

# **1 Sociocultural approaches to understand human interaction: A discussion of new theoretical frameworks, issues, and modern communication technology<sup>1</sup>**

Sara Beth Elson & Alison Dingwall, The MITRE Corporation  
Mansoor Moaddel<sup>2</sup>, University of Maryland, College Park

## **1. Introduction to Sociocultural Understanding**

The sons of Adam are limbs of each other,  
Having been created of one essence.  
When the calamity of time affects one limb  
The other limbs cannot remain at rest.  
If you have no sympathy for the troubles of others,  
You are unworthy to be called by the name of a Human.  
– Sa’adi Shirazi, 13<sup>th</sup> century

No man is an island entire of itself; every man is a piece of the continent, a part of the main;  
if a clod be washed away by the sea, Europe is the less, as well as if a promontory were,  
as well as any manner of thy friends or of thine own were; any man's death diminishes me,  
because I am involved in mankind. And therefore never send to know for whom the bell tolls;  
it tolls for thee.

–John Donne, 1624

That these luminaries from diverse cultures and times arrived at the same conclusion regarding the psychic unity of mankind may lend credence to the premise that all cultures strive for social harmony, compassion, and peace. This eloquently articulated, idealistic view of humanity, however, cannot be easily reconciled with the persistent misunderstandings arising from cultural diversity and the reality that conflicts not only occur among cultures but that cultures themselves contain conflicting elements. The clashes within and between cultures run contrary to the concept of unitary, caring, and compassionate humankind that Sa’adi and Donne celebrated.

---

<sup>1</sup> Approved for Public Release; Distribution Unlimited. Case Number 14-2487

This work was supported by Department of Defense Contract W15P7T-13-C-F601

Copyright © 2014 The MITRE Corporation.

<sup>2</sup> Many thanks to Julie de Jong for her comments on this chapter.

While sociocultural theories of human interaction abound, this chapter focuses on two basic approaches for gaining sociocultural understanding. The first approach uses a theoretical framework to foster medium-term, a period between 10 and 15 years, sociocultural understanding, while the second employs the new technologies of mass communication to enable rapid understanding of how information, ideas, and propaganda flow through various outlets as well as the manner in which the information flow shapes political actions, including extremism.

The theoretical framework underlying the first approach illustrates how sociopolitical and cultural issues serve as indicators of trends in values and ideological change in the medium-term. In describing this framework, we discuss two research methodologies that support it: content analysis and survey research. Content analysis applies to data collected in real time from social media outlets that reveals people's perceptions of issues and events. It provides quick information for understanding changes in the sentiments, emotions, and attitudes of the subject population in the short run and how such changes are linked to different forms of political action. With regard to survey research, we focus on the challenges of conducting cross-cultural surveys and on innovations in survey techniques that are not yet in widespread use.

Section three in this chapter, *Technologies to Promote Near-Term Sociocultural Understanding*, focuses on innovative tools that can help researchers gain understanding of sociocultural issues in near-real time. Many of the tools described are currently in development or are used only by select organizations, but should become more broadly available in the near future. We then discuss gaps in the technologies that research should address in the future. We conclude by describing the actions necessary to bring state-of-the-art technologies and methods into operational use.

We must emphasize the importance of understanding culture as an overall framework for viewing the world. The interaction between society and an individual can influence the development of self-schemas through changing expectations regarding self-awareness (i.e., individualist vs. collectivist societies), social judgments (i.e., adhering to group norms), and social behaviors. But just as our own sociocultural perspectives affect our own thoughts, feelings, and behaviors, it can also color how we view other people or groups. Therefore, it is important to recognize how social context and sociocultural perspectives may influence individual or group behavior. In the section of the book covering detection of sociocultural signatures, a chapter on training presents additional discussion of culture as an overall framework.

## **2. Theoretical Framework for Medium-Term Sociocultural Understanding**

### **2.1. Sociopolitical and Cultural Issues as Indicators of Trends in Values**

"Social change" is an umbrella term used to capture such observable aspects of social life as changes in (a) the principles of social organization (e.g., economics and politics, the form of government, or the relationship between religion and politics); (b) religious beliefs and institutions (e.g., the rise of different religious movements, sects, and cults); (c) values, rituals, identity, and life-style (significant changes in people's religiosity, consumption habits, and style of dress); and (d) the arts and literature. The actual process of change, however, is complex, and its probable forms

and content are hard to detect. In some cases, change appears chaotic, exhibiting few discernible patterns; only multifaceted group conflicts, sociopolitical upheavals, violence, and contradictory events are evident. In other cases, by contrast, change lies hidden beneath the veneer of political stability and smooth economic transition. A sudden emergence of revolutionary movements leading to the breakdown of the existing political order and changes in social relations can interrupt a stable and quiet period of economic development. In the first instance, we know that changes are transpiring, but we do not know the direction of these changes and what cultural pattern will predominate. In the second, we expect the emergent cultural pattern to correspond to the observable indicators of economic development and political formation and are surprised when a strong countercurrent surfaces.

We suggest that certain indicators in public opinion and attitudes, as well as in the expression of intellectual leaders and opinion makers, offer reliable measures of the extent of support for existing sociopolitical and cultural orders. These indicators may change in finite ways, and the direction of changes in these indicators provides clues about the possible content and form of social change in a period encompassing less than a generation—the medium run. We also suggest that historically significant issues function as indicators and barometers of change and that the manner in which issues are resolved signifies the direction of changes in some or all aspects of culture. In some cases, the resolutions entail revolutionary change, causing the existing societal model to be abandoned or modified in favor of another. For example, in the years preceding the Iranian Constitutional Revolution (1905–11), the arbitrary power of the monarch became a political issue among intellectual leaders and activists. Cultural change occurred when the revolution substituted a constitutional modality of politics for monarchical absolutism. In the years preceding the Iranian Revolution of 1979, on the other hand, the pro-Western monarchy-centered secular modality, rather than the monarch's authoritarianism, was the predominant point of contention. The revolution resolved it by substituting clerical absolutism for monarchical dictatorship (Moaddel, 1993; 2005).

In other cases, issue resolution takes the form of a rather gradual shift toward, for example, gender equality or gender hierarchy, democracy or authoritarianism, liberal values or religious government, and national/ethnic or religious identity. These changes may then shape the probability distribution of the success of different political groups, movements, or political parties.

