

Robust Decision Engineering: *Collaborative Big Data and its Application to International Development/Aid*

Steve Chan, Wesley Rhodes, Charles Atencio, Caroline Kuo, Brent Ranalli, Anna Miao,
Simone Sala, Stephen Serene, Robert Helbling, and Sarah Rumbley
Prince of Wales Fellows at the MIT International Development Initiative
Massachusetts Institute of Technology
Cambridge, MA

s_chan@mit.edu, wdrhodes@mit.edu, catencio@mit.edu, caroline_kuo@brown.edu, branalli@mit.edu, annamiao@mit.edu,
salas@mit.edu, sserene@mit.edu, rjhelbl@mit.edu, srumbley@mit.edu

Marc Clement

Institute of Life Science
Swansea University
Swansea, Wales, United Kingdom
r.m.clement@swansea.ac.uk

Lisa Sokol

Office of the CTO, Software Group
IBM Federal
Bethesda, MD
lsokol@us.ibm.com

Loren Gary

Center for Public Leadership
Harvard Kennedy School
Cambridge, MA
Loren_Gary@hks.harvard.edu

Abstract—Much of the research that goes into **Big Data**, and specifically on *Collaborative Big Data*, is focused upon questions, such as: • how to get more of it? (e.g., participatory mechanisms, social media, geo-coded data from personal electronic devices) and • how to handle it? (e.g., how to ingest, sort, store, and link up disparate data sets). A question that receives far less attention is that of *Collaborative analysis* of Big Data; how can a multi-disciplinary *layered analysis* of Big Data be used to support robust decisions, especially in a collaborative setting, and especially under time pressure? The robust *Decision Engineering* required can be achieved by employing an approach related to *Network Science*, that we call *Relationship Science*. In *Relationship Science*, our methodological framework, *karassian netchain analysis* (KNA), is utilized to ascertain *islands of stability* or *positive influence dominating sets* (PIDS), so that a form of *annealed resiliency* or *latent stability* is achieved, thereby mitigating against *unintended consequences*, elements of instability, and “perfect storm” crises lurking within the network.

Index Terms—*Decision Engineering Science; Robust Decision Engineering; Complexity Ceiling; Selection Bias; Compressed Decision Cycles; Gestaltian Closure; Decision-Making; Faster Decisions; Better Decisions; Intelligent Decisions; High Adaptation Cycles; Perfect Storm Crises; Smart Power Times; Velocity, Volume, and Vectors of Big Data; Collaborative Big Data; Bigger Data; Provenanced/Pedigreed Data; Big Compute; Layered Analytics; Content Analytics, Entity Resolution, Predictive Analytics; Complexity Theory; Social Complexity Science; High Performance Computing; Computational Intelligence; Sparse Data; Sparse Networks; Social Influence Network; Participatory Revolution; Network Science; Relationship Science; Relationship Manager; 3D; 5D; Big Insights; Common Operating Picture; Aegis System; Unintended Consequences; Condition-Creating; Cyber-Physical Supply Chain; Karassian Netchain Analysis; Local Community Structures; Network Shapes; Memes; Motifs; Fifth Column; Dualistic Actors; Insider Threats; Sentiment Analysis; Flash Mobs; Islands of Stability; Sandpile Effect; Cascading Failure; Positive Influence Dominating Sets; Civil Society; Democratic Governance; Science of Development; Brittleness; Annealed Resiliency; Latent Stability.*

I. INTRODUCTION

We live in a dynamically evolving world in which the *velocity* (whereby the time needed per stock trade has decreased from several minutes to milli- and microseconds), *volume* (whereby the corpus size has swelled from *Big Data* to *Bigger Data*, via expanded tag clouds and the fusion of taxonomic with folksonomic data), and *vectors* (from symmetric to asymmetric conditions in warfare and business) of change are increasing. Ongoing globalization has increased the number of temporally-dependent involved and influential actors on the world stage, while planned technological evolution and disruptive technologies have increased global complexity, connectivity, and interdependency. This convergence of trends has resulted in a daunting 21st century *decision-making* landscape.

This complex landscape represents both opportunity and potential peril to decision makers. The rapid evolution of disparate global networks of influence has created a myriad of uncertainties (political, military, economic, social, infrastructure, informational, et al), systemic risk, and unanticipated/undesired consequences for decision makers. Within these international networks, highly dynamic interactions have led to both increased interdependencies among disparate actors and the speed of propagation at which the consequences of their activities ripple across the world. Geographic distance has become increasingly irrelevant as messages that one hundred years ago would have taken months to traverse the globe now reach their destination in a fraction of a second. These accelerating worldwide interactions reduce the time available to actively collect, analyze, evaluate, and actuate upon a burgeoning corpus of diverse data. In essence, there is less time to make decisions, and functionally speaking, the resultant “*compressed decision cycles*” pervade this *decision-making* realm.

For example, high-frequency trading (HFT) now accounts for over 70% of all U.S. equity exchange trading volume [1],

and the vast majority of volumes now traded on the New York Stock Exchange (NYSE) and other U.S. exchanges is characterized by the utilization of computer algorithms to analyze quote data so as to detect and exploit trading opportunities (often lasting only milliseconds or even microseconds). In the early 2000s, HFT had an average round-trip order execution time (latency) of several seconds, whereas by 2010, latency had decreased to milli- and microseconds. The implication of this significant decline in latencies is that trading is now so fast that HFT brokers in Chicago cannot know what the most recent quote was in geographically disparate financial centers, such as Wall Street in New York City. After all, “light travels at [approximately] 186 miles per millisecond, while the straight-line distance between New York and Chicago is 711 miles” [2]. Thus, “in 16 microseconds, light can only travel three miles (Wall Street to midtown Manhattan), and it would take 3.82 milliseconds to travel from New York to Chicago [2]. So, if one trade occurs every 16 microseconds, by the time a broker in Chicago learns about a trade in New York, 239 trades would have occurred without the broker’s knowledge” [2]. Since 2006, U.S. brokers have had a codified obligation to use “reasonable diligence” [3] to place trades in the “best market” [3] at the “best possible price” [4] under prevailing market conditions. However, when trades and quotes are changing every millisecond, it is difficult to ascertain what the best price is or even what the “prevailing market conditions” [5] are. A human decision-maker can not keep up with the speed-of-light [6] pace of data accumulation; this is a classic demonstration of how simply piling on more data will not help matters.

