

Quantifying a State's Reputation in the Strategic Competition and Crisis Wargame for the Center for Army Analysis

Luke Braun, Eric Celaya, Nestor Grubb, John Tichenor, Matthew McClary, and Rich Morales

United States Military Academy
West Point, NY 10996

Corresponding Author: luke.braun@westpoint.edu

Author Note: All contributors of this report are members of the United States Military Academy. Cadet Braun is a third-year cadet in the Department of Military Instruction (DMI). Cadet Celaya is a fourth-year cadet in the Department of Mathematical Sciences. Cadets Grubb and Tichenor are fourth-year cadets in the Department of Systems Engineering (DSE). Cadet Chaney Lieberman, third year, also supported this capstone project with research from an independent study. Colonel Ricardo Morales, DSE Department Head, and Major Matthew McClary, DMI instructor, are the capstone group advisors. The client for this project is the Center for Army Analysis.

Abstract: The Center for Army Analysis (CAA) developed the Strategic Competition and Crisis wargame to capture how modern military competition operates in today's world. The CAA tasked our interdisciplinary team with developing a more robust, well-rounded reputation model for quantifying a state's reputation based on another state's perspective using each of our respective specialties within Systems Engineering, Defense and Strategic Studies, and Operations Research. The team leveraged tools within the Systems Decision Process to quantify the theoretical, intangible concept of reputation. Research began with a qualitative value model based on expert stakeholder analysis and a literature review. The team then identified value measures to build a swing weight matrix that produced a reputation score for each state from another state's perspective. That score, alongside an enhanced Game User Interface, can now be integrated into the existing SC2 wargame to provide a more complete, narrative experience that charts player decisions throughout the game.

Keywords: Wargame, Competition, Reputation, Interdisciplinary

1. Introduction

This capstone project aims to develop a reputation model that enhances the existing architecture of the Strategic Competition and Crisis (SC2) wargame to illustrate the impact that military, economic, or political decisions have on perceived international power by leveraging a systems approach to value modeling. Our client, the Center for Army Analysis (CAA), has developed their SC2 wargame "to capture the discussion associated with 'Competition', and how it supports the Army's Global Strategic Framework" (Engelmann & Kearney, 2021). The objective of the game is to take actions that improve one's level of global reputation "based on general perceptions of its strength, reliability, and resolve" (Headquarters, Department of the Army, 2021). Currently, CAA has a model for reputation that factors in diplomatic, military, and economic power. Our capstone group conducted research to develop a new model to include more factors that affect a state's reputation and then be able to successfully apply the model to the SC2 wargame, specifically the four great powers: The United States, United Kingdom, China, and Russia.

2. Background

Robert Keohane, a prominent academic in International Relations, writes that a state's reputation is based on "the effects of their present actions on others' future behaviors" (1999). Alex Weisiger and Keren Yarhi-Milo, political scientists and professors at Columbia University, coined the term "Reputation for Resolve" and defined it as "others' perception of that state's willingness to risk war" (2013).

The definition of reputation appears relatively consistent through the years, but variations do exist. Due to these variants, the capstone team decided to standardize the definition of reputation based on certain aspects that remained constant throughout the research. The definition of reputation consistently includes how a state perceives another and often incorporates past actions or behaviors to predict future events. In this case, we define reputation as a dynamic concept that captures how a state perceives another based the characteristics of credibility, capability, and stability.

Stakeholder analysis was conducted periodically throughout the entire process of developing our qualitative model and defining reputation, especially in the beginning. We conducted interviews with Sebastian Bae, an Adjunct Assistant Professor of Wargaming at Georgetown University, Colonel Matthew Dabkowski, an Academy Professor in the Department of Systems Engineering at West Point, and various Officers in Charge of the Wargaming Club at West Point to ensure we were well prepared. We also conducted multiple In-Progress-Reviews with the Strategic Wargaming team at CAA to ensure we remained on track with our primary stakeholders.

3. Methodology

3.1 Qualitative Value Model

Developing and refining the qualitative value model was a critical component of this research and set the foundation for creating a new reputation model in SC2. The qualitative value model reflects the research and stakeholder analysis by establishing a mutually exclusive and collectively exhaustive value hierarchy of measures to encompass a state’s reputation (Parnell, et. al., 2011). The team developed a model based on a state’s Credibility, Capability, and Stability attributes to expand on the rather myopic method of assessing a state’s reputation solely using the DIME model. The DIME model is a popular method of categorizing a state’s instruments of national power: Diplomacy, Information, Military, and Economy, but it fails to address abstract soft power elements that present themselves in the credibility and stability components of the improved model.