The model proposed here parameterizes issue resolution in terms of the variable features in the dynamic context of ideological debates and religious disputes among both intellectual leaders and the public at large, on the one hand, and the role of the state as well as politically powerful groups in this context, on the other. While culture is conceptualized “as a ‘tool kit’ of symbols, stories, rituals, and world-views” (Swidler, 1986, p. 273), we argue that it hardly constitutes a consistent set of values, norms, rituals, symbols, memories, and institutional principles. Rather, it contains incongruent elements that often clash. In fact, fundamental cultural values derive their meaning from opposition to other cultural values that may exist in the same culture. This means that multiple cultural schemas exist simultaneously in a social environment and can convey contradictory messages about values, goals, and appropriate methods. Given these multiple, potentially conflicting schemas, one cannot claim that any particular culture dictates certain types

of decision making and behavior. Instead, as discussed by Swidler (1986; 2001), cultural models generally provide goals and strategies from which people can choose in making specific decisions. Thus, individuals can follow one set of cultural elements in a particular circumstance and another set of elements in a different circumstance. This is true, because “people, manifestly, care about more than one thing—indeed, are, simultaneously and sincerely attached to values that clash”(Sniderman, 1993, p. 224).

The existence of diverse cultural schemata that shape different, if not conflicting, strategies of action creates the potential for perpetual conflicts in society. This diversity complicates intracultural and intercultural understanding. Nonetheless, sociopolitical stability is ensured when a subset of cultural elements is formalized into a dominant discursive framework, becomes the guiding principles of social interaction, and gains the appearance of permanence and taken-for-granted features of social order. Certain social institutions and watchdogs maintain and protect the dominant discourse, which is ultimately sanctioned by state power.

## **2.2. Cultural Debates, Religious Disputations, Political Conflict, and Issue Resolution**

Cultural change may occur when individuals or groups challenge some or all of the elements of the dominant discourse, these elements are turned into issues, intellectual leaders offer resolutions to the issues, and a significant section of the population, members of powerful groups or social classes, or the ruling elite accept those resolutions. Thus, an important aspect of understanding cultural change entails understanding the actual process involved in resolving issues.

The emergence, debate over, and resolution of issues constitute a non-anonymous and observable process. The process begins when producers of sociopolitical ideas turn such aspects of daily life as people’s social identity, the status of the economy and distribution of economic resources, poverty, social and cultural inequality, arbitrary rule, national security, the perceived status of a nation or community vis-à-vis others, or people’s spiritual needs into topics of debate. What requires explanation is how intellectual leaders offer resolutions to these issues. How, for example, do they resolve issues related to identity and economic difficulties? Definition of identity may be resolved in terms of religion, nationality, ethnicity, or language. Similarly, societies may address economic difficulties in socialist terms at one extreme, or, at the other, by giving tax incentives to the rich in order to promote investment in the country.

The actual resolution of issues may, however, derive directly neither from the nature of these issues nor from the social positions of the individuals involved in the cultural debates. To be sure, social position constrains expression. An intellectual leader from a humble background may be more sympathetic to a socialistic idea than to the idea of competitive capitalism. Nonetheless, that same person living under Soviet totalitarianism might turn into an enthusiast of the free market capitalism. To give another example, people born in Egypt may have only three possibilities for defining their primary identity: Egyptian, Arab, or Muslim/Christian. Nonetheless, it is crucially important to understand how under certain historical circumstances Egyptians define themselves primarily in national territorial terms (i.e., as Egyptians), but under other circumstances in terms of language (i.e., as Arabs) or religion (i.e., as Muslims/Christians).

### **2.3. Ideological Target as a Key Factor in Issues Resolution**

We argue that societies resolve issues within the dynamic context of intellectual debates, religious disputes, and political conflict as each side of the debate structures the sociopolitical expression of the other side. Each side constitutes the target of ideological production for the other side. In a national context, where the key actors are the regime and its opponents, political activists define their collective identity, specify cultural framing, and resolve sociopolitical issues in oppositional relation to the ideology and identity of the state. The latter may be national or foreign, secular or religious. This oppositional context determines whether people define their collective identity in terms of the nation, language, ethnicity, or religion. For example, territorial nationalism in Algeria and Egypt emerged in an oppositional relation to foreign occupation (the French and British, respectively), and anti-clerical secularism and constitutionalism in Iran in the late nineteenth and early twentieth centuries emerged in oppositional relation to *ulama* religious obstructionism and monarchical absolutism (Moaddel, 2005). Some discourses, such as Islamic modernism, fundamentalism, and various forms of secularism—liberal nationalism or authoritarian nationalism—have formed comprehensive cultural movements. Other discourses, such as Arabism and pan-Arab nationalism, have focused on language as the organizing principle of cultural movements.

### **2.4. Methodological Tools for Collecting Cultural or Attitudinal Data**

In collecting sociocultural data, researchers can use various methodological tools, including comparative-historical investigations, participant and non-participant observation in the field, experiments, and discourse analysis. Here, we focus on two basic tools that have proven useful in collecting cultural or attitudinal data in order to understand and predict change. One centers on analyzing the content of the expressions of intellectual leaders, opinion makers, and bloggers. An analysis of these expressions through time generates useful information for reconstructing intellectual trends in a society and predicting the form of future political and cultural movements. For example, the discourse of pan-Arab nationalism that inspired the military coups in Egypt, Iraq, Libya, and Syria in the 1950s through the 1960s originated in the 1920s. In the same way, the discourse of radical Islamism most prevalent today was formulated in the 1950s and 1960s.

The other method applies survey research to assess the attitudes and value orientations of the ordinary public. Surveys are effective ways to generate scientific knowledge about the role of culture in human lives and the causes and consequences of mass belief systems. Longitudinal and panel survey data provide useful information for understanding trends in values in the population of interest. The following subsections outline key topics related to conducting successful surveys.

#### **2.4.1 Challenges in cross-cultural survey research**

The quality of survey data is a function of the extent to which the researcher formulates all three components of survey research correctly: the questionnaire, sampling, and interviews. These criteria are particularly important in the context of cross-cultural survey research. We describe below several methods used to facilitate the conduct of cross-cultural surveys in light of the challenges that may confound the findings and undermine the validity of scientific generalization.

*Questionnaire Development.* The items in the questionnaire must be clearly linked to, and constitute valid and reliable measures of, the theoretical constructs that drive empirical research. To this end, researchers must consider several aspects of questionnaire development, including translation issues, cultural sensitivity, and item organization and context within the instrument itself.

When questionnaires originally written in English are used to measure attitudes and value orientations across multiple non-English speaking populations, researchers must ensure that the translation does not lose or distort meanings. One standard procedure to reduce the influence of linguistic differences entails (a) translation of the questionnaire from English to the vernacular by a competent translator, (b) back-translation of the questionnaire from the vernacular to English by someone who has not seen the original English version, and (c) a systematic comparison between the two English versions. Another approach, which could either complement or substitute the standard procedure to reduce cost, is to poll some non-English respondents regarding their understanding of the questions during the pilot testing of the questionnaire. If responses are consistent with the intention of the question, then the question can be considered a valid measure of the construct.