*Compressed decision cycles*¹ have also helped to precipitate *decision-making* flaws that have had significant international consequences, such as the USS Vincennes shooting down of an Iranian airliner in 1988. It was determined that the U.S. Navy crew had misidentified the airliner as a F-14 fighter aircraft because, within the brief window available for *decision-*

¹These *compressed decision cycles* reside within the rubric of *Decision Engineering*, which centers upon the notion of improving the capability for quickly arriving at quality decisions, efficiently prioritizing organizational resources amidst resource scarcity in an ever-changing landscape, and effectively evaluating the unanticipated short-term and long-term consequences as well as the risks associated with each and every decision. The ultimate goal of *Decision Engineering* is to produce intelligently designed decisions that are both scalable and extensible, thereby overcoming the “*complexity ceiling*” that is anchored in: (1) *complexity theory*, which treats organizations as collections of dynamic structures and adaptive strategies, and (2) *network science*, whose organizational structures are comprised of a dynamic network of interactions whose relationships are not simply aggregations of individual static actors. Furthermore, this *complexity ceiling* notes the contradistinction between: (1) sophisticated specialization of organizational *decision-making* processes and the complexity of the situations for which those decisions must be made, and (2) *intelligence quotient* (IQ) versus *emotional intelligence quotient* (EQ) and the requisite harmonization for designing an *intelligent decision*.

Within the realm of *Decision Engineering*, real-world issues encompass the finely woven fabric of interconnected data and necessitate a holistic framework. By way of example, energy affects climate, which affects food security, which affects health, which increases susceptibility to counterfeit pharmaceuticals, which affects *cyber-physical supply chain* transparency, which affects assured point-of-need-delivery for Humanitarian Assistance and Disaster Relief (HADR) supplies, etc.

making, they had subconsciously interpreted the data in accordance with a predefined scenario [7].

II. HOW TO DEAL WITH THE ABSENCE OF A DIRECT PROPORTIONALITY BETWEEN DATA QUANTITY AND DECISION APPROPRIATENESS

In essence, when confronted with too much data in a compressed timeframe, we oftentimes make mistakes. We are overwhelmed, and unable to process all the data that are coming at us; we process selectively. Interpretation is constitutive of, and thereby limited to, the individual’s life experiences. According to Carl Jung, while things may be apprehended at the conscious level, the archetypes that inform the interpretation exist only at the unconscious level;² these archetypes are representative of unlearned tendencies, similar to the concept of instincts put forth by Sigmund Freud, to experience things in an individualized fashion, and in most cases, the individual’s desire for *Gestaltian closure* [8] leads to an assignment of a lower-order approximation based upon these inherent biases or archetypes.

One strategy to combat this predilection or tendency and improve *decision-making* [9] is to elongate *Gestaltian closure* for the *decision cycles* within the *decision-making* process and to extend our outlook horizon of the problem so as to permit a more comprehensive assessment of its underlying constituent issues. This could be attempted through self-knowledge and self-discipline: becoming aware of and taking account of our own biases, as well as consciously slowing down the *decision-making* process. It could also be automated. For example, decision-support systems could be programmed to present data and prompt for a response only when enough data have accumulated and achieved a specified threshold and level of data robustness.

A second strategy is to make *decision-making* a *collaborative* process. Bringing multiple partners into analyses of data has several advantages. If the team is properly reconstituted, collaboration can constrain and mitigate biases. Second, collaboration can make the analytical process less overwhelming by allowing the team to “divide and conquer” [10]. Different members of the team can focus their attention upon a manageable amount of data, thereby focusing upon those data types and data sources they are most comfortable and familiar with—ideally overlapping, so there are multiple perspectives on any one data set. Groups with cognitively diverse membership will not only challenge members’ biases and compensate for members’ blind spots, but they will also visualize and communicate the same data in a variety of ways, adding depth to the deliberative process as they collaboratively turn over interpretations and settle on the best solutions.

²The societal impact of archetypes is best exemplified by the specious logic attached to New York City Detective Bureau Chief Thomas Byrnes’1886 rogues’ gallery — a collection of pictures/photographs of suspects/criminals utilized by law enforcement agencies for identification purposes — which was entitled, *Professional Criminals of America*. Byrnes, similarly with Italian criminologist and physician Cesare Lombroso in his 1876 *Theory of Crime*, had reasoned that criminality could be determined simply by examining an individual’s portrait!

Collaboration, then, can help “cross the chasm”³ [11] from *Big Data to Big Insights*, and enable the formulation of *better decisions* (rather than just *faster decisions*) in a landscape of “*high adaptation cycles*”⁴ [12]. But there are some pitfalls. *One pitfall* is the concern to ensure that the team is not subject to shared bias. This can never be guaranteed, but the sometimes obvious bias can be identified and dealt with. An example: In 2008, U.S. Secretary of Defense Gates announced a new project named after Minerva, the goddess of wisdom and war [13]. Minerva is a Pentagon initiative that seeks to involve universities in the “Global War on Terror” or “Overseas Contingency Operations.” However, Minerva has proven to be somewhat controversial. Although numerous scholars support Minerva, a number of academic associations, such as the American Anthropological Association, have raised concerns. Reputable scholars, such as George Mason University’s Dr. Hugh Gusterson, worry that “any attempt to centralize thinking about culture and terrorism under the Pentagon’s roof will inevitably produce an intellectually shrunken outcome...The Pentagon will have the false comfort of believing that it has harnessed the best and the brightest minds, when in fact it will have only received a very limited slice of what the ivory tower has to offer—academics who have no problem taking Pentagon funds. Social scientists call this *selection bias*, and it can lead to dangerous analytical errors” [14]. To offset this bias, the Pentagon has sought to become more inclusive, via the movement that is now known as *Collaborative Big Data*;⁵ this movement was especially evidenced during the 2010 Haiti earthquake and 2011 Japan earthquake and tsunami when coordination tools, such as *Ushahidi*, highlighted effective crowd-sourcing and collaborative efforts [15], in what is known as the *participatory revolution*. Hence, *Ushahidi* achieved what Minerva could not—democratize data and make it accessible to academics and long-tail demographic groups worldwide.