Many Defense and Strategic Studies and International Relations experts agree that the DIME model provides a reliable baseline for measuring a state’s reputation, but a complete model should include much more (Worley, 2012). Thomas Schelling wrote that “face is one of the few things worth fighting for,” with “face” being how other countries believe “the country can be expected to behave” (Schelling, 1966). “Credibility” captured this idea of “face” by developing value measures that account for a state’s deterrence and resolve. “Capability” captured a state’s projection of national instruments of power by expanding and refining the already present DIME model. “Stability” utilized a variety of cultural, societal, and structural measures to capture how well a state performs internally. **Figure 1** (below) provides a visual representation of the upper portion of the qualitative value model with the three categories: credibility, capability, and stability. Each value measure was originally placed in one of the categories seen below to conceptualize a state’s reputation more easily before beginning the process of raw data collection.

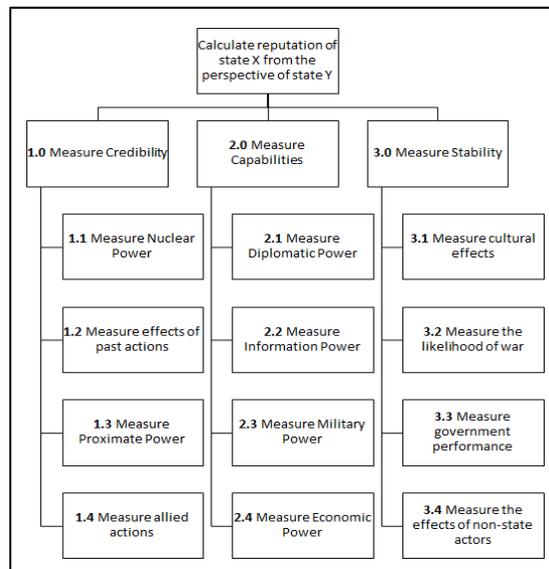


Figure 1. Qualitative Value Model

The value measures were developed by prioritizing what factors could accurately represent reputation, and what factors could be realistically measured with available and credible data. There are endless factors that have an effect on a state’s reputation, however many of them are too intangible to measure. For example, measuring a state’s successful cyber operations would be impossible, because there is not enough open-source information on the topic. While it would benefit the model, measuring this value measure or similar ones is not feasible. The value measures were also chosen to avoid issues with multicollinearity. If the model used value measures that were too similar, it would be saturated with too many similar results, thus skewing the data. For example, one of the value measures used is the “Number of Tanks in Target Country’s Region”.

According to a RAND study conducted in 2020, forward-deployed ground forces provided much more evidence of a deterrent impact than air and naval forces (Frederick et al., 2020). If the model included numerous other metrics about ground forces like field artillery, the model would be inflated by redundant data, therefore leading to skewed results.

3.1.1 Credibility

The Credibility category of our model aims to capture a country's "face" and how that country is expected to act (Schelling, 1966). Four functions capture the credibility of a given country: measuring nuclear power, measuring proximate power, measuring allied interactions, and categorizing the past actions of that state. Measuring nuclear power has been one of the strongest indicators of a state's credibility since nuclear weapons were first used at the end of World War II. The idea of nuclear deterrence developed as the first use of nuclear weapons displayed their destructive capability. Nuclear power is measured by the number of nuclear weapons a state possesses capable of targeting the country in question and number of nuclear armed submarines. The next function of credibility captures how a state's ability to project power declines with distance by measuring that distance between two states. Although there is much academic debate surrounding the idea of proximate power, Stephen Walt, renowned political scientist, makes a convincing case for its existence and significant impact on a state's ability to influence others (Walt, 1985). The third function for this component is categorizing past state actions. Past actions can have either positive or negative implications depending on the action itself and the states involved. These past actions can impact a state's credibility depending on whether the actions had a positive or negative impact on the country in question. These past actions are accounted for by measuring the amount of leadership changes since 1946 and the number of military interventions since 1946. The last function of credibility is measuring allied actions. A state's credibility can be impacted by how well a state supports its allies. This is measured by the number of allied campaigns supported since 1946.