Scholarly flexibility and non-ethnocentrism are also important considerations in questionnaire design. For example, in some cases the version of the questionnaire in the vernacular, although different from the English version, more effectively taps into the respondents' sentiment and intentions. In such a case, the researcher must modify the English version to conform to the vernacular version. For example, in a recent cross-cultural survey of populations from relatively diverse countries such as Egypt, Iraq, Lebanon, and Saudi Arabia, the initial questionnaire included the question: "Given the opportunity, how often do you think that men will try to take sexual advantage of women: always, sometimes, rarely, or never?" Both the Egyptians and Saudis argued that the wording in Arabic referring to sexuality would be too sensitive in their home countries, which led to the revised wording: "Given the opportunity, how often do you think that men will harass women?" The revised question uses less inflammatory language while still accessing the original intention of the question. Additionally, there may be Western biases in the types of values addressed in survey questionnaires. For example, questions on religious tolerance or gender relations may lead to a negative image of Muslim-majority countries, as these countries appear less tolerant of religions other than their own or more supportive of gender inequality and hierarchization.

Studies have shown that how questions are worded and framed, including the choice and order of response categories, can have significant effects on the responses (McColl, et al., 2005). Furthermore, while non-locatable participants and refusals may affect data quality (Braver & Bay, 1992), other kinds of self-selection biases that result from the respondents' cultural outlooks may be more subtle and their effects harder to assess. For example, the connection between watching satellite TV and religious attitudes may become complicated if people with conservative outlooks decided against installing satellite TV in their homes. Finally, over-reporting socially desirable and under-reporting undesirable behaviors/attitudes is another serious problem that may have considerable impact on the quality of the data (Cahalan, 1968; Gordon, 1987; Hyman, 1944; Parry

& Crossely, 1950). In Muslim-majority countries, for example, the social desirability bias emerges most clearly in such questions as those that solicit respondents' attitudes toward homosexuality and prostitution. While anecdotal evidence indicates the presence and even tolerance of (male) homosexuality and prostitution in these countries, at least ninety percent of respondents in surveys of the public from these countries said that such acts are never justifiable, which would probably indicate social undesirability bias. Furthermore, asking people in these countries whether they had ever practiced homosexuality would almost certainly anger the respondents, terminating the interview. Questionnaire design must take into account contextual effects and desirability biases such as those reported in the examples above in order to produce reliable and valid data.

*Sampling.* Samples of country populations must be nationally representative, with all respondents having a known probability of selection. In many non-Western societies, the formulation of an effective sampling frame presents a serious challenge, particularly in the absence of recent census data. Researchers must assess how any changes in the sampling procedures as a result of unforeseen factors might affect the representativeness of the final sample used in the study. These may include where the non-availability or inaccessibility of detailed census data requires using alternative spatial sampling or using the random-walk procedure, or where the outbreak of a war or political violence may make certain areas inaccessible to the interviewers or causes mass migration. When designing samples for use in cross-cultural survey research, researchers must also consider comparability of the samples across countries and across time within countries.

*Interviewers.* Interviewers' skill level, their understanding of the questions, and honesty will make or break the validity of the entire process of data collection. Interviewers must be well trained and well paid, and the field supervisor must review their work. In the context of cross-cultural survey research, comparable interviewer training is also crucial.

## **2.5. Innovation in Survey Methods: Understanding the Causes of Popular Attitudes**

Although surveys have made it possible to gain sociocultural understanding, traditional surveys have been limited in that they do not allow analysts to understand the causes of attitudes that respondents hold. Only experimental methods can pinpoint the drivers of human behavior or the reasons why people hold the attitudes they hold.

Numerous texts, such as Shadish, Cook, and Campbell (2002), contain in-depth discussions of experimental methods. In essence, an experiment is a study in which the researcher deliberately introduces (or "causes") an intervention in order to observe its effects. As an example, a researcher might have a theory regarding the aspect of a message that makes it persuasive in a population of interest—for the purposes of this discussion, some characteristic of the person delivering the message, such as his/her age. To test this theory, the researcher might conduct a randomized experiment. For this experiment, the researcher must first recruit a sample of participants who, ideally, represent the population of interest (e.g., intelligence analysts). The researcher then might use a coin toss or random number generator to assign each participant to one of two conditions. In the first condition, participants view a message ascribed to an older person, while in the second condition participants view exactly the same message, but are told that a younger person delivered

it. By randomly assigning participants to view either the message delivered by an older person or a younger one, the researcher can ensure that both groups are reasonably similar to each other. Therefore, if one group finds the message more persuasive than the other group does, this difference is probably due to the factor being tested (i.e., age of the source delivering the message) rather than to any pre-existing differences between the groups. In this way, researchers can establish a causal link between some aspect of a message (e.g., age of the source) and the persuasive power of that message.

Although experiments have traditionally been conducted in academic labs, technology now enables researchers to embed experimental methods within surveys, thus combining the strengths of each method. Namely, it is now possible to explore causal relationships in human behavior via a survey, enabling researchers to “work at scale” by examining a large and representative sample of an entire population. A population-based experiment uses survey sampling methods to obtain a set of participants representing a target population of interest for a particular theory—whether that population is a country, a state, an ethnic group, or some other group. In such experiments, the researcher randomly assigns participants to conditions, and inserts interventions as in any other experiment. New technologies, such as Computer-Assisted Telephone Interviewing (CATI) and the internet itself, have made it easier to randomly assign participants to conditions without asking them to come to a particular place for initial screening. For this reason, researchers wanting to discover a causal relationship that generalizes beyond a narrow pool of participants are well advised to use population-based survey experiments.

Although these experiments are administered to representative samples of the target population of interest, they need not target nationally representative population samples, and often have not done so. The population of interest might be members of a particular ethnic group, parents of children under the age of 18, people who watch television news, or some other subset of the national population. This ability to tailor the sample constitutes the great advantage of population-based survey experiments: researchers can test theories on samples representative of the populations to which the theories are said to apply.

### **3. Technologies to Promote Near-Term Sociocultural Understanding**

Up to this point, we have discussed a theoretical framework and methodological approaches for gaining sociocultural understanding that have long traditions in the behavioral sciences—and are described in commensurately broad research literatures. However, with the advent of new forms of communication via the internet (e.g., social media), researchers have developed new technologies for collecting the voluminous data generated. In particular, social media outlets such as Twitter, Facebook, YouTube, blogs, chat rooms, and others require a new set of methods and technological prototypes to conduct analysis—particularly when trying to gain sociocultural understanding in the short term. We now turn to these newer methods and technologies, in the hope of introducing the reader to technologies that are ready for use but perhaps not yet widely known. Because these technologies are at different stages of readiness for use, we have grouped them according to whether they are deployed, whether they are in prototype form but not deployed, or whether they are in a research phase.



