Applied computational social choice theory as a framework for new cyber-threats

David M. Perlman, Ph.D.
CoPsyCon.org
San Francisco, California, USA
david@copsycon.org

***Abstract*—Social media and big data have combined to create a new era of marketing, political campaigning, and hostile propaganda. The *tactics*, such as microtargeting of ads, have recently received intense public scrutiny. However, little has been publicly said about the tools and techniques of *strategy*. In this context, Applied Computational Choice (ACSC) refers to a framework for analyzing data, modeling tactics, and planning strategy. Here we describe an ACSC framework derived from the work being done by some of the main actors, and apply it to show how a few simple scenarios can be modeled and realistic behaviors predicted, as well as illuminating possible motivations for certain patterns observed in the real world. We introduce the concept of vulnerability assessment applied to voting systems by analyzing the cost of influence operations on simple model voting systems. We believe this framework reflects those being used by a number of different actors with various goals, and hope that this article helps provide an overview and introduction to the field.**

***Index Terms*—data science, political science, propaganda, influence, information warfare, narrative warfare, weaponized demographic**

I. Introduction

Although adversarial propaganda is as old as war itself, recently new techniques have been implemented with unprecedented power, speed and effectiveness in a number of political contests around the world. Technological advances in the application of Computational Social Choice Theory [1], mass profiling [2], and microtargeting [3], [4], have been developed by a number of sources, including tech media companies, private marketing and campaign data businesses, as well as hostile state and/or non-state actors. In addition to the societal ethical concerns, it is now apparent that hostile actors have developed extensive art and proficiency in using these technologies offensively in ways that are critically relevant to national security and the military [5]–[8]. Broad awareness of this threat is newly dawning, and terminology is not yet standardized; various terms are used including Cyber-enabled Information Operations (CyIO) [6], [9], Information Warfare and Influence Operations (IWIO) [10], or Information/Influence Warfare and Manipulation (IIWAM) [5]. The battlefield of the new information warfare is the information environment [5], [7] especially "social media" [10], the globally pervasive sphere of media-rich personal and social communication that has evolved out of the original handful of social networks, which is perfectly suited to communicating emotional information that bypasses rational filtering. The weapons and tactics are narratives [11]; the delivery vehicles are "memes" (in the original sense, as well as the contemporary meaning), units of information that are tuned for, first, rapid propagation by the humans in the social network, and ultimately, assimilation into the mass mentality for the promotion of disruptive and harmful politicians or agendas, and other political and social goals [12]. With respect to these cutting-edge social media techniques, a large portion of the publicity has recently been focused on Cambridge Analytica (CA), Facebook (FB) and the Internet Research Agency (IRA) in St. Petersburg, but it is virtually certain that actors all over the world are engaged in research and deployment of these techniques.

In part, ACSC is a natural extension of advertising and marketing science developed in the context of this social media sphere and a highly competitive, largely unregulated marketplace. The presence of a significant fraction of the world's population on social networks turns mass psychology and behavior manipulation into computational "big data" problems. Social Choice theory and the models described here originate in the economics and political science literature as far back as 1957 [13], with extensive theoretical work in the 1960s and 1970s [14]–[17]. Computational approaches to social choice theory appeared more recently, [1] and continue in earnest. Practical, technological applications of computational social choice theory became possible only very recently in the age of "Big Data" and social media, and serious academic research has only become prominent since around 2016. There is, obviously, an enormous market for understanding and influencing population psychology and behavior; the most famous companies of our era, such as Google, Facebook, etc. spend much of their effort studying this domain and developing techniques, tools, algorithms, and other kinds of expertise. In addition, it seems clear by now that some state actors have devoted tremendous time and attention to understanding the role of social media and the internet in population influence. However, both the private corporations and the governments doing this work have powerful incentives to keep their innovations secret, and so little has yet been published of the formalism and techniques useful in this domain. When private research has been published, it has generated strident criticism [18], [19]. Thus it is not surprising that research and implementation of applications of these technologies are largely hidden behind a veil of secrecy. Here we share a general framework or formalism for political data science which we believe is representative of how some of these actors may be operating, and then apply the formalism to suggest qualitative outlines of how several plausible scenarios could be conducted.

The information in this article has been inferred from extensive conversations with a number of individuals across various related fields, and synthesized with general principles of data science and linear algebra to form this framework. Some of the contributing individuals have reviewed the framework and, without divulging trade secrets, have agreed that it is compatible with many of the important considerations they would raise. It is my hope that this article will serve as a useful introduction to what might be possible, and to provide a common terminology and framework for those wishing to study ACSC more openly. The goal of this article is to synthesize a framework that is optimized for practical applications in industry and defense, specifically the application of identifying, understanding, and countering large-scale influence operations in the big-data context of online digital platforms. To serve this goal, the primary considerations have been utilitarian rather than theoretical: little of the material is theoretically novel; it has been selected from other sources to be useful for this endeavor, and no effort has been made to comprehensively cover mathematical voting theory or any other existing field. The specific examples given are not intended to prove that any specific activities actually happened in the real world (although we have our suspicions!) but instead are intended to stimulate intuition with respect to what is possible in this domain and encourage and support more researchers wishing to enter this field.

We have intentionally avoided two crucial topics, not because they are unimportant, but because they are already receiving extensive attention: individual psychology, and agent- and network-based simulations of organic social behavior. There is currently a cottage industry around fake news, misinformation, manipulation, and bias, and a thoughtful awareness of much of the most important work in the field is readily available to the public and professionals alike. Likewise, many people are investigating the natural patterns of propagation of information through social networks and the formation of cliques and cults. The goals that we hope to serve with this paper ultimately will rest on a foundation of knowledge of individual cognition and emotion, and emergent aggregate phenomena. Here we address only the edge of the field that we believe to be most critically underserved.

