NAVAL POSTGRADUATE SCHOOL

MONTEREY, CALIFORNIA

THESIS

CROWD-BASED TECHNIQUES TO IMPROVE INTELLIGENCE ANALYSIS

by

Sridhar Srinivasan

September 2018

Co-Advisors: Robert L. Simeral
Erik J. Dahl

THIS PAGE INTENTIONALLY LEFT BLANK

REPORT DOCUMENTATION PAGE			*Form Approved OMB No. 0704-0188*
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instruction, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188) Washington, DC 20503.			
1. AGENCY USE ONLY ***(Leave blank)***	**2. REPORT DATE** September 2018	**3. REPORT TYPE AND DATES COVERED** Master's thesis	
4. TITLE AND SUBTITLE CROWD-BASED TECHNIQUES TO IMPROVE INTELLIGENCE ANALYSIS			**5. FUNDING NUMBERS**
6. AUTHOR(S) Sridhar Srinivasan			
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Naval Postgraduate School Monterey, CA 93943-5000			**8. PERFORMING ORGANIZATION REPORT NUMBER**
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) N/A			**10. SPONSORING / MONITORING AGENCY REPORT NUMBER**
11. SUPPLEMENTARY NOTES The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government.			
12a. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release. Distribution is unlimited.			**12b. DISTRIBUTION CODE** A

13. ABSTRACT (maximum 200 words)

The essential nature of the homeland security enterprise involves making consequential and complex policy decisions under uncertainty. The inputs that policy makers use in making these decisions are facts, analyses, and predictions (which can fit a definition of intelligence)—all of which are subject to significant uncertainty. This thesis seeks to improve analysis by developing a crowd-based analytic methodology to address the problem of intelligence analysis while accounting for, and taking advantage of, the unique characteristics of the intelligence analysis process and the U.S. Intelligence Community culture itself. The thesis's proposed methodology applies learning regarding crowdsourcing and prediction markets–based forecasting in a new context—that of intelligence analysis and the Intelligence Community. If the Intelligence Community implements the crowd-based analytic proposed methodology, which has achieved results in other contexts, it should improve its predictions of real-world events.

14. SUBJECT TERMS intelligence, analysis, prediction markets, crowdsourcing, methodology, superforcasting			**15. NUMBER OF PAGES** 153
			16. PRICE CODE
17. SECURITY CLASSIFICATION OF REPORT Unclassified	**18. SECURITY CLASSIFICATION OF THIS PAGE** Unclassified	**19. SECURITY CLASSIFICATION OF ABSTRACT** Unclassified	**20. LIMITATION OF ABSTRACT** UU

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. 239-18

THIS PAGE INTENTIONALLY LEFT BLANK

Approved for public release. Distribution is unlimited.

CROWD-BASED TECHNIQUES TO IMPROVE INTELLIGENCE ANALYSIS

Sridhar Srinivasan
Management and Program Analyst, Immigration Records and Identity Services,
Department of Homeland Security
BSE, University of Pennsylvania, 1987
MA, Columbia University, 1990
MBA, Columbia University, 1996

Submitted in partial fulfillment of the
requirements for the degree of

**MASTER OF ARTS IN SECURITY STUDIES
(HOMELAND SECURITY AND DEFENSE)**

from the

**NAVAL POSTGRADUATE SCHOOL
September 2018**

Approved by: Robert L. Simeral
Co-Advisor

Erik J. Dahl
Co-Advisor

Erik J. Dahl
Associate Chair for Instruction,
Department of National Security Affairs

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

The essential nature of the homeland security enterprise involves making consequential and complex policy decisions under uncertainty. The inputs that policy makers use in making these decisions are facts, analyses, and predictions (which can fit a definition of intelligence)—all of which are subject to significant uncertainty. This thesis seeks to improve analysis by developing a crowd-based analytic methodology to address the problem of intelligence analysis while accounting for, and taking advantage of, the unique characteristics of the intelligence analysis process and the U.S. Intelligence Community culture itself. The thesis's proposed methodology applies learning regarding crowdsourcing and prediction markets–based forecasting in a new context—that of intelligence analysis and the Intelligence Community. If the Intelligence Community implements the crowd-based analytic proposed methodology, which has achieved results in other contexts, it should improve its predictions of real-world events.

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF TABLES

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF ACRONYMS AND ABBREVIATIONS

ACE	Aggregative Contingent Estimation
Brexit	exit of Britain from the European Union
CDA	continuous double auction
DPM	dynamic pari-mutuel market
EMH	efficient markets hypothesis
IAEA	International Atomic Energy Agency
IARPA	Intelligence Advanced Research Projects Activity
ICPM	Intelligence Community prediction markets
IEM	Iowa electronic market
IPO	initial public offering

THIS PAGE INTENTIONALLY LEFT BLANK

EXECUTIVE SUMMARY

The essential nature of the homeland security enterprise involves making consequential and complex policy decisions under uncertainty. The inputs that policy makers use in making these decisions are facts, analyses, and predictions (which can fit a definition of intelligence), all of which are subject to significant uncertainty. Reduction in the uncertainty associated with these inputs may improve the soundness of decision-making by policy makers. This thesis seeks to improve analysis by developing a crowd-based analytic methodology to address the problem of intelligence analysis while accounting for and taking advantage of the unique characteristics of the intelligence analysis process and the U.S. Intelligence Community culture itself.

The methodology developed in this thesis utilizes prediction markets–based techniques and crowdsourcing techniques that have significantly improved forecast accuracy in other contexts found in the literature. The thesis's particular contribution focuses on understanding the unique characteristics of the Intelligence Community culture and work processes, and it uses this understanding to inform the design of the proposed crowd-based intelligence forecasting methodology. It can be argued that any analytic methodology hoping to improve the predictive accuracy of the Intelligence Community analysts must both reflect and adapt to the underlying Intelligence Community culture. If it does not, it is likely that any new or modified methodology either may be limited in its adoption, or more likely, be ignored by the intelligence analytic community at large.

The thesis's proposed methodology applies learning regarding crowdsourcing and prediction markets-based forecasting in a new context, that of intelligence analysis and the Intelligence Community. This research excludes quantitative probabilistic assessments, quantitative and qualitative models, and polls-based techniques from consideration because others have already done extensive work on utilizing these techniques in an intelligence context.

This thesis discusses the characteristics of the proposed crowd, the proposed structure of the forecasting effort, the proposed incentive structure, the proposed task design, and the proposed prediction market design and associated structural parameters underlying the forecasting effort, as well as the key characteristics of the proposed platform used to implement the prediction market. Additionally, the thesis uses all of these critical concepts to design a methodology—a crowd-sourced forecasting tournament—that the Intelligence Community can use to improve its forecast accuracy. If implemented, the proposed methodology should improve Intelligence Community predictions of real-world events, based on results achieved in other contexts.

The thesis proposes that the utility of the methodology be demonstrated to the analytic branches of intelligence using a pilot program to help get buy-in to the methodology as a whole, as well as to engender participation in the methodology's prediction market from individuals and teams drawn from the analytic community. If positive, the results of the pilot program may also be used to justify the Intelligence Community spending the financial, analytic time based, administrative time based, and other resources to implement the methodology. Finally, the proposed pilot should allow practitioners to test and tweak various aspects of the methodology from outreach to task design to ensure that the implemented methodology does indeed result in the analytic improvements as it seeks to do.

This thesis is just a starting point; the methodology should be subject to several rounds of peer review and revision before implementation even in pilot form takes place. Once this review and revision occurs, practitioners can implement the pilot, and ascertain if the methodology creates consistently more accurate forecasts than traditional methods. If the pilot is successful, the methodology becomes one more tool in the intelligence analysts' quiver.

ACKNOWLEDGMENTS

The author is grateful for all the support and encouragement received from Rebecca Callahan, his wife. The author also wishes to thank his co-advisors, Robert Simeral and Erik Dahl, for their support, guidance, and hard work on his behalf

THIS PAGE INTENTIONALLY LEFT BLANK

I. INTRODUCTION

A. PROBLEM STATEMENT

This thesis seeks to improve intelligence analysis by using relevant practices in crowdsourcing and prediction markets design, implementation, and operation to develop a crowd-based analytic methodology applicable to the problem of intelligence analysis. The methodology accounts for and takes advantage of the unique characteristics of the intelligence analysis process and of the Intelligence Community culture itself. The emphasis on Intelligence Community culture is deliberate as the proposed methodology seeks to improve forecasting by taking advantage of the same cultural characteristics that are problematic in traditional intelligence analysis that negatively affect forecast accuracy, forecast applicability, and forecast utility. The proposed methodology can be implemented, tested, and if successful, adopted by the Intelligence Community in an effort to address these cultural issues. Although the thesis proposes an analytic methodology to improve forecast and analytic accuracy, it leaves the implementation and testing of that methodology to others.

The essential nature of the homeland security enterprise involves making consequential and complex policy decisions under uncertainty. Policy makers use inputs, such as facts, analyses, and predictions (which can fit a definition of intelligence) to make these decisions, all of which are subject to significant uncertainty. Reduction in the uncertainty associated with these inputs may improve the soundness of policy decisions. However, the enhancement of the predictive quality and accuracy of intelligence may not always improve decision quality and outcomes. Even if methodology, such as this thesis proposes, improves the quality and accuracy of intelligence inputs, it may not be possible to assess how these changes impact policy outcomes. Researchers usually perform assessments of policy outcomes well after the fact, such as the cases with the published examination of events leading up to Pearl Harbor and 9/11, which suggests a long lag time between analysis and its impact on the intelligence process. Still, it is unlikely that improvements to the quality and accuracy of intelligence in the short term may negatively impact policy. Indeed, improvements in the quality and accuracy of intelligence leave policy makers in a better position.

Although a number of preliminary efforts have been made to improve the Intelligence Community's forecast accuracy to date, they do not appear to have led to a significant improvement in the ability of the Intelligence Community to anticipate and prevent terrorist attacks and other harmful events. Indeed, tactical and strategic surprises resulting from other actors' actions (e.g., those of North Korea or Russia) are still problematic. However, the prospect of improving policy outcomes by enhancing the predictive quality and accuracy of intelligence does provide a rationale for the Intelligence Community to seek continual improvement in its analyses.

One way to achieve improvements in analysis may be for the Intelligence Community to utilize crowd-based and prediction markets-based forecasting techniques. Indeed, in the past five years, the Intelligence Community has started to explore the potential of these techniques to improve its understanding of the timing, type, and qualitative and quantitative characteristics of events of interest to policy makers, and by definition, intelligence analysis. The Good Judgment Project is the most salient example of these attempts focusing on crowdsourcing in intelligence.

The Good Judgment Project is sponsored by the Intelligence Advanced Research Projects Agency (IARPA) through its Aggregative Contingent Estimation (ACE) program.[1] The project involves implementing crowdsourced prediction techniques for forecasting event outcomes related to questions of interest to the Intelligence Community and its clients. This project also tests the ability of graduate students, faculty, and practitioners from the political science realm to forecast global geopolitical events.[2] Surprisingly, in 2012, see *The Good Judgment Project: A Large Scale Test of Different Methods of Combining Expert Predictions* by Ungar et al. who found that the most successful lay forecasters participating in the project exceeded the success rate of career

---

[1] "About IARPA," Intelligence Advanced Research Agency, accessed February 3, 2017, https://www.iarpa.gov/index.php/about-iarpa.

[2] "About IARPA."

intelligence analysts in predicting geopolitical events by over 30 percent.[3] These successful lay forecasters are called *superforecaster*s. Superforecasters are individuals and teams who are consistently better than the top two percent of all forecasters and make accurate forecasts about events of any type.[4]

Other efforts using prediction markets-based techniques in non-intelligence contexts for the prediction of political, geopolitical, financial, and business related events have met with similar success, with success defined as making significantly more accurate forecasts than alternative techniques, such as surveys, polls, and fundamental analysis. Indeed, studies of the accuracy of prediction market forecasts under different scenarios have found prediction markets make accurate forecasts of events under a wide variety of conditions.[5]

Despite significant evidence of the utility of crowd and prediction markets-based techniques in forecasting, the Intelligence Community does not seem to be utilizing these techniques as part of its analytic toolkit beyond general interest and a few pilot projects, like the Good Judgment Project. However, Kajdasz et al. examined the use of prediction markets in the Intelligence Community and provided direction for such an effort in the future. They say that any "Intelligence Community Prediction Markets (ICPM) should support decision makers, support analysts, identify the best forecasters in the Intelligence Community, and provide a test for future study."[6]

---

[3] Lyle Ungar et al., *The Good Judgment Project: A Large Scale Test of Different Methods of Combining Expert Predictions*, AAAI Technical Report FS-12-06 (Palo Alto, CA: Association for the Advancement of Artificial Intelligence, 2012).

[4] Tam Hunt, "How I Became a Superforecaster," *Slate*, last updated November 19, 2015, http://www.slate.com/articles/technology/future_tense/2015/11/good_judgment_project_how_i_became_a_superforecaster_for_the_intelligence.html.

[5] Kenneth J. Arrow et al., "The Promise of Prediction Markets," *Science* 320 (2008): 877–878; Joyce E. Berg, Forrest D. Nelson, and Thomas A. Rietz, "Prediction Market Accuracy in the Long Run," *International Journal of Forecasting* 24, no. 2 (2008): 285–300.

[6] James E. Kajdasz et al., "An Alternative Analysis Technique: Examining the IC Prediction Market," *Studies in Intelligence* 3, no. 58 (2014): 22–37.

B. RESEARCH QUESTION

The research question for this thesis is how can a crowd-based analytical tool be developed for use by Intelligence Community superforecasters to improve the quality and accuracy of intelligence assessments? To answer this question, this thesis builds on work in the intelligence studies literature on prediction markets, such as the Central Intelligence Agency's journal *Studies in Intelligence*, the *International Journal of Intelligence and Counterintelligence*, as well as works on prediction markets published in other journals or by students at academic institutions.

C. THESIS ARGUMENT

The thesis argues:

- The combination of forecasts, using two independent sources (crowd-based and prediction markets-based techniques) improves forecast accuracy.
- Identification and application of relevant practices in crowd-based and prediction markets design, implementation, and operation drive the improved forecast accuracy.
- Crowd-based and prediction markets-based forecasting techniques can overcome the impact of the characteristics of intelligence community culture that have negative consequences for traditional analytic forecasting. A methodology that adapts these techniques to Intelligence Community culture may result in increased Intelligence Community forecast accuracy, applicability, and utility.

D. RESEARCH DESIGN

The thesis's research design guides the development of the arguments and methodologies that make up the thesis. Careful thought about and creation of a robust research design is critical to address the research question adequately and comprehensively.