On the other side of the coin, practitioners from the development community are, traditionally, wary of utilizing military terms and concepts within a development context—sometimes out of concern for how they are perceived by the

³Geoffrey Moore described the technology adoption lifecycle in his 1991 book, *Crossing the Chasm: Marketing and Selling High-tech Products to Mainstream Customers*. This seminal work described how most new technologies fail to “cross the chasm” between “early adopters,” who represent the small minority of the total market that are interested in technological novelty, and the “early majority,” who are part of the mainstream market and are interested in gaining practical value from the application of new technologies.

⁴The Rapid Reaction Technology Office (RRTO), a component of the Rapid Fielding Directorate (RFD) within the Assistant Secretary of Defense for Research and Engineering (ASD[R&E]), describes insurgent groups and other oncoming adaptive adversaries as possessing an increasingly shorter “cycle of adaptation.” As a result, the decision-making, prototyping, and testing processes to develop effective countermeasures are increasingly more compressed.

⁵This doctrinal (the fundamental principles guiding the rules of engagement for armed forces) shift is as seminal as the aircraft carrier doctrinal difference, which saved the U.S. during the Pacific War of World War II (the Battle of Midway was a punctuating event in that the differences in carrier doctrine between the U.S. and Japan was a turning point in the war; the Japanese carrier doctrine, which did not well take into account enemy carriers that were within striking distance, resulted in the Japanese fleet being caught with hangar decks that were densely packed with fueled and armed aircraft — not unlike the situation that the Japanese took full advantage of at Pearl Harbor).

partner/host nation, and sometimes for philosophical or other reasons. The reality is, however, that the U.S. military, among others, with its relatively sizable budget, is already performing many functions that traditionally belong to the development community, including Humanitarian Assistance and Disaster Relief (HADR), and can be a powerful ally. On account of a predilection to avoid military tools, doctrines, and contexts, the development community may, inadvertently, overlook certain invaluable opportunities to learn from military experience, and opportunities to “piggyback” on an architectural stack of best-of-breed technologies that have already been painstakingly developed within the defense arena and include technologies—particularly those involving *Big Data/Big Insight* analytical capabilities (*Content Analytics, Entity Resolution, Predictive Analytics*, etc.—collectively, *Layered Analytics*)—that could help the development community to readily meet the grand challenges and high-impact development intent outlined by the USAID Administrator during the 2012 John B. Hurford Memorial Lecture to the Council on Foreign Relations [16]. However, there are signs that this resistance is changing. Secretary of State Clinton has consciously borrowed the notion of “*smart power*” from the Department of Defense (DoD) lexicon, defining “*smart power*” as “the capacity of an actor to combine elements of hard power and soft power in ways that are mutually reinforcing such that the actor’s purposes are advanced effectively and efficiently” [17] and USAID Administrator Shah has often asserted, “*smart power* is perfectly aligned with the discipline of development” [18]. Particularly amidst these “*smart power times*,” there is a pressing and pragmatic need to operationalize the “*3D*” (Diplomacy, Development, Defense) construct [19] authored by former Secretary of State Rice and championed by current Secretary of State Clinton [19]; it is through this “*3D*” lens (now extrapolated to be “*5D*”⁶—Diplomacy, Development, Defense, Disaster, and Data) that many development practitioners are now beginning to wear, as a badge of honor, their breadth and depth of experience within the defense sector. They desire to bring to the development community an orchestrated fusion of field-tested approaches that have proven to be successful in arenas that have, traditionally, been considered to be outside the comfort purview of international development practitioners — concepts, such as the *Aegis System*,⁷ a system that effectively assigns priorities with limited resources: “the minimum volume of resources required” [20] to achieve “meaningful, measurable, and lasting impact” [21]. The *Aegis System* was “the first fully integrated combat system capable of simultaneous warfare against air, surface, subsurface, and strike threats [by] detect[ing] incoming missile or aircraft threats, sort[ing] them by assigning a threat value,

⁶Transformational Diplomacy is a diplomatic initiative pioneered by former U.S. Secretary of State Condoleezza Rice. Rice’s “*3D*” (Defense, Diplomacy, and Development) notion was extrapolated to a 4D concept (to include Disaster) as a result of the 2010 Haiti earthquake. Noting the National Science Foundation’s mandate for data management plans as of January 18, 2011, Prince of Wales Senior Fellow S. Chan augmented the 4D notion into a 5D approach (to include Data).

⁷The Aegis Combat System (ACS) is the U.S. Navy’s most capable surface-launched missile system. Aegis was first conceived in the late 1960’s as the Advanced Surface Missile System (ASMS) to maintain combat effectiveness against threats that were increasingly armed with anti-ship missiles. Aegis has evolved into an advanced command and control system-of-systems, integrating phased-array radars, various short-medium-long range missiles, point defense systems, electronic warfare countermeasures, and Command and Decision (C&D) capabilities.

assign[ing] on-board Standard surface-to-air missiles, and guid[ing] them to their targets.” [22]; the same underlying architecture could enable the development community to prioritize threats, let us say, that would potentially jeopardize a stable and prosperous Afghanistan while also monitoring climate, food security, health, educational outcomes, etc.

This increasing willingness to take advantage of solutions pioneered by the military comes at a crucial time for the international development community. On June 22, 2011, it was announced that “the U.S. goal [was] to be substantially out of Afghanistan by 2014”⁸ [23]. At that time, the U.S. will no longer carry the same influence it has had, over the past decade of operations, in Afghanistan. Instead, influence will be partitioned among the many individual nations, partnered with Afghanistan, and the International Organizations (IOs) of the International Community of Action (ICOA) will be the principal actors for continuing to provide humanitarian assistance and support. In the ensuing absence of a single prominent leader, the ICOA and the Afghan government will need to coordinate their efforts so as to maintain long-term stability and growth. Like an orchestra without a conductor, the post-2014 Afghanistan might necessitate a new level of teamwork among the numerous, varied players on the world stage to adequately succeed.