II. The Framework

A. Ideological Space

We will follow the traditional structure of computational social choice theory [14], [16], [17], [20], [21]. We construct a preference space over political ideologies, which we will call "ideological space" or "policy space". Consider a population B_N of N individual agents in a particular society, so that they share some set of K issues of political or ideological interest to them all. Without loss of generality, we can represent each political/ideological issue as a real number in $[-1, 1]$ corresponding to the agent's response to an issue question on a continuous Likert-type agree-disagree scale. Each individual's preferences are represented by a K-dimensional vector $\mathbf{b}$; we will write $\mathbf{b}_i$ for the preference vector for individual i. We will assume a K-dimensional Euclidean coordinate space defined by taking each of the K issues as a dimension. (In some unusual applications, the assumption of a Euclidean space may be limiting, but is a tremendously useful starting point because it facilitates conceptual intuition as well as computation.) Within this Euclidean space, the "ideological space" of all possible configurations of beliefs is P^K, the K-dimensional cube centered on the origin with edges of length 2. Thus the complete population preference set B_N^K is a set of N points inside the K-dimensional cube P^K, $B_N^K = \{\mathbf{b}_i \in P^K : i \in [1, N]\}$.

Of the K issues/dimensions, it is likely that some issues are highly correlated with each other. Assume that it is possible to apply an appropriate dimension-reduction process to determine the actual number of underlying latent ideological dimensions k. This defines a reduced ideological space P^k, a k-dimensional cube in k-dimensional Euclidean space, and a reduced population preference set B_N^k

$$B_N^k = \{\mathbf{b}_i \in P^k : i \in [1, N]\} \tag{1}$$

The k dimensions might be thought of as the fundamental philosophical, moral, emotional, etc. beliefs that form the foundation for individuals' preferences on policy issues. It would require empirical investigation to assign meaning to these k reduced ideological dimensions, and in practice in large-scale applications of this framework it may not be possible to extract dimensions that appear meaningful in human terms. In fact, in big-data applications generally, dimension reduction is a non-obvious problem. It is likely that much work is happening behind closed doors to develop these basic techniques.

B. New Technology for Surveying the Ideological Space

The traditional method of estimating B_N^k is, essentially, opinion polling. Political polls that attempt to model "likely voters" are attempting to estimate a reduced set $B_{n_{vt}}^k$ and a turnout function VT (see equation 2 below). Recently, there has been a great deal of attention given to the possibility of measuring psychological profiles, political preferences, etc. from online social media behavior, search history, and other internet sources [2]–[4], [22]–[24]. Most of the discussion of this topic has focused on the individual level: the privacy implications of collecting and modeling personal information without an individual's knowledge, and the implications for personal autonomy of precisely microtargeted advertising or propaganda. These new techniques may also allow greatly improved speed and accuracy of estimation of B_N^k: increased speed due to the use of massive online databases that contain daily or even faster updates for many people, and increased accuracy due to the freedom from response biases with surreptitious modeling, as well as the very large sample sizes available. This increased power is what makes possible rapid and powerful operations such as short-term manipulation of political preferences before an election, effective on the order of days or weeks [25], leaving no time for any effective tactical response. This is especially true for Western liberal democracies that do not currently have national defense capabilities in this domain at all.

C. The Curse of Ideological Dimensionality

The first insights from the framework can be gleaned by considering the question of the actual dimensionality k of the ideological space P^k for real-world societies. A plausible guess is that it would be comparable to the number of distinct political issues identifiable in the news and other media of the society at any given time. In general, the specification of the analytical framework will be different depending on the goals of the given application, especially the population in question and the time frame in question. The necessary value of k for a useful application of the model will be greater if we wish to study how a population's beliefs evolve over time, because some issues will be forgotten as new ones arise, but the model must contain dimensions for all of them in order to represent this drift. If we wish to study a population over a long stretch of time, then the necessary k may be very large. The necessary k will also be greater if we wish to study a larger and/or more diverse population, because although any individual may only be aware of a small number of issues, across a large population, a larger number of issues will be represented. Importantly, in many real-world societies, it is also likely that k has recently increased as the proliferation of digital news media sources has increased the total number of different issues of which the citizens are collectively aware.

As the number of ideological dimensions k increases, a number of practically-relevant phenomena can be expected based on the "curse of dimensionality" [21], [26], [27]:

1) Individuals disagree with each other more: The expected distance between any two randomly chosen points increases.
2) The population overall becomes more dissatisfied with any platform that specifies a complete set of policies, e.g. the actual government policy at any moment: The average distance between all the points in B^k_N and the single point G^k that represents the platform increases.
3) The potential for insurmountable disagreements increases: The maximum possible distance between any two points increases even faster than the expected distance.
4) A smaller proportion of the ideological space is taken up by moderates, and a greater proportion by the fringe, even for a highly inclusive definition of moderate: We can choose a reasonable definition of "moderate" in the policy space as being "near the center", and represent this with $S^k_{r_{mod}}$, the centered k-ball of radius $rmod$, i.e. the region of policy space that is within the distance r_{mod} of the center. Then the ratio of the volume of $S^k_{r_{mod}}$ to the volume of the policy space P_k goes to zero as k increases, even if we choose $r_{mod} = 1$.
5) Extending that, depending on the population's distribution of beliefs, more and more of the population will find that their values and beliefs fall outside of any of the available political parties: If we represent p political parties as p non-overlapping k-balls of radius r_i, S^k_{i,r_i}, the ratio of the total volume of all the parties' territories $\sum_{i=1}^{p}(S^k_{i,r_i})$ to the total volume of P_k also goes to zero.
6) The two previous points, taken together, implies that if a political party uses this kind of data analysis in their electoral or marketing strategic planning, they will be motivated to expand their ideological-space territory to include more and more of the fringe in an attempt to capture more of the electorate.