## **Deployed Technologies**

### **3.1. Topsy**

Topsy analyzes tweets and web pages gathered from millions of websites, blogs, and social media services. The platform supports all languages, and combines a search engine and analytic tools that index, analyze, and rank content and trends. Topsy's pipeline architecture currently processes over 400 million posts every day from multiple data sources, generating ranked and scored results within a few minutes. Topsy currently operates the world's largest global index of Twitter data and gives access to conversations in near-real time as well as to an archive going back several years. Customers that include U.S. Government agencies, global marketing firms, and large online publishers use Topsy to monitor activity by keyword, domain, and geographic areas and analyze content by media type (links, photos, videos), author, and sentiment. In addition, Topsy's indexing and live-ranking technologies process content from social networks, helping users to determine the relative resonance of various topics and to identify "tipping points" in message exposure. Users can also develop predictive trending metrics and conduct historical analysis to correlate social media activity with real-world events. This analysis provides the ability to view historical activity and to compare activity levels for related topics over long periods of time. The platform supports any language regardless of character set.

### **3.2. InTTENSITY**

InTTENSITY has created a suite of tools to enable sociocultural understanding (inTTENSITY, 2013). One of them, the Social Media Command Center (SMCC), combines a social media harvesting service with extraction and categorization engines; the combination of these provides an analytical capability using a cloud-based approach. The SMCC examines 75 million social media sources, including Twitter, Facebook, and a variety of public blogs and pages. It specifically looks for spikes in social media volume of any sort compared to a normal distribution (as calculated and adjusted periodically by a baselining activity). For example, if normal Twitter frequency for a given day and hour is approximately 1,000 tweets per second and the SMCC detects a spike up to 3,000 that persists for a sustained period within that timeframe, the system identifies and collects statistics about that spike. In this way, researchers can use the SMCC to detect and gather information on any high-impact events, including mob violence, civil disturbances, natural disasters, or any other event where people use their mobile devices to "check in."

SMCC can filter data by geography and keywords, enabling users to focus on an event within a specific area. The system works with unstructured text data and enables users to manipulate the processed data based on specific categories. Furthermore, the classification technology recognizes and displays relationships between data points and divides the results into contextual groups. Its clustering functionality classifies unstructured information into thematic groups, identifies the major topics in documents, and recommends structures for classification.

InTTENSITY has also built the InTTENSITY Analyze tool, which transforms text in social media, emails, surveys, and other sources into insights regarding sentiments and trends. Specifically, InTTENSITY allows users to automatically discover topics of conversation in social media, emails,

surveys, and other sources; analyze coverage in social media; profile social media user behavior and preferences; and measure the effectiveness of information campaigns.

### **3.3. Worldwide Integrated Crisis Early Warning System (W-ICEWS)**

W-ICEWS, developed by Lockheed Martin with funding by the Department of Defense (DoD), consists of technologies that monitor, assess, and forecast the occurrence and evolution of instability events throughout the world (Lockheed Martin, 2013). W-ICEWS provides situation awareness that is designed to help produce planning products, crisis action planning, and post-execution analysis. The system comprises four components that together perform near-real time data extraction and analysis from news reports and information generated by the government Open Source Center (OSC).

W-ICEWS Data Management (iDATA) provisions W-ICEWS models in near-real time with data from over 6,000 international, regional, national, and local news sources (Lockheed Martin, 2013). The software has processed more than 30 million news stories in English, Spanish, and Portuguese from the past 13 years to extract “who, did-what, to-who, when, where.” iDATA uses deep (BBN Serif) and shallow (JabariNLP) parsing technologies to produce over 19 million unique geolocated events with an accuracy of greater than 80 percent. iDATA draws over 300 different types of coded events from the CAMEO (Conflict and Mediation Event Observations) taxonomy, with each event type having an observer-neutral intensity (Goldstein) score that represents how hostile or how cooperative the event is. The actors (country, sector, organization, individual) involved in events come from dictionaries of over 50,000 named and time-indexed entities as well as over 700 generic agents (e.g., police, government official, protestor). The iDATA repository also contains data from 30 different sources that contain primarily quantitative data on over 175 different countries. Because many countries do not publish reliable information (e.g., the gross domestic product of Afghanistan), iDATA uses a type of data imputation known as the hybrid copula method to impute missing data records. A copula is a joint cumulative distribution function that captures the dependence among a set of random variables (Hollenbach et al., 2013). It is a function that binds together two or more univariate marginal distributions of known form to produce a new joint distribution (Trivedi & Zimmer, 2005).

W-ICEWS Trending, Recognition, and Assessment of Current Events (iTRACE) aids situational understanding through analysis and visualization of event history trends and patterns generated by iDATA (Lockheed Martin, 2013). It gives users capabilities to generate time series, map-based views, trends, relationships, matrices, and other visualizations, with drill-down to underlying stories. iTRACE provides an automated capability to monitor political activity around the globe by automatically converting news reports into structured indices that reflect the character and intensity of interactions among key leaders, organizations, and countries—who is doing what to whom, when, where, and how around the world. Users from various military, government, and intelligence communities use iTRACE to create analytical products. iTRACE was transitioned into U.S. Strategic Command’s Integrated Strategic Planning and Analysis Network (ISPAN) framework in the spring of 2012 on both the Secret Internet Protocol Router Network (SIPRNet) and the Joint Worldwide Intelligence Communication System (JWICS). This capability, coupled with other W-

ICEWS components, is available on U.S. Southern Command's unclassified W-ICEWS servers for test and evaluation.

W-ICEWS Forecasting (iCAST) provides the capability to forecast instability events around the globe, using a mixed methods modeling approach to forecast instability (Lockheed Martin, 2013) iCAST combines forecasts from heterogeneous statistical and agent-based models to generate an aggregate forecast with accuracy greater than 90 percent (Ibid). Users can drill down into the underlying model variables and data to gain detail on the forecasts and can experiment with the effects of changes on the model indicators in a "what-if" scenario.

Finally, W-ICEWS Sentiment Analysis (iSENT) measures attitudes and perception about issues, people, and events through sentiment analysis from blogs, tweets, and Facebook (Lockheed Martin, 2013). It also shows sentiment propagation across the internet and identifies key sites and people in shaping opinion dynamics. Open source digital content generated by recent events, such as the Arab Spring and London riots, creates a valuable source of information. Using such information, iSENT can give intelligence analysts the tools to understand emerging regional trends and sentiment, predict threats from groups and individuals or find the proverbial "needle in a haystack." Its analysis algorithms are designed to distinguish between useful information and noise in a rapidly changing content. iSENT developers have conducted experiments with operational organizations, transitioned the capabilities for operational use by intelligence analysts and desk officers, and integrated with other W-ICEWS components and the ISPAN program of record.