This brings us to a *second pitfall* in the idea of *Collaborative analysis of Big Data and decision-making*. This is the difficulty of achieving optimal group performance. Successful collaboration is an art and a science, and dysfunctionality is a commonplace pitfall. How can the ICOA, such as that focused upon Afghanistan, work harmoniously and effectively towards a successful outcome?

The orchestral metaphor is particularly apt. Lessons can be learned from the Orpheus Chamber Orchestra [24], a world-class orchestra that operates entirely without a conductor and has frequently served as a business model for collaborative teaming. The question is, *how is it possible to consistently deliver a coherent, perfectly synchronized performance?* Some would say that Orpheus has achieved its world-class quality and reputation for particularly energetic performances, not in spite of its conductor-less structure, but rather because of it. According to cellist Julian Fifer, who founded the Orpheus Chamber Orchestra, “there is a basic commitment to make things work that really comes out in performance ... because everyone is pitching in all they can. And they do that because the performance is theirs. There is no one for us to blame if things go wrong; and there is no one to take the credit from us when things go well” [25].

⁸During the 2010 North Atlantic Treaty Organization (NATO) Heads of State and Government Summit, NATO officials stated that International Security Assistance Force (ISAF) combat operations would transfer full responsibility over security activities to Afghan military and police forces by the end of 2014. However, in order to ensure long-term stability can be maintained afterwards, NATO declared that there would be an enduring partnership between the alliance and Afghanistan and has tasked ISAF to partner with the international development community to facilitate a gradual transition from military-focused defense operations to civilian-focused development activities.

III. CASE STUDY: U.S. POLICY IN AFGHANISTAN/PAKISTAN

For the past decade, the International Security Assistance Force (ISAF) [predominantly, via the U.S.] has acted as a conductor, leading and directing the rebuilding of the Afghan government. However, this will change as of 2014 when each member of the ICOA will be just one player in the symphony that is Afghanistan’s performance journey. Like a musical performance, if all players were perfectly synchronized, the varied efforts will combine into a beautiful harmony. As with all real-time operations, such precise synchronization may be spontaneous, unrehearsed, and therefore relies on an intimate connection among the ICOA players. To achieve this effect, all the players must have the same understanding of the score or “music” (i.e. the vision) as well as of each other, and they must all strive toward the same goal. Hence, the parallel between ICOA players and Orpheus Chamber Orchestra musicians is drawn, and the principles underlying these strong intra-musician relationships within the Orpheus Chamber Orchestra carries high relevance to the coordination required among the ICOA in post-2014 Afghanistan.

Richard Hackman (a Professor of Social and Organizational Psychology at Harvard University as well as a member of both the Board of Trustees of the Orpheus Chamber Orchestra and Intelligence Science Board of the Director of National Intelligence) has long coordinated research on collaborative teamwork as applied to both the Orpheus Chamber Orchestra and the situation faced by ISAF. Hackman’s research has shown that: (1) the most powerful thing a leader can do to foster effective collaboration is to *create conditions* that help members competently manage themselves; (2) the second most powerful thing a leader can do is to *launch the team well* (have a balanced group comprised of complementary Myers-Briggs Types [26], Thomas-Kilmann Conflict Styles [27], Prospector 360-Degree Combinations [28], etc; establish ground rules, trust, common vision/goals, etc.); and (3) the third essential ingredient for the leader centers around real-time teaching, coaching, and *shaping*. Hackman’s research suggests: (1) “condition-creating” (establishing conducive conditions within the instantiated team framework) accounts for about 60% of the variation in how well a team eventually performs [29]; (2) the quality of the “team launch” accounts for another 30% [29]; and (3) the real-time coaching accounts for only about 10% [29].

It would seem, based upon Hackman’s findings, that by maximizing the time between now and 2014, ISAF (as the pre-established conductor until the 2014 Transition in Afghanistan) [30], would be well advised to focus upon the most impactful consideration—“condition creating” or creating the conditions for effectiveness, including effectiveness in gathering, sharing, analyzing, and acting upon data, particularly *sparse data*⁹ [31] (the effectiveness can be enhanced by utilizing an expanded time horizon for the study—a Janusian perspective).

⁹The occurrence of sparse data sets becomes increasingly prevalent as the *decision-making* window on a battlefield becomes increasingly compressed. This increases the risk of meta-cognitive errors from pattern matching/misjudgment of situations, which tend to lead to operational errors.

Compared with many parts of the world, an extraordinary amount of data is available on conditions in Afghanistan. Field survey data, satellite data, electronic communications data—if ever there was a glut of Big Data that needed sorting, it is the trove of data that have been gathered over the past decade by the international community, both military and civilian. In this brave new world of data vertigo, traditional indicators and proxies seem hardly to apply. Whereas in the past, the nutritional status and food consumption of a community was measured by using proxies like open pit/bucket latrine size—measured by field workers, who counted these types of sanitation facilities—today there are electronic medical records, there are anthropometric¹⁰ and kinematic standards¹¹ [32] that can be applied to images from video feeds, and there are satellites that can verify the timely delivery of food aid. The glut of data makes it all the more urgent for the international community to find ways to collaboratively process and interpret data for smart and timely decisions, rather than getting overwhelmed by the torrent of data.

Contrary to the commonly accepted idea that teams should frequently bring in new players/members to gain fresh perspectives, Hackman suggests that “the longer members stay together as an intact group, the better they do. As unreasonable as this may seem, the research evidence is unambiguous. Whether it is a basketball team or a string quartet, teams that stay together longer play together better” [29]. In other words, Hackman’s research shows that the strength of a collaborative relationship grows with time. As quasi-empirical evidence, the Orpheus Chamber Orchestra’s “roster is fixed, and boasts a low turnover rate” [25], thereby supporting Hackman’s suggestion that “excellent teamwork is based on “strong relationships developed over many years” [33]. ISAF’s remaining 30 months in Afghanistan could allow ample time to guide, shape, and strengthen relationships between the ICOA and Afghan government, and these relationships will, hopefully, via “Relationship Managers,” blossom and persist well beyond 2014.