See the Appendix for mathematical derivations of these effects.

III. Democracy: Policy from Population

A. Voting Models

Democracy, in a very general sense, refers to a system in which the policies and actions implemented by the government are intended to be consistent with the will of the population. Intuitively, this means that the point G^k that represents actual government policy in ideological policy space ought to be "in the middle" of the density of the point cloud B^k_N. We define a policy-choice function g that takes B^k_N as input and yields G^k as output. We further define g to be the composite $g = h(\tau)$, where τ is a turnout function and h is a complete-turnout voting function. The turnout function τ determines a subset of B^k_N, $B^k_{n_\tau}$:

$$B^k_{n_\tau} = \tau(B^k_N) \tag{2}$$

The voting function h takes $B^k_{n_\tau}$ as input and yields a single point for G^k. Thus:

$$G^k = g(B^k_N) = h(\tau(B^k_N)) = h(B^k_{n_\tau}) \tag{3}$$

B. Turnout

A great deal of complexity and uncertainty is hidden in τ. Polling services devote large resources to modeling voter turnout, with limited success. Historically there have been a number of famously embarrassing and disruptive prediction errors based on errors in turnout modeling, such as the classic "Dewey Defeats Truman" headline. The outcomes of elections can be dramatically altered by changing turnout, and there is already reason to believe that hostile actors have engaged in microtargeted social media campaigns primarily oriented around voter suppression. Future expansion of this manuscript will include examples that consider the implications of changing τ in this framework.

C. *Simple Examples*

a) Technocratic Direct Democracy: The intuitive criterion that a democracy should yield actual government policies G^k that are "somewhere in the middle" carries over into h. As the (mathematically, not practically) simplest possible example, one can imagine a hypothetical "technocratic direct democracy" (TDD) where the full population has their policy preferences measured and then G^k is set at the average or centroid of B^k_N:

$$G^k_{\text{TDD}} = \bar{\mathbf{b}} = mean(B^k_N) \tag{4}$$

b) Two-Party Direct Democracy: Now we expand that reductionist model to include one additional element of complexity. Consider now the simplest possible example of a two-party voting system, which we might call "2-party direct democracy" (2PDD). B^k_n is divided into $B^k_{n_1}$ for the n_1 voters of Party 1 and $B^k_{n_2}$ for the n_2 voters of Party 2. Party 1 evaluates the preferences of their constituency and defines a platform G^k_1 as the centroid [1] of $B^k_{n_1}$ and Party 2 likewise defines G^k_2 as the centroid of $B^k_{n_2}$. The implementation of the voting function then yields

$$\begin{gathered} G^k_{\text{2PDD}} = h(B^k_{n_1} \cup B^k_{n_2}) = G^k_j \\ j = \underset{i \in \{1,2\}}{\arg\max}\, n_i \end{gathered} \tag{5}$$

c) Dictatorship: For comparison purposes, we can also describe G^k_{dictator} as a G^k that is dictated without regard to B^k_N. This may be thought of as a voting function g that is constant. Or, if other factors are known and available to be modeled, g may be a function of those other factors.

d) Other examples: We can also describe a number of other simplified example scenarios. Future expansion of this manuscript will describe:

- Indirect democracy: Voting Districts, and an "Electoral College"
- Analyzing the effects of different voting systems such as First Past The Post and Ranked Choice Voting.
- Turnout defined over districts or other clusters
- Turnout as a selector function versus probability field T' over P_k
- Iterative feedback between political parties' selection of issue-space territories and voters' party alignment

IV. Population influence

A. *Influence Cost Function*

Among the three examples of G^k_{dictator}, G^k_{TDD}, and G^k_{2PDD}, we can consider what would be necessary for an influence operation to change policy by looking at how changes in B^k_N affect G_k. To allow comparisons, we can define a metric of "influence cost" for a change from B^k_N to B'^k_N. The simplest metric is based on unweighted sum of Euclidean distances moved by each individual:

$$C_{\text{unweighted}}(B, B') = \sum_{i=1}^{N} |(\left\|B'^k_i - B^k_i\right\|_2)| \tag{6}$$

where $|\cdot|$ is the numerical absolute value and $\|\cdot\|_2$ is the (k-dimensional) Euclidean norm applied row-wise to the differences of the ith rows of $B'^k_i - B^k_i$. This can also be written

$$C_{\text{unweighted}}(B, B') = \left\|B'^k_i - B^k_i\right\|_{1,2} \tag{7}$$

where $\|\cdot\|_{1,2}$ is the $L_{1,2}$ matrix norm, for row-wise data points in a matrix.

We can abbreviate:

$$C_{\text{unweighted}}(\Delta B) = \|\Delta B\|_{1,2} \tag{8}$$

B. *Weighted Cost Functions*

Different individuals will have different susceptibility to influence. To take this into account we can add weights w_i for each individual, represented in an $N \times N$ diagonal weight matrix W:

$$C_W(\Delta B) = \|W \Delta B\|_{1,2} \tag{9}$$

Different preference dimensions of the ideological space may have different "stickiness", as well; some may be easier to change people's minds about than others. To account for this we can add another $k \times k$ diagonal weight matrix V with the weights v_j for each of the k preference dimensions:

$$C_{WV}(\Delta B) = \|W \Delta B V\|_{1,2} \tag{10}$$

C. *Example scenarios*

We will look at some simple "back-of-the-envelope" calculations of the cost of influence operations to explore the possibilities within the framework.

a) Dictatorship: In G^k_{dictator}, g is not a function of B^k_n. The most direct way to change G^k_{dictator} would be to influence the dictator individually. Influence on B^k_n leads to changes in G^k_{dictator} only to the extent that the dictator notices, cares, and reacts to the population change—or, under a coup.