3.1.2 Capability

The Capability category encompasses aspects of the DIME model that were previously used to score reputations in SC2 and aims to capture a state's physical capabilities. Four functions were developed to measure the capability of a state: Measuring diplomatic, information, military, and economic power. Diplomatic power is measured by the number of embassies and consulates from a state in the state in question and the combined GDP of a state's allies. Information power is measured by the number of state-controlled communication satellites in orbit. Military power is measured by total defense spending per capita, number service members, fixed wing assets, blue water vessels, and number of countries with persistent military presence. Finally, economic power is measured by the quantity of available rare-earth elements proportional to the global supply.

3.1.3 Stability

The Stability category aims to capture the various cultural, societal, and structural interactions between two states. We broke this category down into four functions: Measuring cultural effects, likelihood of war, government performance, and effects of non-state actors. Cultural affects were measured using the number of shared religions between the two states, the percentage of the target country's population that speak the country in questions primary language, and the annual number of immigrants taken in from fragile or failed states by the target country. Likelihood of war is measured using the democratic peace theory and the concepts that democracies are less likely to go to war with one another (Russet, 1993). Government performance is measured using several metrics pulled from the CIA World Factbook to include the ability for citizens in a country to express themselves freely, government effectiveness, confidence in government, and the perception of the extent that public power is exercised for private gain (CIA.gov). The effects of non-state actors are measured by the presence of multinational corporations, number of successful terrorist attacks, number of domestic counterinsurgency forces per capita, and the number of years of internal armed conflict since 1946 all within the country in question.

3.2 Quantifying Reputation

3.2.1 The Survey

The survey was derived from the value measures within our qualitative value model. Each question asked respondents to rank each value measure from 1 to 5 in terms of how influential they believe the measure to be on a state's reputation. The definition of reputation being used for this project was placed at the beginning of the survey to set the frame for respondents' thinking and increase the accuracy of their responses. Upon approval from the Institutional Review Board, the capstone team shared the survey with national security professionals within the Department of Military Instruction's Defense and Strategic Studies program and the Modern War Institute at West Point. In addition, members of the client's organization (CAA) participating in the survey as well. The results from this survey were then used to generate the relative swing weights of each value measure in the quantitative value model.

3.2.2 The Quantitative Value Model

Our chosen method for creating our quantitative value model was a weighted scoring matrix. Based on the results of the survey, we assigned a swing weight to each value measure ranging from 0-100. Value measures that were rated to be of the same importance received the same swing weight. To calculate the global weights, we summed together each value measure's swing weight. We then divided each swing weight by this total to get global weights for every value measure. These calculations were done three times – one for each category of the value model.

The next major step in our model generation was the collection of raw data for each state in our model. The team scraped raw data on each of the generated value measures from reputable internet sources such as the CIA World Factbook and the World Bank. This raw data was necessary in calculating base reputation scores for each country. When playing the wargame, facilitators collect game data based on the actions players make in the game. This game data will now replace/supplement the raw data after each turn, generating new reputation scores in real time.

The final step was calculating the total reputation score. The team created value functions for each value measure to make the raw data and game data compatible with one another. These value functions assigned each value measure a score from 1-5 based on the raw data. The value functions prevent the reputation scores from being skewed by large numbers from data points such as GDP. After the data is converted, each data point is multiplied by its appropriate global weight. We then summed these scores together to create a reputation score for each country for all three categories. It is also important to note that the model calculates reputation scores of three of the great powers from the perspective of the fourth. To account for each perspective, there are four versions of the model – one for each great power. The models are identical other than this key distinction.

4. Results and Products

4.1 Reputation Results

After building out the value model, we were able to successfully calculate baseline reputation scores for each of the four great powers from each of their perspectives. These reputation scores are intended to create a base score for each country at the beginning of each play session of SC2. This score is possible because we converted our raw data into a normalized scale using value functions as mentioned in the previous section; therefore, the reputation score is unitless. The calculated reputations scores are illustrated in **Figure 2** (below).