1. Object of Study

This thesis begins with an analysis of the implications of Intelligence Community culture for the accuracy, applicability, and utility of intelligence analysis. The goal of this thesis is to determine more precisely just how crowd- and prediction markets-based techniques can be applied within the Intelligence Community to improve forecast accuracy. Toward this end, the thesis analyzes practices in the application of crowd- and prediction markets-based techniques to forecasting in non-intelligence contexts. This researcher then uses this analysis to develop a crowd- and prediction markets-based forecasting methodology for use by the Intelligence Community that will result in more accurate forecasting and analysis. This methodology is designed to address the cultural drivers affecting forecast accuracy (or inaccuracy) within the Intelligence Community. While this thesis develops a plan for the implementation and testing of the proposed methodology, it leaves the actual implementation, testing, and validation of the methodology to others.

2. Selection Criteria and Rationale

To understand the object of study fully, it requires the selection and examination of the following:

- Relevant information on the unique characteristics of the Intelligence Community, its culture, and its techniques for developing analytic products that policy makers use to support decision making. An understanding of Intelligence Community culture is critical to improving Intelligence Community forecast accuracy because intelligence community culture and its impact on traditional analysis is a key driver of forecast *inaccuracy.*

- Relevant information on the design, testing, accuracy, efficacy, and operational and evaluative processes related to prediction markets, crowdsourcing of analytic inputs in an intelligence analytic and creation environment.

A given set of information is relevant to the thesis if it supports the overarching goal of understanding the cultural context of this thesis and the goal of assessing, selecting, and combining best practices regarding crowd-and prediction markets-based techniques into an overall methodology for the Intelligence Community to use to improve its intelligence forecasts. The goal is not to create new methodologies for each crowd- and prediction market-based technique; rather, the end state involves synthesizing existing best practices in the design, implementation, and operation of each technique with an understanding of Intelligence Community culture in a novel way to create a more accurate combined forecasting methodology.

3. Study Limitations and Scope

Intelligence forecasts and other predictions supporting, affecting, or affected by national policy are the boundaries of the types of forecasts this thesis considers. Qualitative analyses of Intelligence Community culture and its implications for intelligence analysis form another boundary of the thesis scope. Commonly accepted design methodologies for prediction markets and crowdsourcing efforts form the remaining boundaries of this thesis scope. All other contexts and forecasting techniques are out of scope by design. Additionally, the scope of this thesis specifically excludes other forecasting methodologies, such as quantitative and qualitative modeling, polling, social network analysis-based forecasting, big data-based forecasting, or any technique not previously cited. Finally, this thesis proposes evaluation criteria and potential tests of the methodology but does not actually test the methodology.

4. Data Sources and Evidence

The thesis consults the rich existing literature produced in both academic and non-academic contexts on Intelligence Community products, processes, culture, and prediction markets, and crowdsourcing. Finally, this thesis uses only open-source information on Intelligence Community processes, products, and accuracy as closed-source information sources are inaccessible.

5. Preview of Thesis Findings

The thesis finds that the key characteristics of Intelligence Community culture that may drive forecast inaccuracy or indeed even forecasting failure will likely have minimal impact when using crowd-or prediction markets-based forecasting techniques. This low impact results from how those Intelligence Community cultural characteristics manifest themselves in the traditional analytic process. Indeed, by applying best practices for the proposed crowd- or prediction markets-based methodologies, it is possible to use these very cultural characteristics to drive improved analysis and forecast accuracy instead.

E. THESIS OVERVIEW

The thesis develops a practical, actionable, and testable crowd-based methodology to improve the accuracy, applicability, and utility of intelligence analysis through:

- a literature review (Chapter II)
- a discussion of relevant practices in crowdsourcing and prediction markets design, implementation, and operation (Chapters III and IV, respectively)
- a discussion of the implications of Intelligence Community culture for crowd sourced and prediction markets-based forecasting techniques (Chapter V)
- a proposed forecasting methodology (Chapter VI)
- a discussion of the implementation and testing of the proposed forecasting methodology and of areas for subsequent research (Chapter VII)

THIS PAGE INTENTIONALLY LEFT BLANK

II. LITERATURE REVIEW

The starting point for this thesis is a review of academic and nonacademic literature on:

- key characteristics of the Intelligence Community, its culture, its work process, and work products, including an assessment of why it is unique compared to other analytic contexts, such as business intelligence or academic analysis
- combination of forecasts to improve forecast accuracy
- utility of prediction markets in improving forecast accuracy
- utility of crowdsourcing intelligence in improving forecast accuracy

This thesis excludes quantitative probabilistic assessments, quantitative and qualitative models, and polls-based techniques from further consideration because extensive work on utilizing these techniques in an intelligence context has already been conducted. The lack of available relevant information in the case of closed source information led to its exclusion from consideration as well.

A. INTELLIGENCE CULTURE AND ITS IMPLICATIONS FOR ANALYSIS

The emphasis on Intelligence Community culture in this thesis is deliberate and results from a review of the literature on Intelligence Community culture, processes, and products. The proposed methodology developed in this thesis seeks to improve forecasting by taking advantage of the very cultural characteristics problematic in traditional intelligence analysis as cited in the literature. These characteristics include problems related to:

- the driving need for consensus coupled with a bias against sharing information

- an emphasis on tradecraft leading to a notion that intelligence analysis is impervious to understanding based on the scientific method
- inappropriate ways of developing and assessing expertise that lead to analytic sclerosis
- cognitive biases distorting analysis
- time constraints resulting in a focus on the short term
- the focus on current production resulting in inappropriate levels of validity testing and a focus on quantity not quality
- the impact of analysts' rewards and incentives being tied to quantity of production and social standing
- the impact of norms, taboos, and secrecy leading to an inability of analysts to challenge accepted judgment and leading to a belief that secret information is of higher quality than other information
- the impact of analysts' training that results in a lack of a coherent professional identity

These negative cultural characteristics affect forecast accuracy, forecast applicability, and forecast utility of forecasts resulting from traditional Intelligence Community analysis. The salience of culture to forecast accuracy is substantiated in other non-intelligence contexts by the literature on organizational behavior and change.[7]

[7] Susan Cartwright and Cary L. Cooper, "The Role of Culture Compatibility in Successful Organizational Marriage," *The Academy of Management Executive (1993–2005)* 7, no. 2 (May 1993): 57–70.

1. Culture, Intelligence, and the Intelligence Community

The *Oxford English Dictionary* defines culture as the "philosophy, practices, and attitudes of an institution, business, or other organization."[8] The U.S. Intelligence Community has its own unique culture, founded in the notions of intelligence exceptionalism articulated by Turner.[9] Intelligence exceptionalism is the idea that intelligence and intelligence forecasting and analysis as practiced by the Intelligence Community have unique characteristics that set them apart from other types of forecasting and analysis. Indeed, it can also be argued that the negative cultural characteristics problematic in traditional intelligence analysis and previously cited are additional sources of intelligence exceptionalism.

Culturally, fields, such as business intelligence, business forecasting, epidemiological intelligence and forecasting, political analysis, market intelligence and forecasting, election forecasting and the like, are seemingly analogous to intelligence analysis in terms of the types of required analysis and the level of consequence of analytic and forecasting failure. However, they are not considered true analogues to intelligence analysis by the Intelligence Community itself. Indeed, "intelligence culture may be regarded as the ideas, responses and behaviors acquired by intelligence communities and conditioned by history and geography."[10] It is distinct from other organizational cultures based on "ideas of secrecy and the provision of accurate, timely and relevant intelligence,"[11] whereas intelligence is defined as "knowledge…the kind of knowledge our state must possess regarding other states in order to assure itself that its cause nor its undertakings fail because its statesmen and soldiers plan and act in

---

[8] *Oxford English Dictionary*, s.v. "Culture," accessed October 15, 2017, http://www.oed.com/search?searchType=dictionary&q=culture&_searchBtn=Search.

[9] Michael A. Turner, "A Distinctive U.S. Intelligence Identity," International Journal of Intelligence and Counter Intelligence, 17 (2004): 42–61; Mark Phythian, "Cultures of National Intelligence," in *Routledge Companion to Intelligence Studies*, ed. Robert Dover, Michael S. Goodman, and Claudia Hillebrand (Abingdon, United Kingdom: Routledge, 2013), 33–41.

[10] Turner, 42–61; Phythian, 33–41.

[11] Mark Phythian and Peter Gill, *Intelligence in an Insecure World* (Cambridge: Polity Press, 2012), 46.

ignorance."[12] Note that intelligence is not confined to activities bound together by secrecy. Sims updates this idea of what intelligence is when she states:

> Intelligence is best defined as information collected, organized or analyzed on behalf of actors or decision makers… (intelligence) may be collected from open (newspapers, books, radio and television), clandestine (national technical means, agents) and "gray" sources [which] include private citizens or companies willing to divulge information during private conversation.[13]

Lowenthal refines Sims's definition of intelligence when he states, "intelligence is the process by which specific types of information important to national security are requested, analyzed and provided to policy makers."[14] By necessity, this understanding of what intelligence is requires that the cultural focus of the Intelligence Community be on delivering analytic products based upon skilled intellectual effort applied to all manner of information. These products need to have relevance for decision makers, defined as accuracy, utility, and applicability.[15] Furthermore, the creation of intelligence products requires that analysts make decisions in a harsh, unforgiving environment with severe consequences for failure.

Intelligence analysts usually make forecasts based on a significantly incomplete and vague set of facts. The accuracy of the facts is indeterminate, and limited feedback is available to refine the analysts' judgments, which amplifies the notion that intelligence and intelligence analysis are indeed exceptional. It also implies that intelligence and intelligence analysis is more consequential than other forms of analysis.

When coupled with historical Intelligence Community strategic cultural factors extant since during the Cold War, this understanding of intelligence and intelligence work products suggests that the intelligence culture is the preeminent driver of

---

[12] Sherman Kent, *Strategic Intelligence for American World Policy* (Princeton, NJ: Princeton University Press, 2015), 76.

[13] Ernest R. May, Roy Godson, and Gary James Schmitt, ed., *U.S. Intelligence at the Crossroads: Agendas for Reform* (Washington, DC: Brassey's, 1995), 48.

[14] Mark M. Lowenthal, *Intelligence: From Secrets to Policy*, 3rd ed. (Washington, DC: CQ Press, 2006), 10.

[15] Phythian and Gill, *Intelligence in an Insecure World*, 87.

Intelligence Community success or failure in making robust estimates and predictions.[16] The relevant historical strategic cultural factors include:

- a tendency that emerged during the Cold War to oversimplify threats.[17]
- an Intelligence Community work product becoming consensus oriented as a matter of political expediency, which leads to a preference for a "group mindset" or "herd mentality" in the preparation of Intelligence Community estimates.[18]
- an increasingly risk averse culture when it comes to estimates so that making or advancing a position that contradicts or challenges the accepted wisdom of the Intelligence Community even when such challenges come from consumers of intelligence at the highest levels is increasingly unlikely.[19]
- an adherence to the rational actor theory.[20]
- an attitude toward gaps in knowledge that can be summarized in the adage "if you don't know the facts, then make the best educated guess you can rather than admit that you don't know."[21]
- an insistence on the part of every director of the Central Intelligence Agency and now the Director of National Intelligence that they have

---

[16] Matthew M. Aid, "Sins of Omission and Commission: Strategic Cultural Factors and U.S. Intelligence Failures during the Cold War," *Intelligence and National Security* 26, no. 4 (2011): 478–494, doi: 10.1080/02684527.2011.580602.

[17] Aid, 479.

[18] Aid, 480.

[19] Aid, 483.

[20] Aid, 484.

[21] Aid, 488.

unfettered access to the president with all the corresponding implications for the politicization of intelligence.[22]

This list of strategic cultural factors can be mapped to the negative Intelligence Community cultural characteristics affecting forecast accuracy, forecast applicability, and forecast utility mentioned previously and discussed in detail as follows. Indeed, these overarching historical strategic cultural factors are still in play today and when coupled with more granular cultural aspects of the Intelligence Community and its processes (discussed in later sections), they require any proposed Intelligence Community analytic methodology to make cultural compatibility central to its design. The notion of the centrality of culture to effective intelligence analysis is supported by the literature.[23] This assertion is substantiated in other non-intelligence contexts by the literature on organizational behavior and change.[24]

2. The Twin Problems of Consensus and of Information Sharing in Intelligence Community Culture

The Intelligence Community consists of a myriad of players, and all have their own unique (cultural) perceptions of what intelligence is and how analysis should be performed. To quote Boardman in his 2006 thesis, "Overcoming the organizational cultures of multiple, disparate agencies, departments and organizations is critical to solving the problem of sharing information and intelligence such that it may be analyzed and utilized by the people who need it."[25]

Given the costs and benefits of reaching consensus on analytic processes and results, each player in the community will likely be biased toward either jockeying for

---

[22] Aid, 490–491.

[23] Troy Michael Mouton, "Organizational Culture's Contributions to Security Failures within the United States Intelligence Community" (master's thesis, Louisiana State University, 2002), http://digital commons.lsu.edu/gradschool_theses/1121; Satgin S. Hamrah, "The Role of Culture in Intelligence Reform," *Journal of Strategic Security* 6, no. 3 (Fall 2013): 160–171, Supplement, *Ninth Annual IAFIE Conference: Expanding the Frontiers of Intelligence*.

[24] Cartwright and Cooper, "The Role of Culture," 57–70.

[25] Chase Boardman, "Organizational Culture Challenges to Intelligence Community Communication and Interaction" (master's thesis, Joint Forces Staff College, 2006), 7.

primacy to the detriment of collegiality and consensus or toward the lowest common denominator between alternative analytic results. As Lowenthal explains, “the interagency process requires bargaining and negotiation…that requires a great deal of time… [and] gives leverage to an agency that refuses to reach agreement… [and] generates substantial pressure in favor of the lowest common denominator.”[26] George’s description of the underpinnings of why this situation occurs is on point when he states, “The [individual] analyst… is likely to believe that his organization’s view should prevail, in part to reflect the primacy of that office in following the topic in question.”[27] This phenomenon is called “tribal think” by Central Intelligence Agency tradecraft expert Jack Davis, and it reflects each agency’s cultural imperative to preserve its prevailing paradigm and tamp down deviant views to maintain its position as primus inter pares.[28]

Furthermore, the Intelligence Community is not a monolithic agency. When it comes to information sharing, each entity within the community has its own distinct culture. However, some common threads persist, including:

- the practice of limited information distribution and existing extensive compartmentalization practices[29]
- a “need to know” as the basis for information sharing within and outside the Intelligence Community, which by its very nature limits what information is shared and when and how it is shared[30]
- the view that information is a source of power[31]

---

[26] Lowenthal, *Intelligence: From Secrets to Policy*, 6.

[27] Roger Zane George, “Beyond Analytic Tradecraft,” *International Journal of Intelligence and CounterIntelligence* 23, no. 2 (2010): 296–306, doi: 10.1080/08850600903566124.