Call the region of ideological space that represents willingness to act on a grassroots coup Q, and call the minimum number of individuals necessary for a coup n_Q. Then the cost metric for influencing a coup is

$$C_{\text{dictator}}(Q) = \sum_{n_Q} \min(\left\|B^k_i - Q\right\|) \tag{11}$$

Here $\min(\left\|B^k_i - Q\right\|)$ refers to the distance from the point B^k_i to the nearest point in Q.

[1] The centroid or mean of the constituency in the Euclidean space is neither plausibly realistic nor strategically optimal as an actual real-world choice of platform for a party. Any number of other considerations would come into play in the real world, especially turnout, loyalty, and other non-policy effects, and there are also evolutionary and iterative effects in the emergence of parties, see for example [21], [28] as a tiny selection of arbitrarily-selected (and not at all centroidal) examples of greater complexity. Our use of the centroid here is purely motivated by the choice of the computationally simplest starting point for this exposition.

Qualitatively, with a few straightforward assumptions, we can interpret:

1) The cost of influencing a coup is proportional to the number of people who must be induced to participate, which is determined by the strength of the regime.
2) The cost of influencing a coup depends on how far the relevant slice of the population is from the "boiling point" Q^k. In other words, it's easier to induce a coup in a population that is already dissatisfied.

This suggests that it may be possible for an actor with access only to data such as search and social media to remotely estimate the likelihood of regime change with little direct interaction.

b) Technocratic Direct Democracy: Consider the goal of moving G^k_{TDD} to a target Q. According to (4)

$$\begin{aligned} G^k_{\text{TDD}} &= \text{mean}(B^k_N) \\ Q &= \text{mean}(B'^k_N) \\ \Delta B = Q - G^k_{\text{TDD}} &= \text{mean}(B'^k_N) - \text{mean}(B^k_N) \\ &= \text{mean}(B'^k_N - B^k_N) \\ &= \text{mean}(\Delta B^k_N) \end{aligned} \tag{12}$$

The specification of a target does not uniquely determine the influence cost, because we do not know the trajectories of all the individuals; however, the minimum possible influence cost arises in the situation where each individual moves in parallel to the overall movement, in the "forwards" direction:

$$\min C_{\text{TDD}} = N \times \left\|\Delta B^k_N\right\| \tag{13}$$

Under this hypothetical minimal system (and ignoring for the moment the per-person and per-issue weights), there are no influence shortcuts: the cost of influence is proportional to the total population and the magnitude of the targeted change. In future work, we will present calculations suggesting that adding further mechanisms of complexity to the system will lead to more complex influence cost functions, which will have some variables that lead to greater costs and others that lead to less. An actor that performs a detailed analysis of the complete set of mechanisms of a voting system will be able to identify weak points that constitute the influence version of attack surfaces, and design influence campaigns fine-tuned for the maximum socio-political impact with minimum cost. In turn, this tells us that a democracy must perform this same detailed vulnerability assessment of its own voting systems in order to defend effectively against influence attacks which could have devastating, paralyzing consequences.

V. The Overton Hull

A. The Original Overton Window

The Overton Window is a concept first put forth by Joe Overton of the Mackinack Center to refer to the range of public political discourse that is tolerated within a given society's media ecosystem. The original concept referred to the segment on a unidimensional left-right political spectrum that represents the positions that, say, a politician can publicly profess and still expect to be taken seriously. It is important to clarify that the Overton Window is a population-level concept: while it may be reasonable to talk about a "window" that an individual is willing to tolerate, we are interested in studying a society as a whole, so the concept in question relates to the emergent "window" across the society's whole media ecosystem.

B. Extending to k dimensions

The idea of a unidimensional left-right spectrum is certainly used for simple rhetoric, but in order to make the Overton Window practically useful we extend it here to a k-dimensional "blob", the region within the ideological space P^k that represents those views that are acceptable within the media ecosystem of the society in question. Although it is rarely stated explicitly, discussions of the Overton Window universally assume that the Window is a single connected line segment. We can generalize that to define the "Overton Hull" H_{O} as a convex region of P^k that represents the range of political ideological views that are acceptable within the media ecosystem of population B_N.

C. Estimating the Overton Hull

Discussions of the Overton Window usually assume that it is approximately centered around the bulk of the distribution of the population along the political spectrum, or in other words, the majority's political beliefs are within the Window. We can make a first pass at a simple working definition of a measured H_{O} with P^k and B^k_N as our starting point.

First we postulate that B^k_N is a sample drawn from a distribution with probability density function f_B defined over the sample space P^k. Then H_{O} can be defined as the convex hull of the region of P^k in which f_B is over a threshold value d_{thr}:

$$H_{\text{O}} = \text{Conv}\left\{\, \mathbf{a} \in P^k \mid f_B(\mathbf{a}) \geq d_{\text{thr}} \,\right\} \tag{14}$$

VI. Weaponized Demographics

"Useful Idiots" is a term widely used since the Cold War to refer to individuals who are easily manipulated into serving hostile propaganda purposes even though they may not actually support or even understand the issues at stake. In order to study the role of Useful Idiots in influence operations at the population, rather than individual, level, we can describe a "Useful Idiot Demographic" B_{UI} as a subpopulation whose ideological preferences are particularly easy to manipulate. This ease of manipulation can be represented as low values of the cost function weights below a UI threshold $w_i < w_{\text{UIthr}}$ for these individuals:

$$B_{\text{UI}} = \left\{\, \mathbf{b}_i \in B^k_N \mid w_i < w_{\text{UIthr}} \,\right\} \tag{15}$$

With these definitions, we can describe the Weaponized Useful Idiot Demographic (WUID), a mass-influence technique derived from the "Door-in-the-face" (DITF) frequently discussed in the literature on the Overton Window [29]–[31].

We will also describe influencing H_{O} using traditional mass propaganda to provide a baseline for comparison.