### **3.4. Marine Corps Civil Information Management System (MARCIMS)**

U.S. Military Civil Affairs (CA) teams currently lack the ability to efficiently share and analyze critical information and experiences they collect during engagements with the civilian environment. MARCIMS aids effective Civil-Military Operations (CMO) by providing knowledge of the battlespace through mobile computing technologies, semantic information and knowledge management, and Web-based geospatial decision support capabilities. MARCIMS's technical approach begins with data collection using apps on iOS and Android smartphones with configurable forms and real-time submission of data to a central repository through local cellular or wireless networks. In addition, MARCIMS displays field-collected CIM data in context with basemaps and overlays of relevant imagery and geo-data. Users can access geospatial analysis and statistical reasoning tools via standard Web service protocols. Finally, MARCIMS enables users to manage and analyze relationships between field and reference datasets in a Semantic Wiki interface.

## **Prototype Technologies**

### **3.5. TweetTracker**

Arizona State University professor Huan Liu and his colleagues designed the TweetTracker to assist users involved in Humanitarian Aid and Disaster Recovery and other complex operations to identify topics of current concern to crowds (Goolsby & Carley, 2013). Liu collected over 200 million disaster- and crisis-related tweets. U.S. European Command has already instantiated the TweetTracker in its Social Media Dashboard.

TweetTracker helps users to spot rumor and viral information by leveraging the capabilities of news organizations and other sources that collect, sort, vet, and distribute information. For example, TweetTracker can help a user find the top retweeted URL, breaking news, influential people, human sensors on the ground providing on-the-scene reporting, and other items of particular interest to Military Information Support and/to Operations (MISO).

In addition, TweetTracker capitalizes on the fact that many Twitter users gather and generate data professionally and semi-professionally (i.e., newscasters, bloggers, etc.). These users collect vast amounts of detail gleaned from their own informants, discover important events, and distribute vetted information about events. TweetTracker enables users to monitor, leverage, and exploit such broadcast information, and discover capable crowd networks and to analyze information pathways and content for actionable intelligence.

Along with the capabilities described above, TweetTracker allows users to find illustrative tweets, a view of the overall network, the most retweeted actors, the hashtag network, and the most critical hashtags (i.e., those that co-occur most frequently with other hashtags). At present, users can compare the numbers of tweets marked with different hashtags. Furthermore, users can separate English from non-English tweets. Finally, users can display trends in news articles and in Twitter items on the same topic, conduct multi-country comparisons of the number of tweets per hour, and track the number of tweets mentioning a topic across time.

### **3.6. Social Radar**

The MITRE Corporation's Social Radar prototype provides population-centric signal-detection capabilities that can yield early insight into diverse events such as unrest in Africa, natural disasters in East Asia, or disease outbreaks. The Social Radar capabilities enable users to track perceptions, attitudes, beliefs, and behaviors of a population as expressed via Twitter, blogs, and news. Social Radar offers an indications and warning capability, in which users launch mini-applications or "widgets" that may work independently in a modular fashion or may communicate with each other. Within the Social Radar prototype, researchers can utilize two applications to understand the sentiments or emotions expressed regarding a topic.

The Sentimdir prototype (Day, Boiney, Ubaldino, & Brown, 2012) supports detection and tracking of trends in sentiment as expressed in news articles and comparable texts. Users can quantify and visualize the direction and rate of change in sentiment, while controlling for various features of the text such as topic, source, geography, persons named, and more. Users can design highly tailored queries and save them for efficient reuse, updating, and collaborative sharing. They can also drill down to individual sentences or documents, to view annotations of sentiment in context. The technical core of Sentimdir is a set of statistical sequence models that use semantic and lexical features of individual sentences to infer three elements of sentiment expression: (a) *opinion holder*—the person, organization, or metonymous mention of a geopolitical entity, (b) *sentiment expression(s)*—specific terms or phrases, in context, that convey the negative or positive polarity of an opinion holder's sentiment, and (c) *sentiment target*—the concept or entity on which the opinion holder's expression of sentiment is focused.

The MoodMiner tool suite comprises a set of behavioral scientific analytic methods and an automated prototype for understanding trends in emotional expressions. At present, MoodMiner has been applied only to Twitter, but will be extended to other social media. This tool enables users to query tweets and plot trends in emotion levels in real time—both generalized emotions and expressions regarding specific targets. The emotion categories used within the prototype are drawn from social psychological research that examined linkages between the words people use in written texts and the psychological states they experience. Specifically, the research used to develop the Linguistic Inquiry and Word Count (LIWC) framework within social psychology also informed the development of the prototype.

In addition to monitoring the emotions and sentiments expressed toward a topic over time, users may wish to understand—at a glance—the most relevant topics being discussed in a set of text data or to pinpoint the demographic characteristics of those posting tweets. Alternately, users may wish to uncover deception in the tweets. Each of these analyses can be conducted within Social Radar.

Users can draw on Comment Filter (CoFi) to perform exploratory analysis of text data sets in any language by using NLP techniques to group similar comments and to prioritize messages such that the most relevant items can be identified quickly (Doran, Zarrella, & Henderson, 2012). CoFi automatically discovers topics, grouping similar comments, sorting comments by relevance, and providing drill-down and timeline-based visualization. It works with other analytics that detect elevated levels of information sharing behavior, such as retweets, and can partition social media comments by language and country of origin to provide a more targeted perspective to the analyst interested in drilling down to search messages for actionable content.

Users who wish to understand the demographic characteristics of those posting to Twitter can draw on Author DNA to characterize unknown authors along several attributes, including gender, age, and location. Burger, Henderson, Kim, & Zarrella (2011) cast the problem as one of text classification, viewing each author as a set of identifying features derived from the content and metadata associated with all of the author's postings. These features include the sets of words and characters used in the tweet text itself, as well as the author's self-description, screen name, time zone, posting times, text length, emoticons, capitalization and punctuation density, and numerous other attributes associated with a user's social activity. Burger, Henderson, Kim, & Zarrella, (2011) built a statistical profile over these units, and compared it to known samples by measuring the similarity of these distributions. This approach is agnostic about language or writing system, making it applicable to many organizations' problem spaces.

Author DNA has so far focused on machine learning algorithms that can be trained quickly on very large amounts of training data. Henderson, Zarrella, Pfeifer, and Burger (2013) have labeled more than 20 billion tweets with a country of origin, and have performed a series of validation experiments using a sample of Twitter data built over the course of four years. They used blog profile metadata and tweet geo-tags to measure the correctness of their predictions, with results suggesting that the algorithms are roughly 92% accurate at associating a Twitterer with a country.

As a practical matter, Twitterers who produce more tweets are more likely to be tagged more accurately, so the approach is well suited for discovering highly active Twitterers from any region of the globe.