[28] Central Intelligence Agency, *Intelligence Community and Policymaker Integration: A Study in Intelligence Anthology* (Washington, DC: Central Intelligence Agency, 2014), 18, https://www.cia.gov/library/center-for-the-study-of-intelligence/csi-publications/books-and-monographs/intelligence-community-and-policymaker-integration/IC%20and%20Policymaker%20Integration-A%20Studies%20in%20Intelligence%20Anthology.pdf.

[29] George, “Beyond Analytic Tradecraft,” 300.

[30] National Commission on Terrorist Attacks upon the United States, *Final Report of the National Commission on Terrorist Attacks upon the United States* (New York: W. W. Norton, 2004), 471.

[31] James Burch, “The Domestic Intelligence Gap: Progress since 9/11?,” *Homeland Security Affairs* 4 (2008), https://www.hsaj.org/articles/129. Information Sharing.

- the fact that expansion of the pool of those who “need to know” is difficult[32]

Furthermore, information sharing between agencies is not as prevalent as it should be for cultural reasons. As Maras explains:

> Existing organizational cultures in the IC require and encourage secrecy by stressing the necessity to protect their information and clandestine activities. Limited disclosure and secrecy are thus key aspects of IC processes and practices. This leads to limited information sharing... The missions of these agencies [Federal Bureau of Investigation, Central Intelligence Agency, Director of National Intelligence] place them as primary agencies in protecting the United States by engaging in some form of intelligence function. The sharing of information is not explicit or implicit in their missions.[33]

3. Cultural Issues with the Intelligence Analytic Process

The intelligence analytic process itself is subject to intelligence exceptionalism when it comes to culture. The author’s research and analysis suggests that the primary manifestation is seen in the emphasis on tradecraft to the detriment of more scientific analytic methods and unique problems related to the role of experts and expertise. Even when creating and using scientific and technical intelligence or using precision measurement techniques, the Intelligence Community nonetheless relies on idiosyncratic processes to deal with gaps in knowledge or uncertainty. Other less important sources of exceptionalism include unique versions of cognitive bias, such as confirmation bias. According to Hare and Collinson, “extreme time constraints; focus on current production; the rewards and incentives”[34] for analysts; norms, taboos, and the impact of secrecy; and finally, the analyst’s identity and training. As discussed in the next section, each of these

[32] Marie-Helen Maras, “Overcoming the Intelligence-sharing Paradox: Improving Information Sharing through Change in Organizational Culture,” *Comparative Strategy* 6, no. 3 (2017): 187–197, doi: 10.1080/01495933.2017.1338477, 190.

[33] Maras, 190–191.

[34] Nicholas P. Hare and Paul Collinson, “Organisational Culture and Intelligence Analysis: A Perspective from Senior Managers in the Defence Intelligence Assessments Staff,” *Public Policy and Administration* 28, no. 2 (2013): 217–218.

cultural sources of intelligence exceptionalism may have negative consequences for intelligence analysis in general and forecast accuracy in particular.

a. The Impact of the Emphasis on Tradecraft on Intelligence Community Analysis

According to Johnston, an explicit cultural emphasis on treating analysis and the analytic process as tradecraft across the Intelligence Community seems to be evident.[35] Treating analysis and the analytic process as tradecraft implies that analysis is an idiosyncratic process, a black art unknowable to all except to those who have received wisdom from those on the inside. Its implications are:

- The analysis and the analytic process cannot be approached using the rigor of the scientific method.
- The success or failure of the analytic process depends on an intuitive understanding derived from received wisdom coupled with experience and thus cannot be imparted in its most nuanced sense through training.
- The methods and techniques of intelligence analysis are unique, are characterized by being unverifiable, and are unexplainable in some sense.
- The "good" techniques are simply those that have survived through time and are handed down from senior analysts to junior analysts, while lacking comparatively rigorous, testable definitions of what these "good" techniques are.
- The skills of anomaly detection, pattern recognition, and weighing data in terms of its relevance, accuracy, and analytic implications are gained through experience with minimal contributions from training and academic and practitioner research.

[35] Rob Johnston, *Analytic Culture in the U.S. Intelligence Community: An Ethnographic Study* (Washington, DC: Central Intelligence Agency, 2005), 17–21.

- The lessons learned from success or failure are unlikely to be formally captured; instead, they become part of the idiosyncratic lore passed from analyst to analyst.

- The training process becomes subjective, which leads to inconsistencies in analysts' preparation for the tasks at hand.[36]

b. The Impact of the Problem of Expertise Tradecraft on Intelligence Community Analysis

One common cultural factor across the intelligence community is the assessment of someone's level of expertise based on recognition from policy makers for useful written assessments and oral briefs. The perception of success reinforces not only the confidence of the intelligence analysts in their expert judgment but also the confidence of their peers and superiors in said expert judgment. This expert judgment then drives Intelligence Community processes and resulting work products.[37]

This factor contrasts with other fields wherein experts and expertise are defined as those who possess specialized knowledge in a given domain that allows them to: (1) recognize patterns, (2) apply higher order domain specific principles to solve problems more quickly than others, (3) solve problems in their domain with fewer errors than others, (4) possess domain specific short- and long-term memory, and (5) are better at self-monitoring and identifying and filling gaps in domain specific knowledge than others.[38]

When the accuracy of known facts is indeterminate, and limited feedback is available to refine the analysts' judgments, as is the case in the intelligence community, an analyst's application of expert judgment usually involves the creation of a set of mental models based on past successes. This model results in a kind of analytic sclerosis

[36] Johnston, 28–29.

[37] Johnston. 61–62.

[38] Marissa F. McBride and Mark A. Burgman, "What Is Expert Knowledge, How Is Such Knowledge Gathered, and How Do We Use It to Address Questions in Landscape Ecology?," in *Expert Knowledge and Its Application in Landscape Ecology*, ed. Ajith H. Perera, C. Ashton Drew, and Chris J. Johnson (New York: Springer, 2012), 11.

because an analyst constantly attempts to apply these pre-existing mental models to all situations, regardless of whether they are appropriately applied given the relevant range and preconditions inherent in the models. As analysts rely more and more upon their well-honed mental model of the characteristics and behavior of the target of the intelligence effort, the more likely they may miss major gaps or breaks in the continuity of the analysis or key changes that may have occurred in the target of the analytic effort.

Furthermore, the way the Intelligence Community develops, recognizes, and relies on experts and their expertise intelligence analysis and forecasting is contradicted by the academic research on both experts and expertise. According to Tetlock and Gardner, "experts and lay people are sensitive to a range of psychological idiosyncrasies, subjective biases, values, and conflicts of interest."[39] Indeed, experts may know their specific domains but may fail at tasks that reach outside their domains, such as using an interdisciplinary approach to divine the intentions of an adversary. It can also be argued that experts may not necessarily produce the best forecasts. Indeed, Tetlock and Gardner have found that superforecasters are not necessarily accepted as experts in their fields.[40] More often, superforecasters are those who understand that humbleness, an awareness of the complexity of systems of systems, and most crucially, the ability to learn from mistakes are prerequisites for forecasting performance.[41]

c. The Impact of Cognitive Biases and Tradecraft on Intelligence Community Analysis

As George notes, "cognitive bias is inherent to the 'cognition' process every analyst uses to examine an intelligence topic."[42] In the Intelligence Community culture, upon gaining experience, analysts develop patterns of thinking, otherwise known as mindsets, which are working models of how the object of analysis works. Analysts often tend to search for information consistent with or that may confirm existing agency

[39] Phillip Tetlock and Dan Gardner, *Superforecasting: The Art and Science of Prediction* (New York: Penguin Random House, 2016), 18.

[40] Tetlock and Gardner, 81–127.

[41] Tetlock and Gardner, 81–127.

[42] George, "Beyond Analytic Tradecraft," 298.

consensus, or they may select the most probable point of view that is comparatively easy to support, which is known in other contexts as confirmation bias. This does not mean that analysts ignore divergent analytic opinions; however, analysts most often include these as footnotes and usually reflect inter- rather than intra-organizational differences. This search for confirmation on the part of analysts does not necessarily result from a conscious decision; rather:

> It is the result of accepting an existing set of hypothesis, developing a mental model based on previous corporate products, and then trying to augment that model with current data in order to support the existing hypothesis.[43]

At the end of the day, cognitive biases can distort analysis due to mindsets and confirmation biases, among others, In other words, analysts may often discount or downgrade analyses and explanations that do not fit their pre-existing mindset or consensus.

d. The Impact of Extreme Time Constraints on Intelligence Analysis

Intelligence by its nature is time sensitive and perishable. In 2005, Johnston found that time is one of the greatest constraints faced by analysts. This constraint is coupled with the fact that the sheer volume of information (primarily open source but also gray information) that analysts need to integrate into analytic products results in time pressure on analysts that exceeds that of other intellectual endeavors. This reality is exacerbated by the fact that the timeframes of policy makers' decision cycles have become shorter and shorter; the extreme is 24 hours or less. This short lead time leads to the timeframes for analysis shrinking to support the decision cycle adequately. Other intellectually demanding analytic endeavors, such as work on business intelligence or in medicine, face similar time pressures.

When the extreme time pressure analysts face when combined with the consequences for the failure to deliver products on time supports the notion that time pressure in intelligence analytic endeavors is unique. It also results in informal and

[43] Johnston, *Analytic Culture in the U.S. Intelligence Community*, 25.

formal cultural understandings of its reality and how to cope with that reality incorporating into intelligence culture. Indeed, this cultural reality is also driven by changes in the intelligence environment that have resulted in a shift toward short-term issues or problem solving.[44] Both of these factors drive a shift in intelligence analysis resulting in a focus on short-term, tractable problems, and to lessened validity testing to the detriment of longer-term, well-tested, and nuanced analysis.

e. The Impact of the Relentless Focus on Current Production on Intelligence Analysis

The contraction of policy makers' decision cycle, coupled with a huge increase in demand for Intelligence Community products, has resulted in a cultural emphasis on current intelligence production to the detriment of longer term or strategic analytic products. This phenomenon affects both groups' interactions and the analytic process. Groups are often so focused on generating product that validity testing of the group product is less than robust, and the opinion of a single or of a few experts dominates (it is easier to agree to be able to return to individual tasks). In terms of the analytic process, useful techniques, such as Bayesian analysis, scenario development, red teams, simulations, competing hypotheses etc., are superficially applied or not applied at all because of the relentless need to generate analytic product relevant to policy maker's decision cycle. Analysis of the medium- to long-term behavior of any object of intelligence analysis therefore gets the short shrift. As Tyakoff says, "intelligence agencies [are] preoccupied with quantity rather than the quality of finished intelligence."[45]

f. The Impact of Rewards and Incentives on Intelligence Analysis

According to Johnston, analysts' rewards and incentives, namely opportunities for promotion, are directly tied to the amount of analytic product a given analyst produces. In

---

[44] Johnston, 18–19.

[45] Alex Tyakoff, "Counter Terrorism and Systems Dynamics: Modeling Organizational Learning in Postmodern Terrorist Groups," in *Terrorism and Global Insecurity: A Multidisciplinary Perspective*, ed. Klint Alexander (Chicago, IL: Linton Atlantic, 2009), 179–192, quoted in Maras, "Overcoming the Intelligence-sharing Paradox."

addition, rewards and incentives most often accrue to and are a function of the analysts social capital (in terms of their peers and their reputation with policy makers), as well as the level of their influence within the Intelligence Community.[46] Not ultimately a bad thing when assuming that social capital and influence derive from analytic excellence. However, the seeming lack of rigorous backward looking (quantitative not qualitative) analysis of the accuracy and relevance of analytic work product is a significant weakness, as analysts' rewards and incentives are not tied to analytic accuracy and relevance.

g. The Impact of Norms and Taboos and Secrecy on Intelligence Analysis

Norms and taboos are also essential features of the Intelligence Community culture. First among these is the cultural taboo against taking action that goes against the maintenance of the current set of institutional judgments. According to Johnston, "Once any intelligence agency has given its official opinion to policy makers, there exists a taboo about reversing or significantly changing the official or corporate position in order to avoid the loss of status, trust or respect."[47] This tendency is reinforced by perceptions of policy makers, or the perception that changing the official line, even when such actions result from new information, is a manifestation of incompetence or poor performance on the part of the agency. For the agency, the threat of loss of status, funding, or access also accompanies this scenario. Additionally, this threat also directly leads to a cultural norm that requires that the agency's analytic products be decisive regardless of circumstance—as opposed to nuanced, academic, and contradictory products—and results in analysts reworking analysis to be consistent with the requirements of this norm. Another cultural norm relates to the level of secrecy associated with inputs to the analysis. Analysts in the Intelligence Community perceive secret data collected by covert means to have a much greater analytic value than open source or "gray" information. The analysts test the validity of their cognitive model with secret information and use open source or gray information to fill gaps or provide context; indeed, the understanding is more covert information used in the analysis the better. According to the Commission on the

[46] Maras, "Overcoming the Intelligence-sharing Paradox," 16–17.

[47] Maras, 29.

Intelligence Capabilities of the United States Regarding Weapons of Mass Destruction, "Regrettably, all too frequently… 'non-secret' sources are undervalued and underused by the Intelligence Community."[48] These cultural factors have the following consequences for intelligence analysts and their analysis: (1) changing, reversing, or otherwise straying from the agency position, regardless of new or even contradictory information, is highly discouraged, (2) analytic products are generally decisive in nature regardless of whether that decisiveness is justified by circumstance, and (3) non-secret sources of information are systematically undervalued.

h. The Impact of the Analysts' Identity and Training on Intelligence Analysis

Johnston found that analysts' identities revolve around the organization's function or around their own education and background as opposed to revolving around a coherent intelligence analytic culture that treats intelligence analysis as a unique professional endeavor. He also found that their professional identity is more associated with reportage as opposed to being associated with analysis.[49] This association is driven by the perceived shift from medium- to long-term analysis to short-term, tactical analytic efforts. The implication is that analysts in the Intelligence Community lack a coherent commonly held professional identity. This lack of identity, and thus a common frame, has negative implications for group cohesion, inter- and intra-agency interaction and relationships.

In theory, developing intelligence analysts' skills for making forecasts requires that the analysts engage in high levels of effort, gain rewards for experience, and engage

---

[48] Commission on the Intelligence Capabilities of the United States Regarding Weapons of Mass Destruction, *Report to the President* (Washington, DC: Commission on the Intelligence Capabilities of the United States Regarding Weapons of Mass Destruction, 2005), 395. Here after, Commission, *Report to the President*.