A. Mass Propaganda, or The Bulk Move

Consider a target point Q which is outside the Overton Hull H_{O}; using traditional methods of non-targeted mass propaganda operating on the population at large, the attacker wants to move H_{O} to include Q. Let q_{surf} be the closest point to Q on the hull of H_{O} and q_{sq} be the vector from q_{surf} to Q, so that $\|q_{\mathrm{sq}}\|$ is the minimum distance from H_{O} to Q. In order to "Bulk Move" the whole population's average preferences over until Q is just inside H_{O}, we know from the consideration of C_{TDD} above that the cost will be approximately $C_{\mathrm{BM}} \approx N \times \|q_{\mathrm{sq}}\|$. We can now use this a baseline for comparing WUID.

B. The Weaponized Useful Idiot Demographic

The WUID is a formalized variant of the "Door in the Face" technique frequently discussed in the literature on the Overton Window. Let the attacker choose a "dummy target" point Q' near $l \times q_{\mathrm{sq}} + q_{\mathrm{surf}}$, which represents a "more extreme" version of the real target Q, in the sense that it is farther away from H_{O}, with the factor l determining "how much more extreme" it is. Because H_{O} is a convex hull, if the density function f_B can be raised above d_{thr} in even a tiny region around Q', then Q will immediately be included well within the Overton Hull.

To accomplish this, choose a small subpopulation of M individuals from the "useful idiots" demographic, $B_{\mathrm{WUID}} \subset B_{\mathrm{UI}}$, where $M = m \times N$, $m \ll 1$. Although we can reasonably anticipate that the individuals in B_{WUID} will have markedly different individual preferences than the population at large, in the absence of any reason to believe they have a specific direction of political bias, assume that to begin with the average preferences of B_{WUID} are approximately the same as the average for the population at large, $\bar{\mathbf{b}}_{\mathrm{WUID}} \approx \bar{\mathbf{b}}$. Then

$$\begin{aligned} \|q_{\mathrm{WUID}}\| &\approx l \times \|q_{\mathrm{sq}}\| + \|q_{\mathrm{surf}} - \bar{\mathbf{b}}\| \\ &\approx l \times \|q_{\mathrm{sq}}\| \text{ if } l \gg 1 \end{aligned} \tag{16}$$

If the extremeness factor l is great enough, then the target movement distance $\|q_{\mathrm{WUID}}\| \approx l \times \|q_{\mathrm{sq}}\|$. This means that we expect the influence distance to be much greater in this case. However, consider the influence cost. Based on the derivation for C_{TDD}, we can see that

$$\begin{aligned} C_{\mathrm{WUID}} &\approx \bar{w}_{\mathrm{UI}} \times M \times l \times \|q_{\mathrm{sq}}\| \\ &\approx \bar{w}_{\mathrm{UI}} \times m \times N \times l \times \|q_{\mathrm{sq}}\| \\ &\approx \bar{w}_{\mathrm{UI}} \times m \times l \times C_{\mathrm{BM}} \end{aligned} \tag{17}$$

By definition, we know that $l \gg 1$; but $m \ll 1$ and also $\bar{w}_{\mathrm{UI}} \ll 1$. This means that there is ample opportunity for a well-planned operation to have influence cost much lower than that of traditional mass propaganda, $C_{\mathrm{WUID}} \ll C_{\mathrm{BM}}$.

In qualitative terms, we can describe this operation as follows. First, the attacker identifies a particularly gullible demographic of "useful idiots" who are likely to be scattered around the fringes of society in their various beliefs. The attacker uses social media, search history, etc. to profile them and prepare targeted narrative weaponry. The narratives might extensively incorporate the language of conspiracy theories to appeal to the fringe psychology. Next, the attacker uses microtargeted, viral, and mass-media delivery vehicles for the narrative weaponry to "lasso the fringe" into a Weaponized Useful Idiot Demographic (WUID) that the attacker now has some degree of control over. The WUID is induced to create a media-noticeable prevalence of dummy target ideology Q', which immediately opens up the Overton Hull to include Q, thus accomplishing the attacker's goals faster and with much less cost than would be possible with traditional mass propaganda.[2] In fact, the attacker receives even more benefit from the WUID: this is, in essence, a reusable weapon; once the WUID has become accustomed to taking their cues from certain sources, they are likely to remain open to those sources for some time.

C. Defense and Counter-offense

Having considered the potential power of the WUID attack, naturally questions of defense and counter-offense arise. In the long run, the best defense against this attack or any other techniques of influence and propaganda is a well-educated population with a strong sense of national identity founded on principles of tolerance, generosity, openness to diversity, and service to others; especially important are critical thinking skills, the ability to weigh evidence and reject implausible fringe theories, and a realistic respect for the value of established authorities and institutions. The only way to prevent narrative warfare from spreading out from individual victims to mass societal effect is to reduce the systemic vulnerabilities and attack surfaces, the "cracks in our society" that come from ignorance and divisive factionalism. However, in the short run, there is a pressing need for rapidly deployable tactics. We believe that here, as in other forms of narrative warfare, playing defense is a losing strategy. While prevention is the best strategy, once an attack has taken place and the WUID has become entrenched, we believe the most effective tactic will be a counter-offensive. The key observation of the WUID is that in order to be effective, it must remain coordinated. In order for the density spike created in f_B by B_{WUID} to remain high enough to exceed the threshold d_{thr}, the individuals must be clustered close together in the preference space P^k; if they drift apart, then they are no longer an effective weapon. This exposes a weakness in the attacker's weapon that could be exploited by instigating counter-offensive targeted narratives designed to disrupt the unity of B_{WUID} as well as disrupt the narratives the attackers use to direct the WUID. In future work, we will consider possible counter-offensive techniques in greater detail.

[2] If the Weaponized Demographic of Useful Idiots were to be used to influence an election and install a puppet government, perhaps that government could then be referred to as a "Useful Idiocracy".