Given the necessity for, and realistic possibility of, Collaborative analysis of Big Data, what are some of the problems the international development community could tackle? Right next door to Afghanistan (and deeply entwined in its prospects for stability and security), Pakistan is another focal interest of the U.S. Agency for International Development (USAID). USAID spends approximately \$1 billion in Pakistan every year [34]. It funds a variety of projects, among other things, to assist in flood recovery, stabilize the Federally Administered Tribal Areas (FATA) along the Afghan border, modernize energy infrastructure, help entrepreneurs build their businesses, improve gender

equity, and train medical workers and teachers as well as to build clinics and schools for them to work in. Despite all this, USAID funding has not comfortably kept pace with the needs of Pakistan’s education sector, which is “among the world’s least effective [since it] devotes less than 2% of GDP to education [while] nearly one-quarter of primary school age children have no formal education of any kind” [35]. Currently, the agency spends that money to: (1) train teachers, (2) refurbish schools, and (3) fund higher education scholarships [21]. These spending goals are metric-driven and thoroughly tactical, but even so, one simple question still exists: what is the best means of allocating funding among these three areas over, let us say, a five-year timescale?

At first blush, one might suggest that the funding for higher education scholarships should be cut, since the direct perceivable benefits accrue to only a few students. Yet, as the USAID document, *Education in Pakistan Working Paper* notes, these scholarships likely increase Pakistan’s economic development and exist “to increase the number of students that enroll and complete courses in primary, secondary, and tertiary educational institutions” [36]. As that development spreads, more of the country will become electrified, bringing the very real possibility of electricity to the approximately 85% of Pakistani schools that currently lack it [36]. But economic development and electricity are useless unless the aforementioned students are both present (in terms of attendance) and healthy, which may currently be polarized goals, since “50 percent of the public schools in rural areas lack clean drinking water, [and] 37 percent have no latrines” [36]. Continuing along the vein of these recitals of fact, the public schools may spread disease, thereby decreasing attendance. While improving the sanitation in these public schools might improve health, even greater health gains might come from increasing caloric intake and food security, or even from decreasing the atmospheric particulates due to dirty energy production. For that matter, economic development might actually (counterintuitively) increase the power of criminal organizations that promulgate counterfeit medications, thereby exacerbating the health situation and, as a direct result (as an *unintended consequence*), impede efforts to keep children in primary school.

This analysis becomes even more complicated when USAID’s purported goal of advancing a robust education program is seen within the context of broader strategic goals. For instance, the *Education in Pakistan Working Paper* asserts that “the lack of access to a high-quality secular education can contribute to violent extremism” [36] and therefore suggests the leveraging of education spending so as to stabilize “areas vulnerable to extremism” [36]. Yet, if the goal were to reduce the recruitment of children to violent extremism, the question remains as to whether education spending alone represents the best allocation of funds, or whether spending should be carefully be divided between education and those programs that directly prevent the pre-malnutrition, which can leave children particularly vulnerable to disingenuous economic incentives of the underworld (e.g. human trafficking).

The aforementioned scenario offers just one exemplar of how the various pillars of development interact, and these interactions will likely take a very different form in Sri Lanka

¹⁰With regards to anthropometric standards, the dataset provided by the U.S. Army Anthropometric Survey combined with the Civilian American and European Surface Anthropometry Resource (CAESAR) Survey provide a blended corpus (military and civilian body types) of data for more optimal best-fit anthropometric models.

¹¹Kinematic standards refer to observational data centered on human motion. These standards have allowed physicians to diagnose medical conditions through the observation of recorded video of a patient’s range of motion (e.g. walking gait). Such diagnostic methods could conceivably be used to remotely detect the average health or nutritional level of the population in a developing nation.

and Haiti than in Pakistan. Expanding our view, we can see that only by understanding how energy policy affects the climate, the environment, and agriculture can we hope to understand food security. Only by understanding how pollution, climate, and food insecurity affect health can USAID's goal of "treating people...not diseases" [37] be achieved. Only by mapping how nutrition intersects with health and the environment, as well as with energy policy, can we begin to understand how to create the conditions necessary (i.e. Hackman's *condition-creating*) to foment the private-sector driven, broad-based economic growth, which USAID seeks to foster. In turn, only by understanding the interactions of all these characteristics of a society can we hope to truly understand which interventions are the most critical, post-disaster, and what policies will effectively and efficiently cradle the transition to *democratic governance*, and what conditions will lead to an *annealed resiliency* called *latent stability* that can pull fragile and conflict-affected states back from the brink of tragedy. All of this requires data—good [*provenance/pedigree*] data. In many cases, we already have the data; in others, we can readily gather it. Given the oceans of data, what we need is a smart strategy for collaboratively mastering it, and not allowing ourselves to merely drift, carried along by the first most plausible interpretation of whichever data set is most readily at hand.