[49] Commission, 395.

in organized training over time.[50] Many agencies provide basic and advanced training but usually do so independently. Intelligence Community-wide training opportunities are limited, which leads to fragmentation in terms of analytic approaches by analysts in different agencies. Although norms and standards for analysis do exist within the Intelligence Community, the details of how agencies operationalize these norms and standards vary from agency to agency.[51]

The implications of these cultural factors for intelligence and intelligence analysis include: (1) an emphasis on reportage rather than analysis in Intelligence Community products, (2) a lack of a coherent approach to intelligence analysis driven by fragmented training of analysts and inconsistent implementation of intelligence analytic norms and standards, and (3) fragmentation in analytic approaches within particular stovepipes within an agency (technical, tactical, operational, etc.) to the detriment of broader, integrative skill sets. All this fragmentation also means that analysts from different agencies often have difficulty finding, communicating with, and otherwise interacting with analysts outside their parochial purview, which results in all the ensuing consequences in terms of a lack analytic cohesion and conflict during the interagency process.

B. WHY COMBINE FORECASTS?

Although agencies in the Intelligence Community do not seem to combine forecasts across agencies, a number of sources in the literature favor this practice because it increases accuracy. For example, according to Brown and Murphy, "Combining forecasts can improve forecasting performance when one set of forecasts contains

---

[50] Marc Alpert and Howard Raiffa, "A Progress Report on the Training of Probability Assessors," in *Judgment under Uncertainty: Heuristics and Biases*, ed. David Kahneman, Paul Slovic, and Amos Tversky (New York: Cambridge University Press, 1982), 294–305; Manpreet K. Dhami et al., "Improving Intelligence Analysis with Decision Science," *Perspectives in Psychological Science* 10, no. 6 (2015): 753–757.

[51] Office of the Director of National Intelligence, *Analytic Standards*, Intelligence Community Directive 203 (Washington, DC: Office of the Director of National Intelligence, 2015), 1–5.

information…not contained in the other set of forecasts."[52] Similarly in *Superforecasters*, Tetlock and Gardner note forecasts about future events based on combining the forecasts of the most accurate lay predictors turned out to be more accurate than those of trained Intelligence Community analysts.[53] The literature on improving forecasting suggests that the forecast accuracy of the Intelligence Community work product can be significantly improved by combining independent forecasts. The next section discusses empirical demonstrations of the increase in accuracy engendered by combining independent forecasts, as well as the implications of using combined forecasts.

Armstrong summarizes the results of extensive empirical research on combining forecasts (he includes all types of forecasts, including those for natural events, in his studies) and builds on Clemen's seminal work, which reviewed 209 papers on this topic, by reviewing an additional 57 relevant empirical studies.[54] As Armstrong describes:

> Compared to the typical component forecast, the combined forecast is never less accurate. Usually it is much more accurate, with error reductions in the MAPE [mean absolute percentage error, also known as mean absolute percentage deviation, is a measure of prediction accuracy of a forecasting method in statistics] running over 12 percent for the 30 comparisons reviewed. Under ideal conditions (high uncertainty and combining many valid forecasts), the error reductions sometimes exceeded 20%. Also under ideal conditions, the combined forecasts were often more accurate than the best of the components. In short, the combined forecast can be better than the best but no worse than the average.[55]

The appendix contains a table providing a summary of the mean error reductions due to combing forecasts across 30 studies that Armstrong reviewed. Armstrong goes on to provide "rules of the road" for combing forecasts, namely:

---

[52] Barbara G. Brown and Allen H. Murphy, "Improving Forecasting Performance by Combining Forecasts: The Example of Road-surface Temperature Forecasts," *Meteorological Applications* 3, no. 3 (1996): 257–265, doi: 10.1002/met.5060030307.

[53] Tetlock and Gardner, *Superforecasting*, 81–104.

[54] Robert T. Clemen, "Combining Forecasts: A Review and Annotated Bibliography," *International Journal of Forecasting* 5 (1989): 559–583; J. Scott Armstrong, "Combining Forecasts," in *Principles of Forecasting: A Handbook for Researchers and Practitioners*, ed. J. Scott Armstrong (Norwell, MA: Kluwer Academic Publishing, 2001), 417–439, http://repository.upenn.edu/marketing_papers/34.

[55] Armstrong, 15.

- key principles for combining forecasts are to use
- different methods or data or both
- forecasts from at least five methods when possible;
- formal procedures for combining, which are
    - equal weights when facing high uncertainty
    - trimmed means
    - weights based on evidence of prior accuracy
    - weights based on track records, if the evidence is strong, and weights based on good domain knowledge[56]

Combining forecasts is most useful with:

- uncertainty as to the selection of the most accurate forecasting method
- uncertainty associated with the forecasting situation.
- high cost for large forecast errors[57]

More recently, Graefe et al. have found that combining forecasts based on many types of underlying data significantly improved forecasts of how the share of the nationwide popular vote for president was distributed. They state, "Combining [forecasts] yielded error reductions ranging from 16 percent to 59 percent, compared to the average errors of the individual forecasts."[58] This suggestion is amplified by Rothschild in the context of election forecasts.[59]

---

56 Armstrong, 15.

57 Armstrong, 15.

58 Andreas Graefe et al., "Combining Forecasts: An Application to Elections," *International Journal of Forecasting* 30, no. 1 (2014): 43, https://doi.org/10.1016/j.ijforecast.2013.02.005.

59 David Rothschild, "Forecasting Elections Comparing Prediction Markets, Polls, and Their Biases," *Public Opinion Quarterly* 73, no. 5 (2009): 895–916.

The literature strongly suggests that combining independent forecasts generally improves forecast accuracy. This thesis proposes combining forecasts from two independent sources, namely crowd based forecasts and prediction markets based forecasts, to improve intelligence forecast accuracy. The independent forecasts can be combined in a manner consistent with the principles for improving forecast accuracy discussed previously.

C. PREDICTION MARKETS BASICS

Friedrich Hayek elucidated the theory behind prediction markets in his 1945 study on the use of knowledge in society.[60] The following works elaborate on arguments for the utility and accuracy of prediction markets when making forecasts. Refer to Surowiecki's 2004 book, *The Wisdom of Crowds*, Sunstein's 2006 book, *Infotopia*, and Hubbard's 2014 book, *How to Measure Anything: Finding the Value of Intangibles*.[61] The efficient markets hypothesis (EMH) forms the theoretical basis for the demonstrated ability of the prediction market to make accurate predictions. According to the EMH, in a financial market, asset prices fully reflect all publicly available information and instantly change to reflect new public information.[62] Furthermore, as Fama noted in a 1969 article, "the EMH claims that asset prices reflect even hidden 'insider' information."[63]

Since they provide a mechanism to put a price on an outcome (asset), prediction markets are analogous to financial markets. A prediction market can be defined as an exchange-traded market in which participants buy and sell assets that embody the outcome of events. The evolution of the price of the asset until the event actually occurs, or when the asset contract expires, reflects the instantaneous likelihood of the event

---

[60] Friedrich A. Hayek, "The Use of Knowledge in Society," *American Economic Review* XXXV, no. 4 (1945): 519–530, http://www.econlib.org/library/Essays/hykKnw1.html.

[61] James Surowiecki, *The Wisdom of Crowds* (New York: Random House, 2005); Cass R. Sunstein, *Infotopia: How Many Minds Produce Knowledge* (Oxford: Oxford University Press, 2006); Douglas W. Hubbard, *How to Measure Anything: Finding the Value of Intangibles*, 3rd ed. (Hoboken, NJ: John Wiley and Sons, 2014).

[62] Common knowledge in the field of finance.

[63] Eugene F. Fama, "Efficient Capital Markets: A Review of Theory and Empirical Work," *Journal of Finance, Papers and Proceedings of the Twenty Eighth Annual Meeting of the American Finance Association* 25, no. 2 (1969): 383–417.

occurring as assessed at that point in time. Prediction markets participants are in essence buying an asset whose price varies between $0.00 (no likelihood of the event occurring) to $1 (the event is a certainty; a 100 percent chance of the event occurring) at expiry (when it is possible to determine if the event has occurred or not). As participants buy and sell the asset over time, its price varies as a function of supply and demand. The evolution of the price of the asset before expiry reflects the instantaneous likelihood of the event occurring as assessed at that point in time. Prediction market participants can also short the asset, essentially making a bet that the price of the asset will decrease. Shorting involves the sale of a security not owned by the seller but is promised to be delivered in the future. In other words, the sellers are making a bet that the price of the asset will go down because they will be able to obtain the stock at a future date at a lower price so as to fulfill the delivery contract and make a profit. If the price of the asset goes up in the future instead, then the short seller will fulfill the contract at a loss. The opposite of shorting is going long; purchasing the asset in the hope that the value of the stock will go up in the future. The success or failure of the participants in predicting whether the event has occurred is given by the value of the participant's prediction market portfolio at expiry.

D. PREDICTION MARKETS AND FORECAST ACCURACY

The primary reason for the choice of prediction markets as one of the techniques for improving intelligence analysis is the potential for improved forecast accuracy. Indeed, in the past five years, the Intelligence Community has started to explore the potential of prediction markets-based techniques to improve its understanding of the timing, type, and qualitative and quantitative characteristics of events of interest to policy makers; chiefly, intelligence analysis. The body of literature supports the assertion that prediction market-based forecasts are highly accurate in a variety of contexts, such as higher than polls and other techniques. Prediction markets did fail to predict the election of President Trump, as well as the vote affirming the exit of Britain from the European Union (Brexit); however, contemporaneous news accounts do provide plausible explanations for these failures.

Studies on the accuracy of prediction market forecasts under different scenarios have found prediction markets make accurate forecasts of events under a wide variety of conditions.[64] The literature evaluating the accuracy of prediction markets assumes that each individual participating in the prediction market is unbiased and makes forecasts independent of the forecasts of other market participants. Under these assumptions, the aggregated forecast of the group of market participants is uncontroversially better on average than the forecasts made by each individual.[65] However, the limits of the crowd's participating in prediction markets' wisdom are comparatively poorly understood. For example, when group members can compare their predictions to those of other group members, positive correlations between predictions may be expected, which can result in a decline in the group's predictive performance; however, research has shown this assumption is not the case.[66]

Consider the case of sports betting as a prediction market. In sports betting, individuals seem to make systematically biased predictions, and thus, their predictions are strongly correlated. As a result, the accuracy of sports bettors' forecasts in the aggregate can be expected to be reduced compared with those made using other techniques. That is, their predictions are less wise.[67] Dana and Broomell analyze the robustness of crowd wisdom in the face of varying factors, such as bias and diversity or lack thereof, and they find that "a group is wisest, all things equal, when it is maximally 'diverse' in that its members' forecasts are as negatively correlated as possible."[68] Dana and Broomell also state that wise groups should include some members who are better

[64] Arrow et al., "The Promise of Prediction Markets," 877–878; Berg, Nelson, and Rietz, "Prediction Market Accuracy in the Long Run," 285–300.

[65] Armstrong, "Combining Forecasts," 417–439; Clemen, "Combining Forecasts," 559–583; Robert I. Winkler, "Probabilistic Prediction: Some Experimental Results," *Journal of the American Statistical Association* 66, no. 336 (1971): 675–685.

[66] Jan Lorenz et al., "How Social Influence Can Undermine the Wisdom of Crowd Effect," *Proceedings of the National Academy of Sciences* 108, no. 22 (2001): 9020–9025.

[67] Joseph P. Simmons et al., "Intuitive Biases in Choice versus Estimation: Implications for the Wisdom of Crowds," *Journal of Consumer Research* 38, no. 1 (June 2011): 1–15.

[68] Clintin P. Davis-Stober et al., "When Is a Crowd Wise?," *Decision* 1, no. 2 (2014): 79–101.

predictors than the norm and that a diversity of market participants' perspectives is critical to the market's predictive accuracy.[69]

Studies of the accuracy of prediction markets under different conditions appear extensively in the literature, and they have been found to be quite accurate in predicting events in a wide variety of situations.[70] For example, according to Lin, Tung, and Yeh in a 2013 article, "prediction markets have been proven empirically to be remarkably accurate in forecasting future events with a lower prediction error than conventional forecasting methods ex post."[71] Additionally, Berg, Nelson, and Rietz found that when comparing the predictions of polls with those of prediction markets in the U.S. presidential elections from 1998 to 2004, the predictions of prediction markets were closer to the eventual outcome than traditional polls 74 percent of the time.[72] Williams and Reade support this assertion when they determined that they could "conclude that prediction markets appear to provide the most precise forecasts" when compared to polls, expert opinion, and statistical modeling.[73]

Rajakovich and Vladimirov found another example of the effectiveness of prediction markets; they found that when predicting the number of admissions in a health care setting, the prediction of the market participants was 1,158 admissions while the actual number of admittances was 1,154, an error of only 0.3 percent.[74] In a pilot study using prediction markets for forecasting influenza activities in Iowa, North Carolina, and Nebraska in the 2008–2009 and 2009–2010 influenza seasons, Ho, Polgreen, and Prendergast found, "prediction markets achieved high level of forecasting accuracy,

[69] Davis-Stober et al., 79–101.

[70] Arrow et al., "The Promise of Prediction Markets," 877–878"; Berg, Nelson, and Rietz, "Prediction Market Accuracy," 285–300.

[71] Hung-Wen Lin, Chen Yuan Tung, and Jason Yeh, "Multivariate Methods in Assessing the Accuracy of Prediction Markets Ex Ante Based on the Highest Price Criterion," *The Journal of Prediction Markets* 7, no. 3 (2013): 30.

[72] Berg, Nelson, and Rietz, "Prediction Market Accuracy," 285–300.

[73] L. Vaughn Williams and James J. Read, "Forecasting Elections," *Journal of Forecasting* 35, no. 4 (2016): 308–328, doi: 10.1002/for.2377.

[74] David Rajakovich and Vladimir Vladimirov, "Prediction Markets as a Medical Forecasting Tool: Demand for Hospital Service," *Journal of Prediction Markets* 3, no. 2 (2009):78–106.

provide a flexible and effective way to aggregate both objective and subjective information about seasonal influenza."[75] In 2014, Arneson and Bergfjord found that prediction markets outperformed the polls in predicting the outcomes of the 2008 and 2012 U.S. elections.[76] In 2009, Berg, Neuman, and Rietz used a prediction market to estimate Google's initial public offering (IPO) price and found that the prediction market results accurately tracked both the level of IPO oversubscription and Google's first day market capitalization.[77]

According to Slamka, Skiera, and Spann, "Prediction market accuracy depends on its market design, including the choice of market mechanism."[78] Additionally, Gaspoz provides key and comprehensive information on the various factors to consider when designing prediction markets that includes the details of alternative incentive mechanisms, trading processes, clearinghouse parameters, and participant management options.[79] Li, Chen-Yuan, and Chang summarize the impact of design factors on prediction market accuracy, as found in the literature when they explain that:

> Some scholars (e.g., Berg et al., 1997; Gruca et al., 2005) assert, based on trading data of Iowa electronic markets (IEMs), that number of contracts (degree of competition), trading volume and bid-ask price spread are the most important factors. Others (e.g., Forsythe et al., 1999; Oliven and Rietz, 2004) find that number of marginal traders is the major factor for prediction accuracy. Kambil and Heck (2002) and Ledyard (2006) advocate that major factors include large number of traders, sufficient

[75] Anson T. Y. Ho, Phillip M. Polgreen, and Thomas Prendergast, "Prediction Market for Disease Surveillance, a Case Study of Influenza Activity," *Journal of Prediction Markets* 10, no. 1 (2016): 68–82.