VII. Concluding remarks

Narrative warfare and propaganda are as old as warfare itself. Self-propagating units of information are a newer concept in the digital age, and are core to the established field of cybersecurity. Viruses, worms, Trojan horses, etc. are well understood and thoroughly monitored. However, the recent surge of innovation in socio-technical systems such as social networks, weaponized memes, and big data, has opened up a new era of conflict, in which an adversary can, for example, rapidly manipulate popular sentiment to swing an election with only days or hours of lead time, faster than any currently possible response. This paper does not attempt to solve these problems immediately. Rather, our goal has been to describe a framework and a way of thinking about ACSC and CyIO that experienced actors already use to analyze populations and plan operations. By making this introduction widely available to friendly actors we hope to support defensive innovation and lead to improvements in the current situation, in which the U.S. and allies have been severely outpaced in this domain.

Information warfare takes place on a high-dimensional abstract battlefield, which makes monitoring and planning extremely difficult. One promise of this framework is the development of technology for situation awareness and battlefield visualization in near-real-time. There is already a rich field of research and practical application of tools for meaning-extraction and visualization of high-dimensional data sets, which could be adapted to create tactical battlefield displays for real-time awareness, planning and defense against CyIO operations as they unfold.

With moves such as the U.S. Department of Homeland Security designation of elections systems as Critical Infrastructure, the world is acknowledging the need for physical and cyber security in election systems. However, the emergence of powerful CyIO capabilities is a qualitatively new development, and it is currently debatable whether it even falls in the wheelhouse of cybersecurity and cyberwarfare. The concept of national security risk from population-level influence weaknesses, attack surfaces and vulnerabilities not in voting machines[3] but in elections and social choice systems themselves, may not be on anyone's radar screen at all. And yet, these weaknesses may have already been recognized, analyzed, and exploited by hostile actors against democratic systems around the world. We believe that this type of approach can help analyze the vulnerabilities of our own voting systems and recommend improvements. We also hope that a more formal framework for such analysis could generate greater clarity and objectivity about risks and recommendations, and this objectivity in turn might depoliticize election security.

As an example of applying this framework to understand and describe narrative warfare attacks, we explored the idea of a Weaponized Useful Idiot Demographic (WUID). This technique creates a population-level "hammer" that could be wielded with great effectiveness in CyIO operations. The framework of ACSC allows the comparison and evaluation of different operations and tentative measurement of their potential cost and effectiveness. We suspect that WUID-type operations may have already been used with great economy and effectiveness against the U.S., and so it is particularly important to study this type of tactic and develop defenses and counter-offenses.

It is important to note that the assumptions of a linear space and Euclidean metrics are limiting, and in fact there is empirical evidence of subpopulations moving through "wormholes" in the sense that they abruptly shift from one region of P^k to another without seeming to traverse the intervening territory, as well as other strange effects. We introduce only the most elementary election theory here, and acknowledge that there is a huge body of literature that we are glossing over. In particular, recent work critiquing Median Voter Theory is relevant and will certainly inform refinement and changes in this type of framework in the future [32], [33]. Furthermore, the relegation of all of psychology into the weight matrices W and V is a radical oversimplification, to say the least. Nevertheless, the first step towards making any new field theoretically and computationally tractable is to define a mathematical framework that fits reasonably well. This allows the deviations from simple behavior to be quantified, which in turn allows the model to be expanded with appropriately defined weight matrices, locally Euclidean manifold techniques, etc. Our goal with this paper is to help open the doors to this field; we have no illusions of completeness of anything presented here.

Much has been written elsewhere about the greater susceptibility of Western liberal democracies to propaganda and narrative warfare. Many nations censor or block the flow of information, even entertainment, from the U.S. and other Western nations; while we freely allow input from anywhere and enjoy media from every nation of the world. Dictatorships are insulated from the vulnerability of democracy to population-level influence because dictatorial power does not even nominally derive from the will of the people. However, the U.S. and allies do enjoy significant advantages that can potentially be applied in CyIO, especially for defense and/or counter-offense. For one, the people living under dictatorships fueled by lies and corruption usually become inured to any messaging at all from their own government, which may be leveraged by careful introduction of narratives from outside sources. The U.S. tremendously leveraged cultural "soft power" in the defeat of the Soviet Bloc in the previous Cold War [34] and at our best, our values of openness, service and tolerance can inspire and bring much of the world over to our side–a capability which we must regain. At a more technical level, the battlefield of CyIO itself is literally owned by U.S.-based corporations [8] who may, in some cases, be willing to assist in defending against these attacks. In the current climate, increased regulation of social media companies is already all but inevitable; without a solid foundation in the principles of CyIO these regulations are likely to have no beneficial effects

[3] We, of course, are strong supporters of conventional cybersecurity and especially election security and protection from hacking of databases and voter disenfranchisement. However, these are not covered in this paper.

on these risks, or even to make them worse. We hope an improved understanding of the role of social media in CyIO will guide regulatory efforts to be useful and effective towards global peace and security.

Appendix

Here we provide derivations of the various facets of the "Curse of Dimensionality" listed earlier.

1) Individuals disagree with each other more: The expected distance between any two randomly chosen points increases.

The formula for the expected distance between two points chosen on IID uniform distributions in a k-dimensional cube is extensively explored in [35]. The distance has a lower bound of $\frac{1}{3}\sqrt{k}$.

2) The population overall becomes more dissatisfied with any platform that specifies a complete set of policies, e.g. the actual government policy at any moment: The average distance between all the points in B_N^k and the single point G^k that represents the platform increases.

For simplicity we consider IID normal distributions, and place G^k at O. Clearly, the average distance we refer to here is simply $\mathrm{E}(B_i^k)$, which approaches $\sigma\sqrt{k}$ for large k [36].