IV. HOW TO ENSURE NETWORK COORDINATION, BALANCE, AND STABILITY

Having performed a preliminary exploration of the possibilities for Big Data in the international development space, we have some specific ideas to offer the international development community. We call our [*Big Compute*] approach, which is a blend of *social complexity science*, *high performance computing science*, *supply chain science*, and *network science*, "*Relationship Science*." Our *Relationship Science* works for a fairly straightforward reason: it embodies the characteristics of how the world *actually* works. To analyze complex real-world relationships, the methodological framework — "*karassian netchain analysis*" (KNA) — is utilized. This framework differs from traditional netchain analysis [38] (network and supply chain analysis) in three critical ways: (1) it adequately considers the network of informal trust relationships among actors, which may be particularly significant for countries where kinship, social, and business relationships overlap pervasively; (2) it expands the observational space to include the interactions among heterogeneous actors within a given "horizontal" layer of a supply chain; and (3) it captures the latent potential for actors within the horizontal layer to deviate significantly from the average behavior of their peers, which may have ensuing dramatic effects. Furthermore, through KNA, it is possible to identify specific *local community structures* within the supply chain, via discernible "shapes" that correspond to specific economic conditions and/or adaptations amidst various pressure sensitivities. It is critical to successfully identify these *memes* or *motifs*, particularly in the context of HADR — one of USAID's 7 Core Development Objectives (a.k.a. "7 Pillars") outlined in USAID's Policy Framework 2011-2015. Of particular interest are the *network shapes* of involved *dualistic actors* (i.e. *insider threats*), for example: (1) individuals who accept development and humanitarian aid under the auspices of representing the local

Government of the Islamic Republic of Afghanistan (GIROA), but are actually Haqqani¹² [39], [40] fighting against ISAF, or (2) groups that accept monies under the rubric of humanitarian aid, but are instead utilizing the funds to purchase weapons in the Tamil Tigers' fight against the Sri Lankan government. More generally, these *network shapes* could form the basis by which planners and policy-makers can more readily quantify and qualify the happenings within their areas of responsibility for maximum situational awareness. These *computational intelligence* work products can well serve to inform the desired "*common operating picture*" for the full spectrum of USAID's 7 Pillars, thereby segueing to more strategic and impactful *decision-making*.

As applied to international development, our approach (*Collaborative analysis of Big Data*, and the application of *Relationship Science*) can have a tremendous triadic payoff. In the U.S. State Department's FY 2013 Congressional Budget Justification, Secretary of State Clinton identified three goals as priorities: (1) to "secure American interests, strengthen our friends and allies, [and] forge new partnerships;" (2) to engage in the world "shoulder-to-shoulder with our troops—on the frontlines of our efforts to keep America safe;" and (3) to take on "global challenges to human and economic security, including hunger, disease, and the destabilizing effects of climate change." Our approach addresses all three of these priorities—by fostering collaboration among international partners (#1), by fostering open engagement with the military where the military has value-added technologies and lessons-learned to offer (#2), and by tackling the core concerns of the development and aid community (#3).¹³

While the social and physical sciences have traditionally tackled problems by breaking them into constituent parts and simplifying interactions between them, it is now clear within the

¹²The Haqqani Network (HQN) is thought to serve as a proxy of the Pakistan Inter-Services Intelligence (ISI) Directorate and operates from both neighboring Pakistan as well as Afghanistan. Its activities in Afghanistan consist of a complex web of interdependencies between Haqqani foreign military commanders based in North Waziristan, Pakistan, and a vast network of Afghan villages containing an undercover "*fifth column*" of local Afghan government officials, who weave sympathy for the HQN by the provision of essential goods and services for villagers. This warped Haqqani brand of "Bizzaro civil affairs" has built up a lasting Pashtun arc of alliances that have served both as a source of logistical/intelligence support for HQN forces fighting the ISAF as well as an Afghan *civil society* warlord power base that serves as a strategic ally for Pakistan against India. Indeed, the HQN insidious infiltration of the fabric of Afghan *civil society* has resulted in many unseen HQN insidious insurgents gaining the trust of their ISAF and Afghan National Security Force (ANSF) adversaries under the auspices of village leaders.

¹³If applying Decision Engineering, it seems prudent not to center efforts around a specific challenge to the developing world itself, but rather around a specific challenge to those seeking to *understand* it and *shape* its course: Development agencies around the world have identified the importance to development work of the myriad interactions between the different aspects of development, and have noted, as an exemplar, that "achieving a lasting impact in reducing child mortality requires targeted and coordinated efforts in health, education, sanitation, infrastructure, and governance." [42] One would respectfully add that the greater challenge lies in the *shifting interactions* between those many dimensions, signifying that they must be not only approached, but also *understood simultaneously*, from multiple angles. For example, sending pharmaceuticals to a poor region might indeed stabilize a nascent democratic government's legitimacy, but criminal organizations operating in the region might leverage that aid to fuel a counterfeit-drug business, and subsequently turn the resulting profits toward destabilizing factors, such as the *trafficking/exploitation* of children or extreme violence. Meanwhile, a climatic stress, such as drought, might cause migration and subsequent conflict, which the timely delivery of food-aid might have prevented.

development community that for the *science of development*, the emergent patterns that beget predictions will appear only when problems are considered in their full complexity and local context. This will illuminate *cyber-physical supply chains*, and *social influence networks*, as well as ecological and climatic systems (within the multidimensional space of development) that must be: (1) orchestrated to achieve *latent stability*, (2) amalgamated to serve as the backbone of *latent stability*, and (3) leveraged to secure pathways to *latent stability*.

The concept of “*islands of stability*” in *network science* is exemplified by the “*sandpile effect*” (more formally, the Bak-Tang-Wiesenfeld sandpile model of non-equilibrium systems [41]) in which sand is dropped, one grain at a time, onto the same spot on a flat surface, until the addition of one more grain of sand causes an avalanche to slide down the slopes of the growing sandpile. In 1987, physicists Per Bak, Chao Tang, and Kurt Wiesenfeld investigated the “*sandpile effect*” by using a computer to color the sandpile according to steepness—the steepest regions of the pile were colored in red, and the flattest, green; they discovered that a single grain of sand falling onto a red region would instigate an avalanche, which not only caused certain green regions to become red, but also compounded into a cascading series of avalanches that grew in size and intensity as it disturbed other red regions (i.e. cascading failure). Restated in terms of *network science*, instability (e.g. an ill-intentioned *flash mob*,¹⁴ which is difficult to predict—even by robust *sentiment analysis*¹⁵) can spread throughout the entire network, via islands of potentially unstable nodes; these small sets of nodes with the power to influence the entire network are known as *Positive Influence Dominating Sets (PIDS)*. Just as a sandpile avalanche can create instability in previously stable areas, real-world phenomena, ranging from political convictions to diseases to computer malware, to name a few, can originate at just a few nodes (an occurrence of PIDS) and eventually permeate an entire society. Identifying PIDS in a given network requires a detailed knowledge and sophisticated analysis of the network so as to uncover the harbingers of instability and “*perfect storm*” crises lurking within a network, and, on the positive side, to identify opportunities to infuse *latent stability* throughout the network by cultivating and/or influencing PIDS.