[76] Sveinung Arneson and Ole Bergford, "Prediction Markets versus Polls: An Examination of Accuracy for the 2008 and 2012 Elections," *Journal of Prediction Markets* 8, no. 3 (2014): 24–33.

[77] Joyce E. Berg, George R. Neumann, and Thomas A. Reitz, "Searching for Google's Value: Using Prediction Markets to Forecast Market Capitalization Prior to an Initial Public Offering," *Management Science* 55, no. 3 (2009): 348–361.

[78] Christian Slamka, Bernd Skiera, and Martin Spann, "Prediction Market Performance and Market Liquidity: A Comparison of Automated Market Makers," *IEEE Transactions on Engineering Management* 60, no. 1 (2013): 169–185.

[79] Cederic Gaspoz, *Prediction Markets Supporting Technology Assessment* (n.p., Printed in the World, 2011), 57–110.

information as well as incentives for traders to reveal effective information.[80]

Thus far, public prediction markets have been considered. Corporations have used and also currently use private prediction markets. By overarching objective, these private markets include: (1) forecasting markets, (2) markets that revolve around idea genesis and evaluation, and (3) markets that address the problem of innovation by matching research and development problems with researchers and peer-to-peer assistance.

In 2007, Gruca and Berg showed how private prediction markets could be used to tap into private information and unstated knowledge held by stakeholders, such as employees, customers, vendors, etc.[81] In this vein, data from Google concerning its corporate prediction markets suggests that event probabilities predicted by its markets closely approximated actual event probabilities.[82] This data is especially impressive when considering that it covered 2.5 years during which Google ran 270 prediction markets with over 1,400 participants.[83] Hewlett Packard attempted to use prediction markets to estimate future sales and found that the forecasts generated as a result were more accurate than those generated using traditional forecasting processes.[84] Comparably, Intel found that its internal prediction market forecasts were at a minimum as accurate as its official forecasts produced using conventional methods, and in some cases, they were as much as 20 percent more accurate.[85] Davis used an internal prediction market to estimate the cost and schedule performance of Department of

---

[80] Eldon Y. Li, Tung Chen-Yuan, and Shu-Hsun Chang, "User Adoption of Wisdom of Crowd: Usage and Performance of Prediction Market System," *International Journal of Electronic Business* 12, no. 2 (2015): 189.

[81] Thomas S. Gruca and Joyce E. Berg, "Public Information Bias and Prediction Market Accuracy," *Journal of Prediction Markets* 1, no. 3 (2007): 219–231.

[82] Bo Cowgill, Justin Wolfers, and Eric Zitzewitz, "Using Prediction Markets to Track Information Flows: Evidence from Google," in *Auctions, Market Mechanisms and Their Applications: First International ICST Conference, AMMA*, vol. 14, ed. Sanmay Das et al. (Boston, MA: Springer, 2009), 13.

[83] Markus Noeth et al., "Information Aggregation in Experimental Asset Markets: Traps and Misaligned Beliefs" (working paper 1060, California Institute of Technology, Pasadena, CA, 1999), 4–5.

[84] Kay-Yut Chen and Charles R. Plott, *Prediction Markets and Information Aggregation Mechanisms: Experiments and Applications* (Pasadena, CA: California Institute of Technology, 1998), 17.

[85] Jay Hopman, "Using Forecasting Markets to Manage Demand Risks," *Intel Technology Journal* 11, no. 2 (2007): 126–136.

Defense acquisition programs (each program cost and schedule estimate was an asset traded in the prediction market).[86] In a 2011 article, Davis states, "The market was open for 117 days. Within two weeks of opening, on average, the market converged to the right answer [correct estimate of cost and/or schedule slip] for nine [out of 10] assets."[87] Similarly, Buckley reports that some other organizations that have used prediction markets to aid in decision-making include Motorola, Qualcomm, InfoWorld, MGM, Chiron Corporation, TNT, EA Games, Yahoo, Corning, MasterFoods, Pfizer, Abbott, Chrysler, General Mills, and O'Reilly Media.[88]

Dissenting voices about the predictive superiority of prediction markets in making forecasts include Graefe et al., Sjoberg, and Teschner and Weinhardt. Graefe et al. discovered, "prediction markets provided little additional value compared to a simple average of forecasts" when performing a simple quantitative judgment task.[89] Sjoberg looked at multiple different groups of forecasters and forecasts for Swedish elections and did not find evidence of prediction markets generating superior forecasts.[90] Additionally, Teschner and Weinhardt looked at multiple studies on the use of comparing prediction markets to surveys and polls and found that their review "suggests that the relative performance advantage of markets may be small compared to surveys or polls."[91]

Furthermore, recent prediction market failures, such as the failure to predict the election of Donald Trump and Brexit accurately, have thrown the claims of prediction market evangelists in doubt. A 2016 article by Kominers in *Bloomberg View* provides a

---

[86] Danny M. Davis, "Designing a Viable Prediction Market to Forecast Defense Acquisition Cost and Schedule Outcomes," *Defence and Peace Economics* 22, no. 3 (2011): 351–366, doi: 10.1080/10242694.2010.491680.

[87] Davis, 358.

[88] Patrick Buckly, "Harnessing the Wisdom of Crowds: Decision Spaces for Prediction Markets," *Business Horizons* 59, no. 1 (2016): 85–84.

[89] Andreas Graefe and J. Scott Armstrong, "Comparing Face-to-Face Meetings, Nominal Groups, Delphi and Prediction Markets on an Estimation Task," *International Journal of Forecasting* 27, no. 1 (2011): 183–195, http://dx.doi.org/10.1016/j.ijforecast.2010.05.004.

[90] Lennart Sjöberg, "Are All Crowds Equally Wise? A Comparison of Political Election Forecasts by Experts and the Public," *Journal of Forecasting* 28, no. 1 (2009): 1–18.

[91] Florian Teschner and Christof Weinhardt, "A Macroeconomic Forecasting Market," *Journal of Business Economics* 85 (2015): 299, doi: 10.1007/s11573-014-0741-5.

plausible explanation for these failures.[92] In most prediction markets, Trump was consistently trading below 35 cents in the month prior to the election with an average daily closing price around 25 cents, which suggests a 25 percent probability of victory. However, he still had a one in four chance of winning, so his win while improbable, should have happened on average one in four times. For Brexit, the odds of a yes vote were about at around three in 10 in the major prediction markets, so the joint probability of both a Trump victory and Brexit was likely around 7.5 percent. Yet, both happened. An explanation may be that most people betting on prediction markets do not have much contact with the people who voted for Trump and Brexit. If so, no prediction market is likely to give accurate results. If all the traders in the relevant prediction markets are missing a key piece of information, then the market price (remember the conditions for EMH) are likely missing it as well. Even if the market worked as designed, traders leaning toward Trump or Brexit may *not* have been participating in the market. Thus, it is likely none of the market participants had decent information on the scale of Trump's or Brexit's support, and all the trading in the world could not lead to a price that correctly reflected his chance of victory. In his 2016 *Bloomberg View* article, Kominers comments:

> This problem is compounded by the fact that prediction market participants also infer information from the prevailing price—and so may have discounted the signals of Trump's strength that they did receive. Also, total payouts from prediction markets are too low to create a strong incentive for participants to work really hard to become substantially better-informed. This chain of logic suggests that prediction markets could be abnormally bad at forecasting events that will be decided by actions of people who aren't themselves plugged in to prediction markets. And there's a message here about markets more broadly: Even the best-functioning markets don't do a good job of pricing when key players aren't represented.[93]

Finally, the issue of manipulation of prediction markets must be addressed. According to Teschner and Weinhardt, "three types of manipulation [may exist]: action-based (changing the underlying fundamentals), information-based (spreading false

[92] Scott Duke Kominers, "Prediction Markets Didn't Call Trump's Win, Either," *Bloomberg View,* November 15, 2016, https://www.bloomberg.com/view/articles/2016-11-15/prediction-markets-didn-t-call-trump-s-win-either.

[93] Kominers.

information)"and trade-based (buying, selling of shares)."[94] In all these cases, it can be argued that the potential offending manipulator is just another trader. If traders other than the manipulator take advantage of the likely ability to profit from the offending trader's actions, counterintuitively market accuracy may increase. The prevailing opinion in the literature is summarized by Deck, Lin, and Porter in their 2013 review of studies on manipulating prediction markets in which they state, "Research suggests prediction markets are robust to manipulation attacks."[95]

The literature also contains good descriptions of the operation and theoretical basis for prediction markets, as well as evidence and analysis comparing and evaluating alternative prediction market designs. Most sources suggest the centrality of prediction market design to its forecasting accuracy. These design issues include "the choice of participants, the specification of the contracts traded in a prediction market, the trading mechanism, and the incentives provided to ensure information revelation, trader pool size, market termination timing, decision heuristics, market context, and uncertainty."[96] Finally, the literature does include some evidence of the Intelligence Community, including the IARPA Aggregative Contingent Estimating ACE program, using prediction markets in forecasting.[97]

As the literature demonstrates, prediction markets have greater forecasting prowess than other forecasting techniques. Although prediction markets have failed to live up to their promise of increased forecast accuracy in some instances, these failures most likely resulted due to specific design and participation factors unique to the particular markets in question. Prediction markets, if properly designed and implemented, can be a useful, practical part of the intelligence analysts' analytic toolkit.

---

[94] Teschner and Weinhardt, "A Macroeconomic Forecasting Market," 293–317.

[95] Cary Deck, Lin Shengle, and David Porter, "Affecting Policy by Manipulating Prediction Markets: Experimental Evidence," *Journal of Economic Behavior and Organization* 85 (2013): 48–62.

[96] Patrick McHugh and Aaron Jackson, "Prediction Market Accuracy: The Impact of Size, Incentives, Context, and Interpretation," *Journal of Prediction Markets* 6, no. 2 (2012): 22–46.

[97] "Aggregative Contingent Estimation (ACE)," Intelligence Advanced Research Projects Agency, accessed April 29, 2018, https://www.iarpa.gov/index.php/research-programs/ace/baa.

A review of the literature provides the basis for asserting that prediction markets may offer increased forecast accuracy. Indeed, multiple sources suggest that prediction markets significantly outperform more traditional forecasting techniques when concerning forecast accuracy. Furthermore, the salience of prediction markets design, levels, and types of participation to predictions markets forecasting success is clear. The literature includes multiple avenues for further inquiry, as well as multiple case studies of prediction markets-based solutions to real-world forecasting problems in business, politics, marketing, and funding innovation. Studies examining the effectiveness of prediction markets in solving intelligence problems are limited in scope and number, however.

E. CROWDSOURCING BASICS

The definition of crowdsourcing depends on who describes it. For instance, Doan, Ramakrishnan, and Halevy describe crowdsourcing as a system that "enlists a crowd of humans to help solve a problem defined by the system owners."[98] Nakatsu, Grossman, and Iacovu view certain types of peer production, such as open-source software development, as forms of crowdsourcing. They posit, "crowdsourcing is outsourcing to an undefined, anonymous group of people who come together to solve problems and perform tasks once performed by a company's employees."[99] Nakatsu, Grossman, and Iacovu expand their definition and define crowdsourcing as a four-step process in which:

- A requestor (either an individual or organization) identifies a specific task to be performed or problem to be solved.
- The requestor broadcasts the task or problem online.
- The crowd performs the task or solves the problem.
- Depending on the nature of the task, the requestor either

[98] An Hai Doan, Raghu Ramakrishnan, and Alon Y. Halevy, "Crowdsourcing Systems on the World-wide Web," *Communications of the ACM* 54, no. 4 (2011): 87.

[99] Robbie T. Nakatsu, Elissa B. Grossman, and Charalambos L. Iacovu, "A Taxonomy of Crowdsourcing Based on Task Complexity," *Journal of Information Science* 60, no. 6 (2014): 825.

- Sifts through the solutions and selects the best solution (selective crowdsourcing).
- Aggregates or synthesizes the crowd's submissions in a meaningful way (integrative crowdsourcing).[100]

In yet another definition of crowdsourcing, Morschheuser et al. explain "crowdsourcing can be seen as an online, distributed problem-solving approach that transforms problems and tasks into solutions by harnessing the potential of large groups of crowdsourcees via the Web rather than traditional employees or suppliers."[101] The importance and applicability of crowdsourcing as a technique to solve a wide variety of problems can be inferred from the fact that, according to Morschheuser et al., the industry portal crowdurcing.org provides access to information on almost 3,000 crowdsourcing efforts.[102]

Crowdsourcing systems come in four broad categories based on the characteristics of the crowdsourced work being performed, as Morschheuser et al. define:

- Crowdprocessing systems focus on using the crowd to perform large quantities of identical tasks. Crowdprocessing is the objective of Amazon's Mechanical Turk or of Galaxy Zoo.
- Crowdsolving involves using the crowd's diversity to find a huge number of solutions to a given problem, each with different characteristics. Crowdsolving is often used for extremely computationally intractable problems like protein folding or if the problem has no theoretical solution.
- Crowdrating seeks to harness crowd wisdom to perform collective assessments or predictions.
- Crowdcreating uses crowds to find solutions to create new things based on a variety of contributions that are different in scope or type. Examples of

100 Nakatsu, Grossman, and Iacovu," 823–834.

101 Benedikt Morschheuser et al., "Gamified Crowdsourcing: Conceptualization, Literature Review, and Future Agenda," *International Journal of Human-Computer Studies* 106 (October 2017): 26–43.

102 Morschheuser et al., 26–43.

> crowdcreating include all user generated content on sites like YouTube, *Wikipedia*, and the like.[103]

While crowdsolving and crowdrating can be used to provide information about complex problems to decision makers, this thesis focuses on crowdsolving and crowdrating. Simon suggests that the decision maker goes through three steps before coming to a decision, all of which the crowd can address. These steps are:

> Intelligence (information gathering and sharing for the purpose of problem solving or opportunity exploitation, problem identification, and the determination of the problem's importance), design (generating ideas and alternative solutions), and choice (evaluating the generated alternatives and then recommending or selecting the best course of action).[104]

In the intelligence phase of a crowdsourced effort, crowdsourcing can help search for, discover, and aggregate information, gather opinions, make predictions, and accumulate knowledge. In the design phase, crowds can solicit and elicit ideas from employees, customers, and other stakeholders and generate ideas. Simply, crowdsourcees can help generate and evaluate alternatives.