Users who wish to detect deception or covert activity in social media can turn to the Pinocchio widget, which searches for covert information campaigns operated by hidden networks of social media accounts. "Puppet" accounts such as these are often used to conduct fake grass-roots campaigns to distort the online conversation or magnify perceived influence. Pinocchio searches for a constantly-updated list of rare watermarks that are potentially unique to the style or substance of an individual user's messages, such as unusual phrases, hashtags, or URLs. When the widget discovers multiple accounts that repeatedly share the same rare watermarks, it flags those users as potentially suspicious. When those flags start appearing between all members in a much larger set of accounts, at statistically unusual rates, Pinocchio builds up a high-level picture of who is colluding to deceive honest users of social media.

## **Research Phase Technologies**

### **3.7. Program on Social Media in Strategic Communication (SMISC)**

According to the website of the Defense Advanced Research Projects Agency (DARPA, 2013), DARPA launched the Social Media in Strategic Communication (SMISC) program to develop tools that would help human operators to counter misinformation or deception campaigns. The program focuses its research on linguistic cues, patterns of information flow, and detection of sentiment or opinion in information generated and spread through social media. Researchers will also attempt to track ideas and concepts to analyze patterns and cultural narratives. If successful, they should be able to model emergent communities and analyze narratives and their participants, as well as characterize generation of automated content, such as by bots, in social media and crowd sourcing.

SMISC researchers are creating an environment where large amounts of data are collected, with experiments performed in support of development and testing. One example of such an environment might be a closed social media network of 2,000 to 5,000 people who agree to conduct social media-based activities in this network and to participate in required data collection and experiments. This network might be formed within a single organization, or span several. Another example might be a role-player game in which use of social media is central and players have again agreed to participate in data collection and experiments.

## **4. Gaps in Science and Technology**

Although research has made great strides in developing technologies for ingesting and analyzing social media data, certain gaps remain. One technological gap limits the ability to differentiate between people's "public" versus "private" selves from social media data. In addition, given that several of the technologies discussed in this chapter include programs for tracking linguistic patterns in text data, we note two gaps in the area of linguistics and culture. A fourth gap concerns the use of virtual reality games (or lack thereof) in research designed to gain sociocultural understanding. Finally, a fifth gap lies in the scientific tradecraft available for utilizing the diverse

methods of data collection available (e.g., survey research, social media analysis, and others). Specifically, a question arises as to whether insight gained from social media is analogous to, or comparable with, insight gained from surveys, content analysis, focus groups, Mechanical Turk studies, or other forms of data collection. This question is central to determining whether a user can obtain “ground truth” when utilizing any one of these forms of data collection in isolation. We will address each of these gaps in turn.

#### **4.1. Detecting Public versus Private Selves**

As the internet becomes ever more popular worldwide, people increasingly encounter “alternate” social environments and, therefore, opportunities to experience different self-schemas. Although people seek to maintain a consistent view of themselves, the social environment can elicit alternate behaviors and expressions depending on whether it triggers their “public selves” or “private selves.” An individual’s private self engages in thoughts, feelings, and ideas that form his/her self-schema. The public self, however, is the identity revealed to others, and can be context specific. Oftentimes, the person attempts to conceal the private self, while the public self includes the actions that others can see and acknowledge. Both personas can affect behavior. For example, people may follow social norms at work or in certain social contexts; if they do not, they risk being ostracized by their group. In addition to the physical problems that result from being rejected by a group, ostracism can lead to adverse psychological effects, including on one’s overall sense of social well-being (Williams, 2001).

Virtual environments, however, make it difficult to decipher whether an individual is displaying her/his public self or private self. Thus, anyone wanting to tap underlying beliefs would need technology that can distinguish when people are displaying public selves versus private selves. Such technology does not exist at present.

#### **4.2. Psycholinguistics within Foreign Languages**

Several of the automated technologies discussed in this chapter track linguistic patterns in text data—whether these patterns consist of the frequencies of words or of phrase structures (such as subject-verb-object “triples”). Although considerable basic research in the English language has linked the usage of specific words with the emotional/cognitive states that people are experiencing (Tausczik & Pennebaker, 2010), very little such research has focused on non-Western languages and cultures. Thus, it is not always straightforward to interpret the linguistic patterns spotted in languages other than English or what these patterns signify regarding people’s psychological states. Considerable cross-cultural research will be needed to fill this gap in scientific understanding.

#### **4.3. Tracking Linguistic Patterns in Foreign Languages**

Another set of challenges arises because the “languages of the internet” are often combinations of different languages thrown together. For example, Twitter users in the Arab world often combine an Arabic dialect with English (to form what is known as “Arabeezy”) and even throw in Twitter slang and emoticons. The combination of such diverse languages as Arabic and English alone poses a formidable gap that yet-to-be-invented technologies may someday address.

As languages, in and of themselves, have evolved through the centuries, the “internet languages” also evolve—quite possibly at a faster rate. Therefore, any technology used to process language must account for the constantly shifting linguistic patterns of the internet. Whether this requires a “human-in-the-loop” to monitor new language patterns or whether research will produce an automated way of detecting new patterns, anyone wishing to analyze social media communication in the long term will have to grapple with “constant change” on the internet.

#### **4.4. Virtual Reality Games for Gaining Sociocultural Understanding**

Virtual reality simulations are available for training purposes, e.g., to help troops gain sociocultural understanding before deploying overseas, yet few virtual reality games exist for gaining sociocultural understanding. Such games could be designed and implemented in the context of internet experiments (e.g., using the experimental methodology described above). In this kind of experiment, researchers could examine the decisions that participants make during a game and the thoughts they might communicate for clues as to their sociocultural perspectives. Such games have the advantage of being dynamic and engaging, as well as representing a consistent and systematic mechanism for data collection, since they do not depend on a human telephone interviewer to speak exactly the same way in every interview. For a full discussion of the training tools available for gaining sociocultural understanding, see the Training chapter in the Understand section of this book.

#### **4.5. Conducting Multi-Method Analysis**

More than ever before, operational end-users have a choice of data collection methods and analytical processes for gaining sociocultural understanding. These methods include those described in this chapter as well as focus groups, in-depth interviews, on-the-ground observations, and many others. At present, however, few studies have attempted to compare results across different methods—especially when it comes to examining non-Western cultures. A few studies have compared social media results with surveys when it comes to American domestic issues (Mitchell & Hitlin, 2013; O’Connor, Balasubramanyan, Routledge, & Smith, 2010).

This indicates a need for systematic research to understand the factors that may cause different methodologies to converge on the same findings or to diverge—especially with regard to foreign populations. For example, if a social media analysis and a survey project within one culture produce identical results, this “multi-method convergence” suggests that the results are true. However, if the findings diverge, where does “ground truth” lie? At present, little conventional wisdom points the user toward a process for making sense of diverging results. In one culture, people may be afraid to speak the truth in a telephone survey, because they fear being wiretapped, but will readily state their opinions on the internet. However, in cultures that are more open, different methods may yield more convergent findings. In yet another scenario, surveys might tend to reach one demographic group, while social media analyses may cover a different demographic. Again, this pattern of demographic coverage may hold true in some cultures but not in others.