These concepts are being utilized by ISAF-Afghanistan, where this new form of *Relationship Science* and management aims to create a new paradigm for coordination and collaboration with an increased number of large donors and a proliferation of smaller donors, as well as with developing country governments. One goal is to understand fundamental patterns and constraints that arise from those interactions, based upon the preliminary hypothesis that successful, sustainable coordination arises most readily out of the PIDS surrounding mid-level contacts, rather than from *sparse networks* of high-level contacts (hence, *brittleness*); the reason for this is that the mid-level contacts comprise bureaucratic *latent stability*. Another goal is to create a

holistic, cumulative-context picture of those countries, then utilize that picture as the basis for statistical models that integrate expertise in development, social science, and machine learning, which will identify patterns and use them to predict the outcomes of development interventions, as well as identify opportunities to create *latent stability*, and the indicators of lurking instability that are the harbingers of crisis. To effectuate this goal, one must iteratively focus on key areas of development (food, health, climate, energy and economics, and supply chains, etc.) to integrate existing datasets, expertise, models, and best-practices from those fields into our thinking and our computational architecture, while continually integrating those discipline-specific perspectives into an overarching model that also captures how *democratic governance* and the advancing *science of development* combine to create *latent stability*.

V. CONCLUSION

In summary, the art and science of *Decision Engineering* necessitates a sophisticated approach to analysis of *Big Data*. The inability of any one individual to digest more than a limited amount of data can be addressed in part by attempting to elongate *Gestaltian closure* so that maximum situational awareness is achieved by the individual, and in part by making data analysis a *collaborative* effort. We illustrate the pitfalls and possibilities of Collaborative *analysis* of Big Data with examples from the field of international development and aid. From our research in this area, we conclude that the depth and breadth, which *Robust Decision Engineering* requires, can best be achieved by employing an approach related to *Network Science* that we call *Relationship Science*. In *Relationship Science*, a methodological framework, *karassian netchain analysis (KNA)*, is utilized to ascertain *islands of stability* or *positive influence dominating sets (PIDS)*, so that a form of *annealed resiliency* or *latent stability* is achieved, thereby mitigating against *unintended consequences*, elements of instability, and “*perfect storm*” crises lurking within the network.

ACKNOWLEDGMENTS

S. Chan and W. D. Rhodes thank Dr. David McQueeney (IBM Vice President, IBM Research), Bob Griffin (IBM Vice President and CEO of i2, an IBM Company), Professor Marc Clement (Executive Chairman of Swansea University’s Institute of Life Sciences), Professor Thomas Levenson (Director of the MIT Graduate Program in Science Writing), Laura Sampath (Director of the MIT International Development Initiative), Jane Obbagy (Vice President of the Cadmus Group, Inc.), Weston Fisher (Principal of the Cadmus Group, Inc.), Professor J. Richard Hackman (Department of Psychology, Harvard University), and Jessica Lewis (Research Director of the Institute for the Study of War [ISW]) for their ongoing invaluable support. Special thanks goes to our inspiring force and co-author of the report, which spawned this series of publications, Colonel/Dr. Michael Meese (Professor and Head of the Department of Social Sciences at West Point); S. Chan is deeply grateful and indebted to Dr. Meese. Additional special thanks goes to Major General Sean MacFarland (Deputy Chief of Staff for Operations at International Security Assistance Force [ISAF]), Major General Michael Shields (Director of Future Operations [FuOPS] at

¹⁴A flash mob refers to a group of people, who suddenly convene in a location, perform a seemingly random act for a brief period of time, and then disperse quickly.

¹⁵Sentiment analysis or opinion mining refers to the utilization of technologies (i.e. natural language processing, computational linguistics, content-based information retrieval, latent semantic analysis, etc.) to ascertain the attitude of a speaker or writer of a tweet, social media posting, etc.

ISAF Joint Command [IJC]), Colonel James Brown (Commanding the 95th Civil Affairs Brigade), Dr. Neal Rappaport (FuOPS Policy Advisor at IJC), Dr. Kathleen Kiernan (Founder and CEO of Kiernan Group Holdings), and our Advisory Board Members for the Prince of Wales Fellowship at MIT and namesakes of our overarching research theme (Semantic Surge 3.0): Dr. Fred Kagan (Director of AEI's Critical Threats Project) and Dr. Kimberly Kagan (President and Founder of the ISW).