Stottelemyre states that crowdsourced intelligence requires that (1) someone acts on the behalf of a national security organization, (2) someone acquires, not collects, the relevant information, (3) a national security organization receives the information, and (4) the intelligence questions are asked directly of a group of potential sources.[105]

F. CROWDSOURCING AND FORECAST ACCURACY

The literature suggests that crowdsourcing of intelligence may lead to improvements in forecast accuracy. Indeed, in the past five years, the Intelligence Community has started to explore the potential of crowd-based techniques to improve its

---

[103] Morschheuser et al., 26–43.

[104] Herbert A. Simon, *Decision Making and Problem Solving, Research Briefings 1986: Report of the Research Briefing Panel on Decision Making and Problem Solving* (Washington, DC: National Academy Press, 1986), 42.

[105] Steven A. Stottlemyre, "HUMINT, OSINT, or Something New? Defining Crowdsourced Intelligence," *International Journal of Intelligence and CounterIntelligence* 28, no. 3 (2015): 578–589, doi: 10.1080/08850607.2015.992760.

intelligence analysis efforts. Both of these facts are the basis for the argument in this thesis that crowdsourcing intelligence be included as one of the techniques for improving intelligence analysis.

The literature makes a strong case that the crowds are indeed accurate, or "wise," based on the relative accuracy of crowd-based prediction models. According to Bagherpour, "the U.S. Intelligence Community has created more than a half-dozen forecasting programs over the last few years through its research unit, the Intelligence Advanced Research Projects Activity (IARPA)."[106] For example, one forecasting program run by IARPA features a tournament between hybrid teams made up of both humans and machines. It is designed to exploit evidence that the best forecasting results when predictions are generated by computer algorithms augmented with human guidance. This program is a follow on to an early attempt at crowdsourcing intelligence that was shut down in 2013, namely a program called FutureMap, which used a terrorism futures market in which participants placed bets on aspects of future terrorist acts. These were then aggregated to generate probability estimates for such acts. In part, this thesis is an attempt to examine whether and how intelligence can be crowdsourced, and if doing so may lead to improved forecasting by the Intelligence Community.

In his seminal 2004 book, *The Wisdom of Crowds*, Surowiecki states four conditions are a prerequisite for crowds to be wise: (1) opinion and backgrounds in the crowd must be diverse, (2) members of the crowd are independent in terms of how they arrive at their judgments, (3) the crowd is decentralized (people are able to specialize and draw on local knowledge), and (4) a mechanism aggregates crowd judgment.[107] Since Surowiecki published his book, a large amount of research has been conducted on why crowds are wise and on how to extract and apply the wisdom of crowds via crowdsourcing.

The literature contains several examples of crowdsourcing efforts aimed at solving intelligence community analytic problems to include the Good Judgment Project,

---

[106] Nathan B. Moncton, "U.S. Using Canadian Games to Improve Its Intel," *The Times*, July 3, 2017.

[107] Surowiecki, *The Wisdom of Crowds*, 1–106.

efforts to crowdsource in real-time analysis of the identities and motivations of the Boston bombers, and the work the *Arms Control Wonk* does in Syria and Iraq. A description and analysis of each of these efforts is provided in the following sections.

1. The Good Judgment Project

The most salient example of crowdsourcing intelligence is the Good Judgment Project, which is sponsored by the IARPA.[108] The project tests the ability of graduate students, faculty, and practitioners from the political science realm to forecast global geopolitical events.[109] Moreover, the project works by:

> Recruit[ing] over 2,000 forecasters ranging from graduate students to forecasting and political science faculty and practitioners. Each forecaster was randomly assigned to one of the three trainings (none, probability, or scenario training) and to one of the four different modes of information sharing (individual predictions in isolation, individual predictions seeing what others predict, a prediction market, or team predictions). Predictions were evaluated using the Brier scores…Brier scores for each problem on each day were averaged over all of the days the problem was open, and then the scores for all the problems were averaged. Individuals or, in the team setting, teams were encouraged to minimize their Brier score. No financial reward was given, but there was a "Leader Board" making public the most successful people... [The study] compared a variety of aggregation methods, looking at combinations of different:
>
> - weightings of forecasters based on their personality and expertise attributes, averaged either using a weighted mean or a weighted median
> - down-weightings of older forecasts using exponential decay
> - transformations of the aggregated forecasts to push them away from 0.5 and towards more extreme values[110]

Refer to Ungar et al.'s *The Good Judgment Project: A Large Scale Test of Different Methods of Combining Expert Predictions* that found that the most successful

[108] "About IARPA."

[109] "About IARPA."

[110] Ungar et al., *The Good Judgment Project.*

forecasters generally are characterized by significant levels of political knowledge and general intelligence, and they did not need access to classified material or more than modest training in probability and statistics to exceed the success rate of career intelligence analysts in predicting geopolitical events by over 30 percent.[111] This improved accuracy did not result from systemic failure on the part of the career intelligence analysts or from extraordinary abilities on the part of the citizen participants. Rather, it was the result of the impact of probability and statistics in action. The more forecasts are used to make a given prediction, the more likely that the mean of those forecasts will reflect reality better than individual forecasts or even small groups of forecasts. As Spiegel explains on a 2014 *NPR* segment:

> In other words, there are errors on every side of the mark, but there is a truth at the center that people are responding to, and if you average a large number of predictions together, the errors will end up canceling each other out, and you are left with a more accurate guess.[112]

Examining the results of the Good Judgment Project further, Mellars et al. found that the best forecasters benefitted from formal training in probability and statistics, worked in environments characterized by teamwork, and took their predictions seriously in that they spent significant amounts of time developing and updating their forecasts.[113] Commenting on the work of Ungar et al., Bisogno describes:

> Working in groups greatly improves prediction accuracy. The question of how to utilize the wisdom of the crowds…is more difficult to answer than whether or not that wisdom is valuable: "Although the 'wisdom of the crowds' and the power of predictive markets are widely recognized, it is less clear how to best make use of that wisdom."[114]
>
> An important dynamic the study observes is the risk of group-think when experts are able to discuss their predictions. While the study acknowledges

---

111 Ungar et al.

112 Alix Speigel, "So You Think You Are Smarter than a CIA Agent," *NPR*, April 2, 2014, https://www.npr.org/sections/parallels/2014/04/02/297839429/-so-you-think-youre-smarter-than-a-cia-agent.

113 Barbara Mellers et al., "The Psychology of Intelligence Analysis: Drivers of Prediction Accuracy in World Politics," *Journal of Experimental Psychology: Applied* 21, no. 1 (2015): 1, doi: 10.1037/xap0000040.

114 Ungar et al., *The Good Judgment Project*, 18.

the inverse is also possible—that better arguments can be formed this way—this thesis theorizes there may be a risk of groupthink in homeland security and government enterprises *unless* outside perspectives are considered due to cultural biases and organizational tendencies.[115]

2. The Boston Bombing-Crowdsourcing Gone Awry

The Boston Marathon bombing in 2013 provides an example of crowdsourcing that went awry, in part. According to Bisogno:

> [A] student was wrongly suspected as one of the bombers and became the victim of a "digital witch hunt." Before authorities had officially identified a suspect, independent websites posted the student's photo online, and he was subsequently followed by private citizens (not investigators).[116]

The news media unquestioningly reported the false identification without verifying the truthfulness of information coming from the independent parties online.[117] Subsequently, the media disseminated the identification to the nation in a way that suggested that the information was both validated and certain. Effectively, individuals online, bystanders, and the media performed a criminal investigation without the training to do so.[118] Moreover, they also effectively ignored legal considerations and legal constraints, as well as the rights of the accused and critical contextual information.[119] Tapia, LaLone, and Kim describe what the actions some mainstream online groups took, and they characterize the results of these actions as "dangerous and perhaps criminal."[120] As events unfolded, several other innocent individuals were characterized as suspects by these groups based on unvetted and unvalidated information.

---

[115] Tarun Wadhaw, "Lessons from Crowdsourcing the Bostom Bombing Investigation," *Forbes*, April 22, 2013, http://www.forbes.com/sites/tarunwadhwa/2013/04/22/lessons-from-crowdsourcing-the-bostonmarathon-bombings-investigation/#1416d38312b5, quoted in Raymond Bisogno, "Problem Solving in Homeland Security and Creating Policy Conditions for Enhanced Civic Engagement: An Examination of Crowdsourcing Models" (master's thesis, Naval Postgraduate School, 2017), 18.

[116] Bisogno, 31.

[117] Andrea H. Tapia, Nicolas LaLone, and Hyun-Woo Kim, "Run Amok: Group Crowd Participation in Identifying the Bomb and Bomber from the Boston Marathon Bombing," in *Proceedings of the 11th International ISCRAM Conference* (Rio de Janeiro, Brazil: Information Systems for Crisis Response and Management, 2014), 265–274.

[118] Tapia, LaLone, and Kim, 265–274.

[119] Tapia, LaLone, and Kim, 265–274.

[120] Tapia, LaLone, and Kim, 266.

Yet, this crowdsourced effort was somewhat successful. The information gathering effort, whereby the crowd's pictures and videos of the event provided to the authorities, was a resounding success, while the attempt to crowdsource the criminal investigation proved a dismal failure. Individuals are a good crowdsourcing resource in terms of both providing information and reporting events. However, crowds fail when it comes to crowdsourcing taking action. When only providing information to the authorities, individuals are effective at augmenting the resources of the authorities. According to Tapia LaLone, and Kim:

> Seattle's Police Department runs a program where citizens can receive tweets about and report when they spot stolen cars. German police have experimented with posting sketches of wanted criminals on Facebook FB +1.51%, where citizen's identifications have already led to several arrests. In another example, a Broward County Sheriff has leveraged his 10,000 Facebook friends to successfully track down stolen goods.[121]

3. Arms Control Wonk

Refer to Lewis' blog, which provides two additional examples of crowdsourcing in an intelligence context.[122] In 2011, *Arms Control Wonk*, a blogging community, analyzed imagery of what was purported to be a textile factory near the Syrian town of Al Hasaka. This alleged factory had attracted the attention of the International Atomic Energy Agency (IAEA) as a potential nuclear site. The bloggers evaluated open-source imagery (photos, video, and satellite images) of the area and also interpreted Arab-language media reports describing the facility. Based on this analytic work, *Arms Control Wonk* successfully determined that the facility was as a textile mill built with East German assistance decades earlier.[123]

Arms Control Wonk also used crowdsourcing to analyze four videos of the remains of a probable nuclear reactor undeclared to the IAEA at Al Kibar, also in Syria. The Syrian Opposition obtained these videos and posted them to YouTube. The reactor

---

[121] Tapia, LaLone, and Kim, 266.

[122] Jeffrey Lewis, "FSA Overruns Al Kibar," *Arms Control Wonk* (blog), February 25, 2011, http://www.armscontrolwonk.com/archive/206309/fsa-overruns-al-kibar/.

[123] Lewis.

site had been bombed by the Israeli Air Force. The videos showed details of the site that seemed to indicate the site's capture by the opposition. Participants in the crowdsourcing effort (including former IAEA official Olli Heinonen) confirmed that the videos, which were taken with mobile phones, were authentic and that the videos were consistent with what was known at the time about the general details of the site. On further examination, the videos allowed *Arms Control Wonk* to determine that a building on the site contained at least five stationary Scud-type missile launchers, which were designed to be fired through openings in the roof. In addition, *Arms Control Wonk* successfully confirmed the firing Scud-type rockets on cities in the northern parts of the country.[124]

A review of the literature provides the basis for a solid understanding of crowdsourcing. Indeed, multiple sources suggest alternative taxonomies of crowdsourcing types and examples of rules for effective crowdsourcing. The literature includes multiple avenues for further inquiry, as well as multiple case studies of crowdsourced solutions to real-world problems in business, the sciences, marketing, funding, and studies examining the effectiveness of crowdsourcing in solving intelligence problems. However, the literature includes few studies of the effectiveness of crowdsourcing for making predictions in an intelligence context as compared to alternative techniques.

G. CONCLUSION

The literature review suggests that the creation of a combined methodology based on prediction markets and crowdsourcing of analytic inputs that improves the accuracy of intelligence analysis and forecasting is possible. Furthermore, the literature review has identified characteristics of intelligence culture that negatively affect the accuracy of intelligence analysis and forecasts. The review suggests that a crowd and prediction markets-based methodology can address these sources of forecast inaccuracy.

Generally speaking, the literature review resulted in few surprises. However, three issues are of concern:

---

[124] Lewis.

- Comparative dearth of recent scholarly analysis of Intelligence Community culture is a concern in that much of the review of Intelligence Community culture in this thesis dates to the first half of the 2000s and may be out of date as circumstances may have changed.

- Literature search did not identify scholarly analyses of past applications of crowd- and prediction markets-based techniques by the Intelligence Community beyond a few instances, despite multiple expressions of general interest by the Intelligence Community, funding by the IARPA of the ACE program, and the existence of the Good Judgment Project as a pilot.

- Activity on scholarly work on prediction markets slowed significantly starting in 2013.

Each of these issues cited previously has implications for the thesis. In the case of the lack of recent scholarly research on Intelligence Community culture, the thesis proceeds from the assertions that: (1) Intelligence Community culture takes a very long time to change, and so although dated, the available scholarly research is relevant, and (2) the few examples of recent work in this area do not suggest significant changes in Intelligence Community culture that may falsify previous work in the field. As for the lack of past applications of prediction markets-based techniques by the Intelligence Community, given that the literature on the use of these techniques in other contexts is so rich, moving the thesis forward by reasoning by analogy is quite possible and indeed justified. As for activity on scholarly works on prediction markets slowing significantly since 2013, the key thing to consider is that the literature search unearthed comparatively little recent research falsifying the claims of increased forecast accuracy using the technique.

Based on the results of the Good Judgment Project, and on the results of applications of similar methodologies in other non-intelligence contexts, it is likely that intelligence assessments may be improved through the adroit application of crowdsourcing and prediction markets-based techniques to the problem of intelligence forecasting. Toward this end, the thesis analyzes practices in the application of crowd and prediction markets-based techniques to forecasting in other, non-intelligence contexts and uses that analysis as the basis for developing a crowd and prediction markets-based forecasting methodology for use by the Intelligence Community.

The next chapter discusses relevant practices in crowdsourcing drawn from multiple disciplines. These practices serve as a foundation for subsequent work on prediction markets and for the development of the intelligence analytic and forecast methodology that is the central aim of this thesis.