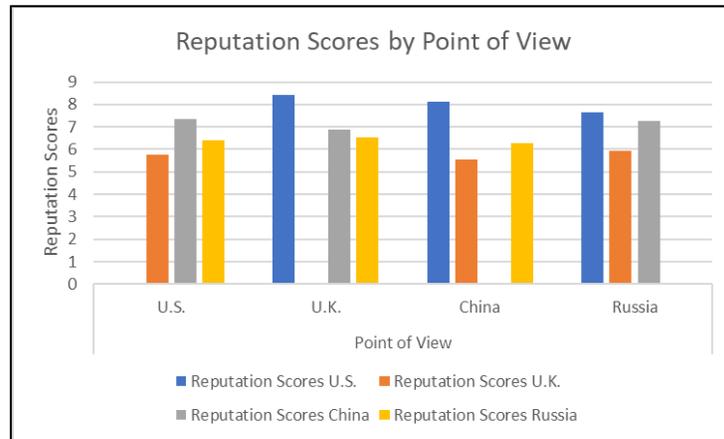


Figure 2. Comparison of Baseline Reputation Scores for the Four Great Powers in SC2

From our results, we can see how each state likely views each other in terms of reputation. Since we have four different perspectives, we can see how reputation changes based on point of view. What is interesting is that the U.K. seems to have the lowest reputation score across all points of view. Even from the U.S.'s perspective, the U.K. is rated much lower than Russia and China. However, from the U.K.'s perspective, the U.S. is the overwhelming leader. This difference of perspectives and reputation is one of the key aspects of international relations we are trying to capture.

To verify that we built our model correctly, we consulted our operations research expert on the team, CDT Eric Celaya, to check the calculations of the model to ensure they are correct. The model uses several simple mathematic formulas and excel

functions. The final reputation values are the sum of many layers of excel referencing, therefore, it was imperative that we tested the model using subject matter expertise.

For validation, we compared our model to the Lowy Institute Asia Power Index. The Power Index is a mathematical representation of the global reputation of Asian states that compiles data based on 131 indicators within eight measures to assign Asian nations a power score, comparable to the reputation score our model outputs for wargame purposes. We compared our model with the Power Index and found that our model produced similar results. Therefore, we assess that our model is valid, since the Lowy Institute Asia Power Index is an extensive and credible project (Lemahieu and Leng, 2021).

4.2 Sensitivity Analysis

To verify our model, we conducted sensitivity analysis on the value measures. This process involved varying the global weights of every value measure on a range from high to low based on the respective swing weights. We used China's reputation model for our sensitivity analysis calculations. After we concluded our analysis, we found that value measure 2.10 – Number of tanks in target countries region – was sensitive to change. From China's perspective, Russia's reputation surpassed the U.K.'s reputation as the global weight of this value measure increased. This insight is important to realize which of our value measures are the most sensitive to change. Sensitive value measures may have a larger effect on reputation as the turns play out in SC2. Identifying these possibilities is important to assessing our model's accuracy.

5. Future Research

The current capstone team has developed three main lines of effort for future research and work that will help guide future capstone groups in further developing this reputation model and SC2: (1) fully integrating the model into SC2, (2) developing a narrative experience for the player, and (3) improving the value measures in the model. The team began this project with little knowledge about wargaming and reputation. Given the research conducted this year and the improved reputation model we developed, the next capstone team will be able to utilize the lines of effort we provide to take this project to the next level.

5.1 Integration of the Model

One focus of future research on SC2 is restructuring the wargame to fit the new reputation model and calculations. The proposed solution is to utilize a new game user interface (GUI) that would allow for improved playability and further data analysis. Using a transition from a Microsoft Form to Microsoft Excel, the GUI guides player actions without restricting decision making and frees up game facilitators to focus on the players in the game. The calculations done in Excel can then be changed in real time and displayed clearly for the players to see. After each game, the Excel file serves as an archive of player actions and effects on reputation that can then be analyzed further through visualizations and decision analysis.

Combining the two efforts of the new model for reputation and the new GUI is limited due to the current format of the game. The current format of SC2 features three measures, diplomatic, military, and economic on scales of 1 to 6. The new reputation model factors in smaller value measures to produce an overall reputation score. To merge the two efforts, the new reputation value measures will be grouped into the three categories of diplomatic, military, and economic and then rescaled to fit the current 1 to 6 scale. This will allow for the gameplay to continue as it currently does with the new GUI while the new reputation model provides more comprehensive starting values for reputation. Further research will focus on changing the game mechanics to support changes to the values measures of the new model.