REFERENCES

- [1] C. Matthews, "High Frequency Trading: Wall Street's Domsday Machine?" Time Business, Time, 8 August 2012.
- [2] P. Kumar, M. Goldstein, F. Graves, and L. Borucki, "Trading at the Speed of Light: The Impact of High-Frequency Trading on Market Performance, Regulatory Oversight, and Securities Litigation," Finance, The Brattle Group, Washington, D.C., issue 2, 2011.
- [3] "Notice to Members: SEC Approves Amendments to Rule 2320(a) Regarding Best Execution and New Interpretive Material 2320," National Association of Securities Dealers (NASD), October 2006.
- [4] J. Anderson, A. Murray-Jones, S.G. Sims, P. Brandt, "FSA Publishes Guidance on Market Maker Order Flow Payments to Brokers," Skadden, Arps, Slate, Meagher and Flom LLP, 23 May 2012.
- [5] "Notice to Dealers that Use the Services of Broker's Brokers," Rule G-43, Municipal Securities Rulemaking Board (MSRB), Effective on 22 December 2012.
- [6] D.B. Sullivan, "Speed of Light From Direct Frequency and Wavelength Measurements," National Institute of Standards and Technology (NIST), January 2001.
- [7] LTJG B. Ozkan, N. Rowe, LT S. Calfee, J. Hiles, "Three Simulation Models of Naval Air Defense," Institute for Modeling, Virtual Environments, and Simulation (MOVES) and Computer Science Department, U.S. Naval Postgraduate School, Monterey, CA, Proceedings of the 2005 International Command and Control Research and Technology Symposium, McLean, VA, 2005.
- [8] N. Jordan, "Decision-Making Under Uncertainty and Problem Solving: A Gestalt Theoretical Viewpoint," Logistics Department, The RAND Corporation, P-2156, 1 December 1960.
- [9] D. Ariely, D. Zakay, "A timely account of the role of duration in decision making," Acta Psychologica, 108, 2001.
- [10] S.R. Waterhouse, "Divide and Conquer - Pattern Recognition using Mixtures of Experts," Doctoral Dissertation, Jesus College, University of Cambridge, 14 February 1997.
- [11] G.A. Moore, "Crossing the Chasm: Marketing and Selling High-tech Products to Mainstream Customers," HarperCollins, 1991.
- [12] "Experimentation and Rapid Prototyping in Support of Counterterrorism," Committee on Experimentation and Rapid Prototyping in Support of Counterterrorism; National Research Council, 2009.
- [13] R. Gates, "Speech to the Association of American Universities," Office of the Assistant Secretary of Defense (Public Affairs), U.S. Department of Defense, 14 April 2008.
- [14] H. Gusterson, "When Professors Go to War," Foreign Policy, The Foreign Policy Group, 21 July 2008.
- [15] S. Chan, B. Ranalli, C. Kuo, K. Intrator, C. Atencio, A. Miao, "Humans as sensors: Fusion of participatory mechanisms and computational innovations to monitor climate change and its consequences," African Centre for the Constructive Resolution of Disputes (ACCORD), South Africa, 2012 (To Be Published)
- [16] R. Shah, "Transforming Foreign Aid: A Conversation with Rajiv Shah," Council on Foreign Relations, 7 March 2012.
- [17] E. Wilson III, "Hard Power, Soft Power, Smart Power," The Annals of the American Academy of Political and Social Science, 616, March 2008, p.110-124.
- [18] R. Shah, "Speech by USAID Administrator Dr. Rajiv Shah: Approach to High-Impact Development," Council on Foreign Relations, 5 May 2010.
- [19] S. Epstein, "Foreign Aid Reform, National Strategy, and the Quadrennial Review," Congressional Research Service, 15 February 2011.
- [20] "ADS Chapter 200 - Introduction to Programming Policy," Policy, Planning, and Learning (PPL), U.S. Agency for International Development (USAID), 10 February 2012.
- [21] "USAID Policy Framework 2011-2015," U.S. Agency for International Development, U.S. Department of State, 2011.
- [22] W. Bridger, and M. Ruiz, "Total Ownership Cost Reduction Case Study: AEGIS Radar Phase Shifters," MBA Professional Report, Naval Postgraduate School, Monterey, CA, December 2006.
- [23] CJ Radin, "US withdrawal from Afghanistan: the plan for 2012, 2013, and 2014," The Long War Journal, The Foundation for Defense of Democracies, 18 March 2012.
- [24] H. Seifter, "The Conductor-less Orchestra," Leader to Leader, no. 21, Summer 2001.
- [25] A. Kozin, "An Orchestra that Needs No Conductor," The New York Times, 12 December 1982.
- [26] R. Lewis, L. Spears, and B. Lafferty, "Myers-Briggs and Servant-Leadership: The Servant-Leader and Personality Type," The Spears Center for Servant-Leadership and Ralph Lewis Associates, August 2010.
- [27] E. Hallmark, "Thomas Kilmann Conflict Styles," National Policy Consensus Center, Portland State University, 7 May 2012.
- [28] "Prospector 360-Degree Assessment," Center for Creative Leadership, 2011.
- [29] J.R. Hackman, "Six Common Misperceptions about Teamwork," HBR Blog Network, Harvard Business Review, 7 June 2011.
- [30] "Afghanistan: The Timetable for Security Transition," International Affairs and Defense Section, House of Commons, United Kingdom, 9 July 2012
- [31] C. Ntuen, "The Knowledge Structure of the Commander in Asymmetric Battlefield: The Six Sights and Sensemaking Process," Army Center for Human-Centric Command & Control Decision Making, 2006.
- [32] S.A. Sisto, "An Overview of the Value of Information Resulting from Instrumented Gait Analysis for the Physical Therapist," within "RRDS Gait Analysis in the Science of Rehabilitation," Edited by J.A. DeLisa, Department of Veterans Affairs, Veterans Health Administration, 1998.
- [33] J. Scott, "Emerging Patterns from the Dynamic Capabilities of Internet Intermediaries," University of Texas at Austin, March 2000.
- [34] "U.S. Senators Vote to Tie Pakistan Aid to Supply Routes," Reuters, 22 May 2012.
- [35] S. Epstein and K.A. Kronstadt, "Pakistan: U.S. Foreign Assistance," Congressional Research Service, 10 April 2012.
- [36] "Education in Pakistan Working Paper," USAID, April 2011.
- [37] "The Summit to Save Lives," George W. Bush Presidential Center, Washington, D.C., 13-14 September 2011.
- [38] E. Talamini, G. Murad Velloso Ferreira, "Merging network and social network: Introducing the 'social net chain' concept as an analytical framework in the agribusiness sector," African Journal of Business Management, vol.4(13), pp.2981-2993, 18 October 2010.
- [39] M. Mazzetti, S. Shane, and A. Rubin, "Brutal Haqqani Crime Clan Bedevils U.S. in Afghanistan," The New York Times, 24 September 2011.
- [40] M. Ryan, and S. Cornwell, "U.S. says Pakistan's ISI supported Kabul embassy attack," Reuters, 22 September 2011.
- [41] P. Bak, C. Tang, and K. Wiesenfeld, "Self-organized criticality," Physical Review, vol. 38, no. 1, p.364-375, 1 July 1988.
- [42] "ADS Chapter 200 - Introduction to Programming Policy," Policy, Planning, and Learning (PPL), U.S. Agency for International Development (USAID), 10 February 2012.