III. CROWDSOURCING PRACTICES

A. INTRODUCTION

Chapter II introduced prediction markets and crowd-based techniques for performing analysis and making forecasts, and it discussed these techniques in the context of intelligence analysis and forecasting. Chapter II also introduced intelligence culture and examined the implications of intelligence culture for analytic and forecast accuracy. This chapter is a more detailed discussion of relevant practices in crowd-based problem solving, including the solution of analytic and forecasting-based problems, and it covers practices used in the design of crowd sourced problem solving efforts drawn from multiple disciplinary contexts. Given the primary goal of the thesis to develop an effective crowd-based forecasting and analytic methodology, it is necessarily to select and apply known practices in the design of crowdsourcing efforts as a precursor. Areas of interest are those practices central to the effectiveness of the crowdsourcing effort and include:

- practices in making the decision to crowdsource
- practices in crowdsourced task design
- practices in finding the right crowd
- practices in managing the crowdsourcing process
- practices in screening and aggregating the results of the crowds' work

B. WHEN SHOULD A TASK BE CROWDSOURCED

Crowdsourcing is useful under certain circumstances according to Chiu, Liang, and Turban:

> Organizations deploy crowdsourcing when they have a problem they need to solve, when they want to exploit opportunities, or when they need a

> large amount of inexpensive labor to perform small tasks (microtasks) that they cannot or do not want to do in-house.[125]

Schenk and Guittard provide a taxonomy of tasks suitable for crowdsourcing, including: (1) simple (routine) tasks with low costs per task and large economies of scale, (2) complex tasks for which the crowdsourcer either lacks the requisite skills or lacks satisfactory in-house solutions, and (3) creative tasks where creativity and uniqueness have value.[126] Crowdsourcers are further characterized by Hossaini et al. (summarized in Table 1) by the terms of the incentive mechanism they adopt, how they recruit, and how they incentivize the crowd, as well as crowdsourcers' ethicality and the level of privacy they provide to crowdsourcees.[127]

Table 1. Features of Crowdsourced Efforts[128]

The Crowdsourcer Features	Short Descriptions
1. Incentives Provision	Providing stimulation for the participants
1.1. Financial incentives	Providing monetary incentives
1.2. Social incentives	Providing community recognition
1.3. Entertainment incentives	Providing gamified and enjoyable experience
2. Open Call	Providing an open audition for participation
3. Ethicality Provision	Providing and following ethical practices
3.1. Opt-out procedure	Providing a method for participants to opt out
3.2. Feedback to crowd	Providing feedback about participants' performance and results
3.3. No harm to crowd	Providing a physically and mentally safe environment
4. Privacy Provision	Providing privacy options for participants

[125] Chao-Min Chiu, Ting-Peng Liang, and Efraim Turban, "What Can Crowdsourcing Do for Decision Support?," *Decision Support Systems* 65 (September 2014): 43.

[126] Erik Schenk and Claude Guittard, "Towards a Characterization of Crowdsourcing Practices," *Journal of Innovation Economics and Management* 7, no. 1 (2011): 93–107.

[127] Mahmood Hosseini et al., "On the Configuration of Crowdsourcing Projects," *International Journal of Information System Modeling and Design* 6, no. 3 (July 2015): 27–45.

[128] Adapted from Hosseini et al., 27–45.

Of these crowdsourcer characteristics, the incentives provision characteristic is the most critical to the crowdsourcing effort, as discussed in detail in the next section. Ethicality and privacy characteristics are self-explanatory with one exception, the need to provide feedback to the crowd. The literature provides a few instances of the impact of feedback mechanisms. Generally, providing feedback can be both an incentive for crowdsources and a way for improving the effectiveness of the crowdsourcing effort.

C. CROWDSOURCED TASK DESIGN PRACTICES

A critical part of any crowdsourcing effort is the design of the crowdsourced tasks. The requesters (people requesting the crowdsourced effort) first need to estimate the workforce required, and second, they have to break the task down into subtasks so that the individual subtasks and the overarching task are tractable. The requesters must ensure that the execution of each subtask does not affect the performance of any other subtask. Hossaini et al. provide typical crowdsourced task characteristics as listed in Table 2.[129] Crowdsourced tasks that have some to many of the characteristics listed in Table 2 are suitable for analysis or execution using crowd-based techniques.

Table 2. Crowdsources Task Characteristics[130]

The Crowdsourced Task	Short Descriptions
1. Traditional operation	How the crowdsourcing task is conventionally performed
1.1. In-house	Task performed by employees
1.2. Outsourced	Task performed by outside organizations
2. Outsourcing Task	The true nature of a crowdsourced task is that it can be outsourced
3. Modularity	How task can be broken up into smaller tasks
3.1. Atomic tasks	Task is indivisible
3.2. Divisible to micro tasks	Task can be divided into micro tasks
4. Complexity	The complexity measurements of the task

[129] Hosseini et al., "On the Configuration of Crowdsourcing Projects," 27–45.

[130] Adapted from Hosseini et al., 27–45.

The Crowdsourced Task	Short Descriptions
4.1. Simple tasks	Task is simple and straightforward
4.2. Complex tasks	Task is difficult and not straightforward
5. Solvability	How task is solved
5.1. Simple for humans	Task is simple enough to be solved by individuals
5.2. Complex for computers	Task is too complicated to be solved by computers
6. Automation Characteristics	How task can be automated
6.1. Difficult to automate	Task is difficult (if not impossible) to automate
6.2. Expensive to automate	Task is expensive to automate
7. User-driven	The individuals' perception of the task
7.1. Problem solving	Task is a problem to be solved
7.2. Innovation	Task needs individual's innovation
7.3. Co-creation	Task needs individuals collaboration in production
8. Contribution Type	How task can be performed
8.1. Individual contribution	Individuals perform on their own to reach a solution
8.2. Collaborative contribution	Individuals need to collaborate to reach a solution

Crowdsourced tasks can be implemented sequentially, in parallel, or by divide and conquer implementation. According to Chitilapilly et al. in sequential implementation, the tasks are divided by the crowdsourcer into "small subtasks" and are "executed in sequence," by taking the output of a given task as "input to the next task."[131] In parallel implementations, tasks are divided into independent subtasks, run together in parallel, and later, the crowdsourcer merges them together to "form the final output."[132] In divide and conquer implementations, the overarching problem is "recursively split into smaller, far easier problems."[133] Once the crowd solves them, the crowdsourcer merges the solutions back to "generate the final problem solutions."[134] The incentive structure used to reward participation and accuracy in tasks or subtasks must

[131] Anand Inasu Chittilappilly, Lei Chen, and Sihem Amer-Yahia, "Survey of General-Purpose Crowdsourcing Techniques," *IEEE Transactions on Knowledge and Data Engineering* 28, no. 9 (2016): 2246–2266.

[132] Chittilappilly, Chen, and Amer-Yahia, 2246–2266.

[133] Chittilappilly, Chen, and Amer-Yahia, 2246–2266.

[134] Chittilappilly, Chen, and Amer-Yahia, 2246–2266.

also be addressed as part of the design. Details on incentive structure and its implications are in Section B.

D. PRACTICES IN FINDING THE CROWD

Refer to Geiger et al. who considered two mechanisms for selecting a proper crowd, namely qualification-based mechanisms and context-specific mechanisms.[135] Under a qualification-based participant selection system, potential crowdsources have to demonstrate certain knowledge or skills before they are allowed to contribute. In a context-specific election mechanism, the crowd is selected by the crowdsourcer based on the decision context. For example, the crowd could consist of individuals who bought a washing machine in the last year if a company was deciding on the attributes of the next year's model.

Finding the right crowd is critical to the success of any crowdsourced effort. The crowd must be diverse, comparatively large, and well-motivated. The crowd for a given task "may include different populations (non-experts, experts, informal members, customers, business partners, etc.)" and may vary in "size, composition, uniformity, and level of expertise."[136] As listed in Table 3, Hossaini et al. provide some key features of effective crowds.[137] Of these characteristics, the diversity and suitability characteristics are critical to the success of the crowd-based effort.

[135] David Geiger et al., "Managing the Crowd: Towards a Taxonomy of Crowdsourcing Processes," in *Proceedings of the Seventeenth Americas Conference on Information Systems* (Detroit, MI: Association for Information Systems, 2011), https://pdfs.semanticscholar.org/d134/065587b5276bec1b0e93695edd673d0bfc10.pdf.

[136] Chiu, Liang, and Turban, "What Can Crowdsourcing Do?" 43.

[137] Hosseini et al., "On the Configuration of Crowdsourcing Projects," 27–45.

Table 3. Features of the Crowd[138]

The Crowd Features	Short Descriptions
1. Diversity	The state or quality of being different or varied
1.1. Spatial diversity	Diversity in location (geographical, department, etc.)
1.2. Gender diversity	Diversity in gender (male or female)
1.3. Age diversity	Diversity in age
1.4. Expertise diversity	Diversity in skills, knowledge, or proficiency
2. Unknown-ness	The condition or fact of being anonymous
2.1. Not known to crowdsourcer	Being anonymous to the crowdsourcer
2.2. Not known to each other	Being anonymous to other individuals in the crowd
3. Largeness	Consisting of big numbers
3.1. Number fulfils the task	Enough individuals to solve the problem
3.2. Number not abundant	Enough individuals to avoid confusion or management issues
4. Undefined-ness	Not being determined, random
5. Suitability	Suiting a given purpose, occasion, or condition
5.1. Competence	Ability and expertise in performing a task
5.2. Collaboration	Working together with other individuals
5.3. Volunteering	Offering capabilities to perform a task
5.4. Motivation	The inspiration to perform a task
5.4.1. Mental satisfaction	Joy of performing a task
5.4.2. Self-esteem	Feeling proud and confident
5.4.3. Personal skill development	Developing individual’s abilities
5.4.4. Knowledge sharing	Distributing the personal information
5.4.5. Love of community	Caring about one’s community

Crowdsourcing platforms like Amazon Mechanical Turk or CrowdFlower may be a good starting point for building the crowd necessary for the crowdsourced task(s). Other options include inviting the participation of members of a given community, say political science graduate students, or from communities that have a stake in the outcome of the task.

138 Adapted from Hosseini et al., 27–45.

E. PRACTICES IN STRUCTURING INCENTIVES FOR CROWDSOURCEES

It is a crowdsourcing axiom (supported by research) that an active, diverse, and comparatively large crowd of participants is central to successful crowdsourcing efforts. Thus, the motivation of crowdsourcees is crucial. Upon review of the literature, Morschheuser et al. found a plethora of studies on what motivates crowdsources, and they list a wide variety of intrinsic and extrinsic motivations for participation.[139] Intrinsic motivators ranged from indulging participants' creativity, to allowing participants to enjoy autonomy, to helping participants develop their own skills and feel competent, to enabling participants to enjoy a pastime, or to achieve social recognition. Extrinsic motivators included financial payoffs or external social reasons.[140] According to Liu, "Evidence shows that prizes and rewards can increase participation rates, but opportunities for learning and skill building are essential for enhancing the quality of participants' contributions."[141]

Gamification is another way of increasing the likelihood that the crowd will indeed participate in the crowdsourcing effort by making the crowdsourced work take on the characteristics a game, which thus provides an incentive to participate other than that of monetary compensation. The purpose of gamification is to change crowdsourcees' motivations from those of an extrinsic gain-seeker individual to those of an intrinsically self-motivated individual. Hamri, Koivisto, and Sarsa, as well as Seaborn and Fels, reviewed research on gamification and found that gamification was likely to lead to increases in crowd participation.[142]

---

[139] Hosseini et al., 27–45.

[140] Winter Mason and Duncan J. Watts, "Financial Incentives and the Performance of Crowds," *ACM SigKDD Explorations Newsletter* 11, no. 2 (2010): 100–108, doi: 10.1145/1809400.1809422.

[141] Helen K. Liu, "Crowdsourcing Government: Lessons from Multiple Disciplines," *Public Administration Review* 77, no. 5 (2017): 656–667.

[142] Juho Hamari, Jonna Koivisto, and Harri Sarsa, "Does Gamification Work? A Literature Review of Empirical Studies on Gamification," in *Proceedings of the 47th Hawaii International Conference on System Sciences—HICSS* (Waikoloa, HI: IEEE, 2014), 3025–3034, doi: 10.1109/HICSS.2014.377; Katie Seaborn and Deborah I. Fels, "Gamification in Theory and Action: A Survey," *International. Journal of Human Computer Studies* 74 (February 2015): 14–31, http://dx.doi.org/10.1016/j.ijhcs.2014.09.006.

Gamified experience designs often include combining points with leaderboards to create competition between crowdsourcees. Points can also be combined with other elements, such as time limits, the level of crowdsourcees participation, rewards for cooperation, badges, and missions visualizing specific goals. Crowdsourcees' behavioral outcomes as engendered by gamification often revolve around the participation level of crowdsourcees in the gamified effort. Several studies report increases in (long-term) participation, increases in the quality of output, as well as reductions in cheating behaviors. However, financial incentives have the greatest impact. Simple gamification using points and leaderboard replace financial incentives when it comes to incentivizing crowdprocessing. Indeed, for such tasks, a review of the literature by Chittilappilly, Chen, and Amer-Yahia finds "Monetary incentives are the best and easiest way to motivate."[143]

F. PRACTICES IN MANAGING THE CROWDSOURCING PROCESS

Issues related to the crowdsourcing process include process governance, process design, legal issues, and the characteristics of the crowdsourcing platform. Process designs for crowdsourcing systems center around the type of problem to be addressed. Common critical configuration items for crowdsourcing efforts focus on complex tasks or sentiment elicitation, as determined by a survey of experts conducted by Hossaini et al., which include choices regarding diversity and financial or other incentives to spur motivation, the crowdsourcing platform's ease of use, presence or absence of feedback, and types of feedback mechanisms, largeness, and competence of participants.[144] The design of each task and subtask involves choices (as discussed previously), the sequencing of tasks, and the parameters of the task itself. Simple estimation tasks are designed differently from conditional estimation tasks, which are yet again different from tasks requiring the application of expert judgment. According to Luz, Silva, and Novais, crowdsourcing workflows consist of (1) selecting workers and distributing the task(s), (2) assigning tasks to workers, (3) task performance (4) assessment of task results; (5)

[143] Chittilappilly, Chen, and Amer-Yahia, "Survey of General-Purpose," 2249.

[144] Mahmood Hosseini et al., "Recommendations on Adapting Crowdsourcing to Problem Types," in *IEEE 9th Conference on Research Challenges in Information Science* (Athens, Greece: IEEE RCIS, 2015).

aggregation of task results, and (6) giving workers rewards consistent with worker incentives.[145] These workflows are best managed using dedicated Web 2.0 platforms as described in the next section.

Chiu, Liang, and Turban speak to the importance of the choice of a crowdsourcing platform when they posit, "Proper matching between platform functions and task types can enhance the performance of crowdsourcing."[146] Hossaini et al. provide an overview of key crowdsourcing platform characteristics, saying, "A crowdsourcing platform would typically need to offer four main facilities; facilities that deal with the crowd, facilities that deal with the crowdsourcer, facilities that deal with the crowdsourced task, and facilities that are related to the platform itself."[147] The details of each key crowdsourcing platform characteristic Hosaini et al. determined are given in Table 4.