Ultimately, research comparing results across different data formats can clarify how to interpret the results of one type of study. It can also suggest whether or not it would be adequate to conduct only one or another type of study in a given culture, or whether it is more effective to conduct



multiple types of studies. Such research can ultimately contribute toward an analytic tradecraft and best practices for interpreting the results of different types of studies.

## **5. Transition to Operational End-Users**

Although numerous methods and tools now exist for analyzing social media data, developers must meet a series of challenges to transition these tools effectively to operational environments (Allen, unpublished manuscript). For example, a new technology may have unforeseen second-order impacts on business processes and capabilities. To understand the effects on an analytic business process, potential users must outline the business process functions at the most basic level and then postulate specific ramifications ahead of time.

As one example, a new technology provided to intelligence analysts might improve the efficiency of gathering data from intelligence sources. This, in turn, would simplify and speed up the overall super-task of gathering intelligence from multiple sources, permitting analysts to go beyond “Source A” alone to other sources. However, collecting intelligence from any source comprises a set of smaller implicit tasks, such as “understand the data type,” “understand the context in which the data was produced,” “understand the technical limitations of the collector,” and “judge the relevance of the data source to the question.” If the analyst was already familiar with Source A, each of these sub-tasks was encapsulated in the super-task. By collapsing several iterations of the super-task (i.e., “Gather from Source B, ...C, ...” etc.), the new technology exposes the finer-grained tasks, because the analyst is less familiar with the new sources. As a result, while the new technology may consolidate the data-gathering tasks, it extends the processing for validity and relevance enormously—an unintended consequence (Allen, 2007).

Another set of challenges to effective transition of technologies lies in the skill sets that analysts need in order to utilize new technologies effectively. Potential users should carefully consider whether the technology requires advanced disciplinary or specialized training (such as a Master’s or PhD in a particular discipline) in order to understand or use it. If the technology makes use of behavioral principles or mathematical processes, for example, agencies may need either to ensure that someone trained in the relevant disciplines is on staff to help analysts navigate the technology, or to provide elaborate training, manuals, and ongoing support as analysts learn the underlying principles for using the technology appropriately.

As one example, analysts must understand the nuances and limitations of particular data types in order to synthesize multiple data sources effectively, especially because most analysts “grow up” within a particular intelligence discipline (Allen, 2007). Developing an analytic strategy across different types of data without having familiarity with all of them, as occurs when huge repositories become available, adds substantially to an analyst’s tasks. Furthermore, as described above, altering business processes to incorporate such “new” tasks has a significant impact upon timing, planning, and production processes.

Furthermore, analysts must understand how to develop query strategies in order to pull the data sets they need from the voluminous amounts of data available (Allen, 2007). To date, the typical

analyst is not well versed in more than Boolean or frequency-based querying. Development of query strategies appropriate to the amount and type of data, the particular problem set, and the technical requirements of the storage solution requires better understanding and manipulation of semantic, statistical, and conceptual clustering strategies. To this point, tool developers have made significant strides on developing novel exploitation algorithms, but bridging the gap to the user has been difficult.

The final challenge relates to analysts' overall comfort level with using technology in general, and in particular, with going beyond Microsoft Office. Analysts might require careful training in the use of any new technology, or may need to rely on excellent "cultural brokers" to facilitate the use of that technology. Ultimately, the adoption of new technologies requires adjustments to the new tools, techniques, and processes. Challenges concerning the skill levels of analysts may be especially difficult to address, because they involve changing *people* as opposed to systems or processes.

Separate barriers to effective technology transition lie in the dissemination processes currently used by government agencies. For example, data-sharing agreements made at one agency may incorporate some limitations regarding the types of users allowed to receive the data. This situation creates problems for analysts working in one unit but supporting a Joint or Coalition activity. Furthermore, some regulations stipulate that data can only be retained for a specified length of time, causing problems when performing tasks that require historical data. Similar challenges arise when data sets may be available to answer a question but are not available for review by another organization. For example, an analyst within one unit might solve an analytic problem, but cannot release the relevant detail to the supported unit because of the nature of the information used.

## **6. Summary**

In this chapter, we provided an overview of theoretical frameworks and technologies that researchers can use to gain sociocultural understanding. Although these frameworks and technologies are rigorous, they are not yet widely used in operational contexts. We outlined technologies on the horizon for gaining sociocultural understanding in the medium term and for conducting rapid analysis of large volumes of data for near real-time understanding. This latter mission brings with it the challenge of converting massive amounts of data into tractable units for analysis. Although the technologies available do an impressive job of exactly that, we highlighted gaps in these technologies and in the science of sociocultural behavior sensemaking underlying them. We also highlighted some possible shortcomings that may limit attempts to transition technologies into operational contexts. We hope that the reader will come away with a greater understanding of the avenues available for sociocultural analysis as well as a realistic understanding of the challenges involved in pursuing these avenues.