5.2 A New Narrative

Another focus of future research should be to create a more tangible narrative experience for SC2. SC2's primary purpose is to simulate real world scenarios that decision makers may encounter and provide them an opportunity to gain understanding in specific aspects of modern military competition and crisis. Future work should create a way for players to easily see the "story" of the wargame as it unfolds throughout the gameplay. This could take form in a digital story board format where the GUI would display the actions each player, and what effects resulted. As the game progresses, those actions and effects would be mapped across subsequent actions and effects, providing players with an easily understood timeline/map of how their actions affected their state's reputation from turn to turn. An easy-to-understand visualization of the entire game is critical to ensuring the wargame's purpose is met and players receive maximum key takeaways. Improving the way SC2 tracks and displays the narrative of the game will greatly improve the facilitation of player learning.

5.3 Improved Value Measures

The final proposed focus for future work pertains to improving the model as it is. While the current model contains a well-developed list of value measures, there remains many other factors affecting reputation that deserve investigation. Some areas of interest include energy-related value measures such as oil and gas. Another is the degree to which a state intends to use its nuclear weapons. Future research should conduct another round of rigorous stakeholder analysis via improvements to the existing survey. Further efforts could also focus on finding more credible datasets and metrics. The current methodology provides a solid framework, but future capstone groups could develop this reputation model into an asset for improving military senior leaders' understanding the effects of military activities on state reputation in the modern military competition environment.

6. References

- CIA.gov. The World Factbook. The World Factbook – Explore All countries. <https://www.cia.gov/the-world-factbook/countries/>.
- Engelmann, K. & Kearney, P. (2021). Global Strategic Competition and Crisis Strategy Game [Slide 26]. Center for Army Analysis.
- Frederick, B., Watts, S., Lane, M., Doll, A., Rhoades, A. L., & Smith, M. L. (2020). *Understanding the Deterrent Impact of U.S. Overseas Forces*. Santa Monica, Calif.: RAND Corporation.
- Headquarters, Department of the Army. (2021). *Chief of Staff Paper #2: The Army in Military Competition*.
- Keohane, R. O. (1999). International Relations and International Law: Interests, Reputation, Institutions. *Proceedings of the Annual Meeting (American Society of International Law)*, 93, 375–379. <http://www.jstor.org/stable/25659335>.
- Lemahieu, H. & Leng, A. (2021). *Lowy Institute Asia Power Index: Key Findings 2021*. Lowy Institute. <https://power.lowyinstitute.org/downloads/lowy-institute-2021-asia-power-index-key-findings-report.pdf>.
- Lupton, D. L. (2020). How leader-specific reputations form and change across repeated interactions. In *Reputation for resolve: how leaders signal determination in international politics*, 47-69. Cornell University Press.
- Parnell, G., et al. (2011). *Decision Making in Systems Engineering and Management*. Wiley Series. 2011.
- Opsahl, T., Agneessens, F., & Skvoretz, J. (2010). Node centrality in weighted networks: Generalizing degree and shortest paths. *Social Networks*, 32(3), 245–251. <https://doi.org/10.1016/j.socnet.2010.03.006>.
- Russett, B. (1993). The Face of Democratic Peace. In *Grasping the Democratic Peace*. Princeton University Press.
- Schelling, T. C. (1966). THE MANIPULATION OF RISK. In *Arms and Influence* (pp. 92–125). Yale University Press. <http://www.jstor.org/stable/j.ctt5vm52s.6>.
- The World Bank Group. (2020). *Worldwide Governance Indicators | DataBank* [Dataset]. <https://databank.worldbank.org/source/worldwide-governance-indicators#>.
- Walt, S. M. (1985). Alliance Formation and the Balance of World Power. *International Security*, 9(4), 3–43. <https://doi.org/10.2307/2538540>.
- Worley, D. R. (2012). *Orchestrating the Instruments of Power: A Critical Examination of the U.S. National Security System*, Lulu Press Inc, 275-276.
- X. Li, J. Zhao, H. Li, M. Cheng and Z. Fan. (2013). Research on systems modeling and analysis method based on wargaming, *2013 IEEE International Conference on Information and Automation (ICIA)*, 242-247, doi:10.1109/ICInfA.2013.6720303.
- Yarhi-Milo, Keren. (2013). In the eye of the beholder: how leaders and intelligence communities assess the intentions of adversaries. *International security*, 7-51.