3) The potential for insurmountable disagreements increases: The maximum possible distance between any two points increases even faster than the expected distance.

This is simply the distance between opposite corners of a k-dimensional hypercube, which can easily be seen to be $2\sqrt{k}$.

4) A smaller proportion of the ideological space is taken up by moderates, and a greater proportion by the fringe, even for a highly inclusive definition of moderate: We can choose a reasonable definition of "moderate" in the policy space as being "near the center", and represent this with $S^k_{r_{mod}}$, the centered k-ball of radius $rmod$, i.e. the region of policy space that is within the distance r_{mod} of the center. Then the ratio of the volume of $S^k_{r_{mod}}$ to the volume of the policy space P_k goes to zero as k increases, even if we choose $r_{mod} = 1$.

The volume of the unit k-ball goes to zero rapidly for large k, as shown in e.g. [37], and so necessarily the volume ratio to the unit cube goes to zero even more rapidly. Scaling to radius $rmod < 1$ simply adds a factor of $rmod^k < 1$ which makes the convergence even more rapid.

5) Extending that, depending on the population's distribution of beliefs, more and more of the population will find that their values and beliefs fall outside of any of the available political parties: If we represent p political parties as p non-overlapping k-balls of radius r_i, S^k_{i,r_i}, the ratio of the total volume of all the parties' territories $\sum_{i=1}^{p}(S^k_{i,r_i})$ to the total volume of P_k also goes to zero.

This can be seen to follow obviously from the previous item: the sum of several volumes, all of which converge to zero, also converges to zero.

6) The two previous points, taken together, implies that if a political party uses this kind of data analysis in their electoral or marketing strategic planning, they will be motivated to expand their ideological-space territory to include more and more of the fringe in an attempt to capture more of the electorate.

If the effective value of k is large, then expansions of a party's "ball of appeal" by a radius ratio $1+\epsilon$ increase the volume by $(1+\epsilon)^k$. More realistically, consider expansion of the "ball of appeal" along only one dimension. This expansion is equivalent to adding a k-dimensional cylindrical chunk of volume to the original ball. For simplicity consider a party that expands its appeal in two specific dimensions by a distance equal to the diameter of the ball; this is qualitatively what you might consider becoming "twice as fringey" in two topics alone while leaving all others the same. We choose the diameter rather than the radius because this adds a cylinder that perfectly circumscribes the original ball, and we choose two dimensions rather than one because the derivation of the formula for the volume of a k-ball [38] has a simple mathematical ratio:

$$V_k^B = \frac{2\pi}{k} V_{k-2}^B \tag{18}$$

The new volume added into the territory is equal to the volume of a cylinder that circumscribes the original sphere, where the cylinder has 2 "straight" dimensions and $k-2$-dimensional balls as the "ends". The volume ratio of the $k-2$-dimensional ball to the k-dimensional cylinder generated by translating the ball twice is, by simple geometry, $V_k^C = 4V_{k-2}^B$. Finally we can see that the ratio of the volume of the original k-ball to the volume of the added cylinder is:

$$\frac{V_k^B}{V_k^C} = \frac{\pi}{2k} \tag{19}$$

If $k = 10$, for example, the ratio is about 0.16. If we proceed boldly with the formalism, this implies that if a party expands their reach by just 16% on only 2 out of 10 salient issues, they can double the volume of issue-space that their territory encompasses; or conversely, likewise with $k = 10$, if a party doubles their "fringeyness" on only 2 of the 10 issues, they now encompass more than 7 *times* as much volume in their territory! [4] It is somewhat frightening to consider the possible superadditive effects of these incentives for fringe expansion, together with the power of fringe manipulation available through the WUID paradigm.

Acknowledgments

I would like to thank MAJ Mike Klipstein, Ph.D. of the Army Cyber Institute; Lt Col Jennifer J. Snow, USAF, USSOCOM Donovan Group and SOFWERX; and CDR Pablo C. Breuer, USN, USSOCOM Donovan Group and SOFWERX.

[4] This is a rule of thumb of the curse of dimensionality: spheres are weak; almost anything other than inflating a sphere will do a better job of capturing more territory.

All of them provided invaluable insights, perspective, and feedback, as well as much-needed support and encouragement. I would also like to thank C.P. Frost, Ph.D. for his expertise and feedback on voting and policy. A few individuals who wish to remain anonymous provided critical insights into the principles and mathematical terrain of online profiling, targeting, and population analysis, which made this work possible at all. Finally, I would like to thank my father, Dr. Michael D. Perlman, Ph.D. for teaching me math and statistics as a child, so that I was inclined to view and question the world through that lens.

References

[1] F. Brandt, V. Conitzer, U. Endriss, J. Lang, and A. D. Procaccia, Eds., *Handbook of Computational Social Choice*. New York: Cambridge University Press, 2016.

[2] M. Kosinski, D. Stillwell, and T. Graepel, "Private traits and attributes are predictable from digital records of human behavior," *PNAS*, pp. 5802–5805, Apr. 2013.

[3] Z. Tufekci. (2014, Jul.) Engineering the public: Big data, surveillance and computational politics. [Online]. Available: http://firstmonday.org/ojs/index.php/fm/article/view/4901/4097

[4] D. G. Wilson, "The ethics of automated behavioral microtargeting," *AI Matters*, vol. 3, no. 3, pp. 56–64, Oct. 2017.

[5] H. Lin and J. Kerr, "On Cyber-Enabled Information/Influence Warfare and Manipulation," *SSRN*, Aug. 2017.

[6] C. Watts, "Cyber-Enabled Information Operations," Apr. 2017.

[7] R. Waltzman, "The weaponization of information: the need for cognitive security," Apr. 2017.

[8] J. C. Inglis, "Statement of Chris Inglis before the Senate Armed Services Committee," Apr. 2017.