## References

- Allen, S. (2007). *Analytic implications of the technical development of the US Army intelligence and security command's information dominance center*. Unpublished manuscript. McLean, VA: The MITRE Corporation.
- Braver, S. L., & Bay, R. C. (1992). Assessing and compensating for self-selection bias (non-representativeness) of the family research sample. *Journal of Marriage and the Family*, 54(4), 925–939.
- Burger, J. D., Henderson, J., Kim, G., & Zarrella, G. (2011, July). Discriminating gender on Twitter. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing* (pp. 1301-1309). Association for Computational Linguistics. Edinburgh, Scotland.
- Cahalan, D. (1968). Correlates of response accuracy in the Denver validity study. *Public Opinion Quarterly*, 32, 607–621.
- Cleveland, W. L. (1971). *The making of an Arab nationalist: Ottomanism and Arabism in the life and thought of Sati' al-Husri*. Princeton, NJ: Princeton University Press.
- DARPA. (2013). *Social Media in Strategic Communication (SMISC)*. Retrieved from: [http://www.darpa.mil/Our\\_Work/I2O/Programs/Social\\_Media\\_in\\_Strategic\\_Communication\\_\(SMISC\).aspx](http://www.darpa.mil/Our_Work/I2O/Programs/Social_Media_in_Strategic_Communication_(SMISC).aspx)
- Dawisha, A. (2003). *Arab nationalism in the twentieth century: From triumph to despair*. Princeton, NJ: Princeton University Press.
- Dawn, C. E. (1973). *From Ottomanism to Arabism: Essay on the origins of Arab nationalism*. Chicago, IL: University of Illinois Press.
- Dawn, C. E. (1988). The formation of pan-Arab ideology in the interwar years. *International Journal of Middle East Studies*, 20, 67–91.
- Day, D., Boiney, J., Ubaldino, M., & Brown, T. (2012, July). Multi-channel sentiment analysis. In *Proceedings of the 4th International Conference on Applied Human Factors and Ergonomics (AHFE) International Cross-Cultural Decision Making, Focus*, San Francisco, CA.
- Doran, C., Zarrella, G., & Henderson, J. C. (2012, June). Navigating large comment threads with CoFi. In *Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies: Demonstration Session* (pp. 9-12). Montreal, Canada.
- Elson, S. B., Yeung, D., Roshan, P., Bohandy, S., & Nader, A. (2012). *Using social media to gauge Iranian public opinion and mood after the 2009 presidential election*, TR-1161-RC. Santa Monica, CA: RAND Corporation.
- Gershoni, I., & Jankowski, J. P. (1995). *Redefining the Egyptian nation, 1930–1945*. Cambridge, UK: Cambridge University Press.
- Gibson, J. L. (1992). Alternative measures of political tolerance: Must tolerance be 'least-liked'? *American Journal of Political Science*, 36(2), 560–577.
- Gordon, R. A. (1987). Social desirability bias: A demonstration and technique for its reduction. *Teaching of Psychology*, 14(1), 40–42.
- Granato, J. & Scioli, F. (2004). Puzzles, proverbs, and omega matrices: The scientific and social significance of empirical implications of theoretical models (EITM). *Perspectives on Politics*, 2, 313–323.
- Haim, S. G. (1955). Islam and the theory of Arab nationalism. *Die Welt des Islams, New Series*, 4(2–3), 124–149.
- Haim, S. G. (1962). *Arab nationalism: An anthology*. Berkeley, CA: University of California Press.
- Heath, A., Fisher, S., & Smith, S. (2005). The globalization of public opinion research. *Annual Review of Political Science*, 8, 297–333.
- Henderson, J., Zarrella, G., Pfeifer, C., & Burger, J. D. (2013, June). Discriminating non-native English with 350 words. In *Proceedings of the Eighth Workshop on Innovative Use of NLP for Building Educational Applications, Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, (pp. 101–110), Atlanta, GA.
- Hollenbach, F.M., Metternich, N.W., & Ward, M.D. (2013). *Fast and easy imputation of missing data*. Retrieved from <http://mdwardlab.com/biblio/fast-and-easy-imputation-missing-data>.
- Hyman, H. H. (1944). Do they tell the truth? *Public Opinion Quarterly*, 8, 557–559.
- inTTENSITY. (2013). *Social Media Command Center*. Catonsville, MD: inTTENSITY. Retrieved from <http://www.inttensity.com/products/social-media-command-center/>.
- Khalidi, R., Anderson, L., Muslih, M., & Simon, R. S. (Eds.). (1991). *The origins of Arab nationalism*. New York, NY: Columbia University Press.
- Khoury, P. S. (1983). *Urban notables and Arab nationalism: The politics of Damascus, 1860–1920*. Cambridge, UK: Cambridge University Press.

- Khoury, P. S. (1987). *Syria and the French mandate: The politics of Arab nationalism, 1920–1945*. Princeton, NJ: Princeton University Press.
- Kuran, T. (1995). The inevitability of future revolutionary surprises. *American Journal of Sociology*, 6, 1528–1551.
- Lockheed Martin (2013). *Worldwide Integrated Crisis Early Warning System*. Retrieved from [http://www.lockheedmartin.com/us/products/W-ICEWS/W-ICEWS\\_overview.html](http://www.lockheedmartin.com/us/products/W-ICEWS/W-ICEWS_overview.html).
- Marsot, A. L. (1968). The role of the ulama in Egypt during the early 19th century. In P.M. Holt (Ed.), *Political and social change in modern Egypt* (pp. 264–80). Oxford, UK: Oxford University Press.
- McColl, E., Jacoby, A., Thomas, L., Soutter, J., Bamford, C., Steen, N., . . . Bond, J. (2001). Design and use of questionnaires: A review of best practice applicable to surveys of health service staff and patients. *Health Technology Assessment*, 5(31) (Executive summary): n.p.
- Mitchell, A., & Hitlin, P. (2013). *Twitter reaction to events often at odds with overall public opinion*. Washington, DC: Pew Research Center. Retrieved from <http://www.pewresearch.org/2013/03/04/twitter-reaction-to-events-often-at-odds-with-overall-public-opinion/>
- Moaddel, M. (1993). *Class, politics, and ideology in the Iranian revolution*. New York, NY: Columbia University Press.
- Moaddel, M. (2005). *Islamic modernism, nationalism, and fundamentalism: Episode and discourse*. Chicago, IL: University of Chicago Press.
- Moaddel, M. (2009). The Iranian revolution and its nemesis: The rise of liberal values among Iranians. *Comparative Studies of South Asia, Africa, and the Middle East*, 29(1), 126–136.
- Moaddel, M., de Jong, J., & Dagher, M. (2011). Beyond sectarianism in Iraq. *Contexts*, 10(3), 66–67.
- Mutz, D. C. (2011). *Population-based survey experiments*. Oxford, UK: Princeton University Press.
- O'Connor, B., Balasubramanyan, R., Routledge, B.R., & Smith, N. A. (2010, May). From tweets to polls: Linking text sentiment to public opinion time series. In *Proceedings of the International AAAI Conference on Weblogs and Social Media (ICWSM 2010)* (pp. 122–129), Washington, DC.
- Parry, H. J. & Crossley, H. M. (1950). Validity of responses to survey questions. *Public Opinion Quarterly*, 14, 61–80.
- Pennebaker, J.W., Chung, C.K., Ireland, M., Gonzales, A., & Booth, R.J. (2007). *The development and psychometric properties of LIWC2007*. Austin, TX: LIWC.net.
- Servi, L. D., & Elson, S. B. (2012). A mathematical approach to identifying and forecasting shifts in the emotions of social media users. MITRE Technical Report MTR 120090. Bedford, MA: The MITRE Corporation.
- Shadish, W. R., Cook, T. D., & Campbell, D.T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Belmont, CA: Wadsworth, Cengage Learning.
- Snyderman, P. M. (1993). The new look in public opinion research. In A. W. Finifter (Ed.), *Political Science: The State of the Discipline II* (pp. 219-245). Washington, DC: The American Political Science Association.
- Swidler, A. (1986). Culture in action: Symbols and strategies. *American Sociological Review*, 51(2), 273–286.
- Tauscik, Y., & Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology*, 29, 24–54.
- Topsy (2013). *Government capabilities*. San Francisco, CA: Topsy.
- Tripp, C. (2000). *A history of Iraq*. Cambridge, UK: Cambridge University Press.
- Trivedi, P. K. & Zimmer, D. M. (2005). *Copula modeling: An introduction for practitioners*. Boston, MA: Now Publishers Inc.
- Williams, K. (2001). *Ostracism: The power of silence*. New York, NY: The Guilford Press.
- Zeine, N. (1973). *The emergence of Arab nationalism*. Delmar, NY: Caravan Books.