Gibbons, D. E., & McDonald, M. L. (2021). Toward Better Management of Potentially Hostile Crowds. *Proceedings of the 2021 Winter Simulation Conference*.

Aros, S. K., & Gibbons, D. E. (2018). Developing an agent-based simulation model of the use of different communication technologies in inter-organizational disaster response coordination. *Proceedings of the 2018 Winter Simulation Conference.*

Aros, S. K., & McDonald, M. L. (2023). Squashing Bugs and Improving Design: Using Data Farming to Support Verification and Validation of Military Agent-Based Simulations. Forthcoming in *Proceedings of the 2023 Winter Simulation Conference.*

Berk, R. A. (2006). An Introduction to Ensemble Methods for Data Analysis. *Sociological Methods & Research*, 34(3), 263–295.

Cioppa, T. M., & Lucas, T.W. (2007). Efficient Nearly Orthogonal and Space-Filling Latin Hypercubes. *Technometrics*, 49(1), 45–55.

Department of the Army. (2019). Stability. *Army Doctrine Publication No. 3-07*.

Devore, J. (2014). *Probability and Statistics for Engineering and the Sciences*. 9th ed. Cengage.

Hastie, T., Tibshirani, R., & J. Friedman. (2008). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction.* 2nd ed. Springer.

Hughes, W. P. (1997). *Military Modeling for Decision Making*. 3rd ed. Military Operations Research Society.

Macal, C. M., & North, M. J. (2005). Tutorial on agent-based modeling and simulation. *Proceedings of the 2005 Winter Simulation Conference*.

Nisbet, R., G. Miner, & K. Yale. (2017). *Handbook of Statistical Analysis and Data Mining Applications*. 2nd. ed. Academic Press, Inc.

Sanchez, S. M., P. J. Sanchez, & H. Wan. (2020). Work Smarter, Not Harder: A Tutorial on Designing and Conducting Simulation Experiments. *Proceedings of the 2020 Winter Simulation Conference*.

Vieira, H., S. M. Sanchez, K. H. K. Kienitz, and M. C. N. Belderrain. (2013). Efficient, Nearly Orthogonal-and-Balanced, Mixed Designs: An Effective Way to Conduct Trade-off Analyses via Simulation. *Journal of Simulation* 7(4):264–275.