[9] D. Reynolds. (2018) Cyber-enabled information operations: The battlefield threat without a face. Jane's Defence Weekly. [Online]. Available: https://www.janes.com/images/assets/438/77438/Cyber-enabled_information_operations_The_battlefield_threat_without_a_face.pdf

[10] H. Lin. Developing Responses to Cyber-Enabled Information Warfare and Influence Operations. Lawfare Blog. [Online]. Available: https://www.lawfareblog.com/developing-responses-cyber-enabled-information-warfare-and-influence-operations

[11] A. K. Maan and P. L. Cobaugh, *Introduction to Narrative Warfare: A Primer and Study Guide*. Narrative Strategies, LLC, Jun. 2018.

[12] M. B. Prosser, "Memetics-a growth industry in US military operations," Master's thesis, United States Marine Corps School of Advanced Warfighting, 2006.

[13] A. Downs, *An economic theory of democracy*, New York, 1957.

[14] O. A. Davis, M. J. Hinich, and P. C. Ordeshook, "An Expository Development of a Mathematical Model of the Electoral Process," *The American Political Science Review*, vol. 64, no. 2, pp. 426–448, Jun. 1970.

[15] O. A. Davis and M. J. Hinich, "A mathematical model of policy formation in a democratic society," in *Mathematical Applications in Political Science*, J. L. Bernd, Ed. Dallas: Southern Methodist Press, 1967.

[16] R. Axelrod, "The structure of public opinion on policy issues," *Public Opinion Quarterly*, vol. 31, pp. 363–371, Spring 1967.

[17] R. D. McKelvey and R. E. Wendell, "Voting Equilibria in Multidimensional Choice Spaces," *Mathematics of Operations Research*, vol. 1, no. 2, pp. 144–158, May 1976.

[18] A. D. I. Kramer, J. E. Guillory, and J. T. Hancock, "Experimental evidence of massive-scale emotional contagion through social networks," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 111, no. 24, pp. 8788–8790, Jun. 2014.

[19] Goel, Vindu, "Facebook Tinkers With Users' Emotions in News Feed Experiment, Stirring Outcry," *New York Times*, Jun. 2014.

[20] F. Brandt, V. Conitzer, U. Endriss, J. Lang, and A. D. Procaccia, "Introduction to Computational Social Choice," in *Handbook of Computational Social Choice*, F. Brandt, V. Conitzer, U. Endriss, J. Lang, and A. D. Procaccia, Eds. New York: Cambridge University Press, 2016, pp. 1–29.

[21] D. Xefteris, "Multidimensional electoral competition between differentiated candidates," *Games and Economic Behavior*, vol. 105, pp. 112–121, Sep. 2017.

[22] M. Kosinski, S. C. Matz, S. D. Gosling, V. Popov, and D. Stillwell, "Facebook as a research tool for the social sciences," *American Psychologist*, vol. 70, no. 6, pp. 543–556, Sep. 2015.

[23] R. J. Gonzalez, "Hacking the citizenry?" *Anthropology Today*, vol. 33, no. 3, pp. 9–12, Jun. 2017.

[24] K. K. Roberts, "Privacy and Perceptions," *The Elon Journal of Undergraduate Research in Communications*, vol. 1, no. 1, pp. 24–34, Mar. 2010.

[25] The Economist Data Team. Support for Britain's exit from the EU is waning. The Economist. [Online]. Available: https://www.economist.com/graphic-detail/2018/06/22/support-for-britains-exit-from-the-eu-is-waning

[26] B. D. Bernheim and S. N. Slavov, "A Solution Concept for Majority Rule in Dynamic Settings," *The Review of Economic Studies*, pp. 33–62, Dec. 2008.

[27] R. E. Bellman, *Dynamic Programming*, Princeton University Press, 1957.

[28] D. Schreiber, "The Emergence of Parties," *Political Research Quarterly*, vol. 67, no. 1, pp. 136–151, Mar. 2014.

[29] Door-in-the-face technique. Wikipedia. [Online]. Available: https://en.wikipedia.org/wiki/Door-in-the-face_technique

[30] C. Brown. Morning feature: Crazy like a fox. Daily Kos. [Online]. Available: https://www.dailykos.com/stories/2009/11/5/800804/-Morning-Feature:-Crazy-Like-a-Fox

[31] J. Lehman. A brief explanation of the overton window. Mackinac Center. [Online]. Available: https://www.mackinac.org/overtonwindow

[32] T. B. Edsall, "Nothing in Moderation," *New York Times*, Oct. 2014. [Online]. Available: https://www.nytimes.com/2014/10/29/opinion/nothing-in-moderation.html

[33] E. Levitz, "Democrats Can Abandon the Center — Because the Center Doesn't Exist," *New York Magazine*, Jul. 2017. [Online]. Available: http://nymag.com/daily/intelligencer/2017/07/dems-can-abandon-the-center-because-the-center-doesnt-exist.html

[34] J. S. Nye, "Squandering the U.S. 'Soft Power' Edge," *International Educator*, pp. 4–6, Jan. 2007.

[35] E. W. Weisstein. Hypercube Line Picking. Mathworld–A Wolfram Web Resource. [Online]. Available: http://mathworld.wolfram.com/HypercubeLinePicking.html

[36] V. Chandrasekaran, B. Recht, P. Parrilo, and A. S. Willsky, "The convex geometry of linear inverse problems," in *eprint arXiv:1012.0621v3*, Apr. 2012.

[37] E. W. Weisstein. Ball. Mathworld–A Wolfram Web Resource. [Online]. Available: http://mathworld.wolfram.com/Ball.html

[38] P. A. Mariano. The volume of the unit ball in n dimensions. University of Connecticut Department of Mathematics. [Online]. Available: http://www2.math.uconn.edu/~mariano/research/MathClubsp14%20.pdf