Table 4. Crowdsourcing Platform Characteristics[148]

The Crowdsourcing Platform Facilities	Short Descriptions
1. Crowd-related Interactions	Facilities in the platform that relate to the crowd
1.1. Provide enrolment	Means to enroll the individuals
1.2. Provide authentication	Means to authenticate the individuals
1.3. Provide skill declaration	Means to help the individuals declare their skills
1.4. Provide task assignment	Means to assign tasks to the right individuals
1.5. Provide assistance	Means to help the individuals during the performing of the task
1.6. Provide result submission	Means to help the individuals to send their results
1.7. Coordinate crowd	Means to coordinate performers in a certain task
1.8. Supervise crowd	Means to supervise individuals during their performance

[145] Nino Luz, Nuno Silva, and Paulo Novais, "A Survey of Task Oriented Crowdsourcing," *Artificial Intelligence Review* 44, no. 2 (2015): 187–213.

[146] Chiu, Liang, and Turban, "What Can Crowdsourcing Do for Decision Support?" 45.

[147] Hosseini et al., "On the Configuration of Crowdsourcing Projects," 27–45.

[148] Adapted from Hosseini et al., 27–45.

The Crowdsourcing Platform Facilities	Short Descriptions
1.9. Provide feedback loops	Means to give feedback to individuals about their performance and about the results
2. Crowdsourcer-related Interactions	Facilities in the platform that relate to the crowdsourcer
2.1. Provide enrolment	Means to enroll the crowdsourcers
2.2. Provide authentication	Means to authenticate the crowdsourcers
2.3. Provide task broadcast	Means to broadcast the task to the right individuals
2.4. Provide assistance	Means to help the crowdsourcers for announcing the task
2.5. Provide time negotiation	Means to help crowdsourcers negotiate time requirements with the individuals
2.6. Provide price negotiation	Meanstohelpcrowdsourcersnegotiateperformanceprices withtheindividuals
2.7. Provide result verification	Means to verify whether submitted results meet the needs of crowdsourcers
2.8. Provide feedback loops	Means to give feedback to crowdsourcers about individuals' performances
3. Task-related Facilities	Facilities in the platform that relate to the task
3.1. Aggregate results	Means to collect and unify submitted results
3.2. Hide results from others	Means to hide individuals' results from each other for privacy reasons
3.3. Store history of completed tasks	Means to keep a history of the completed tasks and related information (such as who completed them, the spent time, etc.)
3.4. Provide quality threshold	Means to guarantee the required quality of results
3.5. Provide quantity threshold	Means to guarantee the required number of responses
4. Platform-related Facilities	Facilities in the platform that relate to the platform itself
4.1. Online environment	Means to keep the platform online and accessible to individuals
4.2. Manage platform misuse	Means to report if there are instances of platform misuse
4.3. Provide ease of use	Means to keep the platform simple to use
4.4. Provide attraction	Means to keep the platform attractive to use
4.5. Provide interaction	Means to keep the platform interactive
4.6. Provide payment mechanism	Means to enable crowdsourcers to pay individuals in their preferred way

G. PRACTICES IN SCREENING AND AGGREGATING THE CROWD'S RESULTS

Once the crowdsourcees have performed their work or are in the process of performing their work, genuine answers must then be separated from those of cheaters using different quality control methods (the work must actually be done in a credible way), and then the workers' results need to be aggregated to create the final crowdsourced answer. In other words, the alternative solutions, or other output from the crowd, have to be evaluated. Often, the output of individuals in the crowd is aggregated when doing so and the quality of work judged in the light of the task's stated goals.

H. RELEVANT PRACTICES IN THE DESIGN OF THIS CROWDSOURCING EFFORT

The relevance of a given crowdsourcing practice to the thesis must be determined based on the overall vision for the analytic and forecasting methodology proposed in this thesis. As discussed in Chapter I, the proposed methodology combines prediction markets techniques with crowd-based techniques. It can be argued that a prediction market is a specific implementation of a crowdsourced problem solving methodology.

The relevant crowd-based practices discussed in this chapter are both an adjunct to and compatible with the relevant prediction markets practices discussed in Chapter IV. In this context, the broad design features of the crowdsourced task (the prediction market) are discussed in the following sections. Note that the thesis only proposes a design for the crowdsourced prediction markets-based analytic and forecast methodology; implementation and testing of the design is left to others.

Key features of the crowdsourced task proposed in this thesis parallel those the tables in this chapter depict. First, choices made from the incentive related options in Table 1 include (1) the use of incentives for participation in the task centered on social and entertainment based incentives, (2) the provision of feedback, and (3) the maintenance of privacy. Social and entertainment-based incentives are appropriate for this effort given the issues surrounding the use of financial incentives (discussed in detail in Chapter IV). Second, social imperatives, such as peer recognition and the spirit of competition, also will incentivize participants. Third, the crowdsourced tasks are also gamified, discussed as follows, which drives the entertainment value of participation. Fourth, feedback is provided in an effort to maintain engagement in the task by playing to the desire of most individuals to do well, as measured against their peers by learning from feedback, and by providing tools that can enable greater participant performance self-improvement, to which most individuals aspire. Fifth, participant privacy is maintained to ensure that participants can freely share their opinions without negative consequences and so that poor performance on the task does not serve as a disincentive for participation, as may be the case if the identity of the poor performer is publicly known. Sixth, open call is not used to allow the sponsor of the crowdsourced effort to control the composition of the community attempting the crowdsourced task. This control is essential to achieve the appropriate levels of diversity of opinion and the right mix between insiders and outsiders in terms of access to information not available to the public.

Choices made from the task related design options in Table 2 include task designs that are (1) modular and atomic, (2) complex, (3) solvable, (4) difficult to automate, (5) based on user driven problem solving and innovation, and (6) for some tasks, co-creation (team built). The need for task modularity and atomicity is driven by the notion that the intelligence problems the crowd will be asked to address have definable, measureable, and clear outcomes, and thus, avoid the need for a hierarchy of subtasks with ambiguous outcomes that crowd members must complete first. By definition, the tasks the crowd will solve are complex and difficult to automate; otherwise, the need for the effort will be negated. Tasks will also be designed to be solvable, in that participants will be able to come up with an answer to the question being posed, but the accuracy of that answer will depend of the participant’s forecasting ability. Furthermore, it goes without saying that the whole purpose of the crowdsourced effort is to bring problem solving skills and innovation to the task of intelligence analysis and that the effort will allow teamwork, that is co-creation, for some tasks as a way of testing the effectiveness of teams in performing intelligence analysis tasks.

The choice of crowd is based on a consideration of the options in Table 3 and determined by the sponsor. Based on the need for predictive accuracy, the crowd is to be diverse, as assessed by gender, background, age, expertise, and also suitable, as defined by background, competence, and skill. The crowdsponsor is periodically reassessed to determine participant suitability based on a predictive track record, as well as ongoing levels of participation. The crowd is to be known to the crowdsponsor but not to each other to ensure adequate diversity in the case of the former and independence of opinion in the case of the latter.

In terms of platform design, commercially available platforms for crowdsourcing efforts (in general) and prediction markets (in particular) address the entire range of crowdsourcing platform requirements listed in Table 4. Finally, the prediction markets that make up the methodology proposed in this thesis are a special case of crowdsourcing problem solving that by their nature aggregate and screen results. As a result, particular attention does not have to be paid to aggregating and screening results in this context.

I. CONCLUSION

Design characteristics central to the effectiveness of the crowdsourcing effort include:

- practices in making the decision to crowdsource
- practices in crowdsourced task design
- practices in finding the right crow
- practices in managing the crowdsourcing process
- practices in screening and aggregating the results of the crowds' work

Although many possible variations of crowdsourcing effort design are possible, this chapter reviewed some of the most salient characteristics of crowdsourced efforts and selected several design parameters for inclusion in crowdsources design methodology that this thesis develops. The researcher selected these parameters based on their likely impact on the performance of the crowdsourced effort and their impact on participant involvement and diversity (an indirect driver of crowdsourcing effectiveness). This design effort is crucial to both the design of the prediction market developed in Chapter IV and to the overall success of the analytic and forecasting methodology proposed in subsequent chapters. The next chapter builds on this chapter by considering and choosing prediction markets design options consistent with the crowd sourcing practices previously discussed and that reflect the design requirements of this thesis' methodology.

IV. PREDICTION MARKETS PRACTICES

A. INTRODUCTION

Chapter II introduced prediction markets and crowd-based techniques for performing analysis and making forecasts, and it discussed these techniques in the context of intelligence analysis and forecasting. Chapter II also introduced intelligence culture and examined the implications of intelligence culture for analytic and forecast accuracy. This chapter has a detailed discussion of relevant practices in designing prediction markets for forecasting that covers design practices drawn from multiple disciplinary contexts. Given the primary goal of this thesis discussed in Chapter I, the selection and application of known practices in prediction markets design is required as a precursor to the development and presentation of the thesis' intelligence analytic methodology in subsequent chapters. Areas of interest this chapter covers are those central to the accuracy of prediction markets and include incentive mechanisms, trading processes, clearing house design, and investor (participant) management processes. Additionally, this chapter reviews existing practices in each of the aforementioned areas of prediction markets design and identifies relevant practices for implementation in the analytic methodology developed in this thesis.

B. PREDICTION MARKET DESIGN

Prediction market design in large part determines how accurate the prediction market is in forecasting future events. This section introduces and discusses the key prediction market design parameter choices that must be considered and selected for the greatest forecast accuracy.

1. Design of Prediction Markets Incentive Mechanisms

Properly designed incentive mechanisms ensure both that participants in a given prediction market are invested in giving their best analytic efforts as they make their predictions and that participants engage in the market at a sufficient level so that the market is updated as new information becomes available.[149] Table 5 contains factors related to prediction market incentive structures.[150]

Table 5. Prediction Market Incentive Mechanisms

Prediction Markets Design Factors: Incentive Mechanisms		
Performance	Performance evaluation	Wealth: portfolio value based Accuracy: best predictors Effort: trading behaviors minimum number of trades
Reward	Reward type	Non-monetary (prize) Monetary Corporate support
	Reward base	Tournament: performance based Lottery: luck based
Involvement		Trading sessions Workshops Training Other

The incentive structure embodied within successful prediction markets revolves around motivating serious participant engagement as exemplified by the frequency and quality of participants' trades. Ideally, participants in the prediction market should find the participation incentive sufficient to motivate engagement, the level of work required for participation manageable, and adequate reciprocity occurs between market sponsors

149 Justin Wolfers and Eric Zitzewitz, "Prediction Markets," *Journal of Economic Perspectives* 18, no. 2 (2004): 107–126.

150 Gaspoz, *Prediction Markets Supporting Technology*, 82–83.

and participants. Active, public commitment by participants' management (in the case of private prediction markets) to the prediction market effort also helps.[151]

Three design factors in Table 5 deserve further discussion. First, no proof exists that using real money as a reward leads to superior accuracy in making predictions.[152] Second, in terms of the reward base, the rank order tournament or fixed payouts for accuracy seem superior to other pay for performance schemes in serving as incentive mechanisms.[153] Finally, it is possible to use various different schemes to initiate and maintain involvement in public and private prediction markets. No one or group of involvement mechanisms is clearly superior.[154]

2. Design of Prediction Markets Trading Process

Table 6 presents the design factors related to prediction market trading processes.[155]

---

151 Carolin Decker, Isabelle M. Welpe, and Bernd H. Ankenbrand, "How to Motivate People to Put Their Money Where Their Mouth Is: What Makes Employees Participate in Electronic Prediction Markets," *Technological Forecasting and Social* 78, no. 6 (2011): 1002–1015; Tung H. Ho and Kay Y. Chen, "New Product Blockbusters: The Magic and Science of Prediction Markets," *California Management Review* 50, no. 1 (2007): 144–158.

152 Tom W. Bell, "Private Prediction Markets and the Law," *Journal of Prediction Markets* 3, no. 1 (2009): 89–110; Emile Servan-Schreiber et al., "Prediction Markets: Does Money Matter?," *Electronic Markets* 14, no. 3 (2004): 243–251, doi: 10.1080/1019678042000245254.

153 Stefan Luckner, "How to Pay Traders in Information Markets: Results from a Field Experiment," *Journal of Prediction Markets* 1, no. 2 (2007): 147–156.

154 Luckner, 147–156.

155 Gaspoz, *Prediction Markets Supporting Technology*, 1.

Table 6. Prediction Market Trading Process Characteristics

Prediction Markets Design Factors: Trading Process		
Fees		Trading fees Expiration fees IPO Fees
Trading Mechanisms	Double auction	Open order book Closed order book
	Market maker	Market scoring rules Dynamic pari-mutuel Continuous double auction
	Trading time	24/7 Trading sessions Selected opening hours

The effect of trading fees on prediction markets does not seem to have been studied in detail in the literature. Most extant prediction markets do not charge any of the types of fees Table 6 lists.[156] In terms of prediction market trading mechanisms, the double auction, or its close cousin, the continuous double auction (CDA), are both analogous to the mechanisms used in financial markets and are also the most commonly used prediction market trading mechanism.

A double auction occurs when traders submit buy or sell orders directly executed against opposite orders (each buy order at a given price is matched to another trader's sell order at that price or is stored in an order book and then processed iteratively as countervailing sell orders come in). Unexecuted orders expire after a given period of time. This approach is advantageous because new information is incorporated into the market continuously as the prices at which participants place buy and sell orders change as the new information is digested by market participants. Market liquidity (the ability to find counterparties to a given trade) can be an issue. However, this issue can be addressed

[156] Gaspoz, 83.

by adding participants, known as market makers. The market makers charge a premium to buy and a discount to sell from their own inventory, called a bid-ask spread, and by adjusting the amount of the bid-ask spread, the market makers ensure that the market clears (trades take place).This process also allows trades to occur at different times (that is a CDA) with the market makers trading against their own inventory until a counterparty can be found or the bid-ask spread is adjusted to clear the market. Market liquidity is always an issue for the market makers. Hanson found that CDAs work only in "thick market" scenarios (with a lot of trades and traders relative to the number of contracts traded).[157] This issue, discussed in the next section, can be addressed by using a market scoring rule to govern the actions of the market makers.

A market scoring rule takes the CDA, as described above, and modifies the behavior of the market and the market maker in particular to address liquidity concerns. Using a logarithmic market scoring rule requires that we first agree that there only two outcomes are possible: (1) Traders can only buy or sell shares of (bet on or against), and (2) only one of the two outcomes is guaranteed to occur over the course of time. The market maker tallies the total number of shares have purchased at a given time for each outcome. The market maker also uses a cost function that records how much money traders have spent in total on each outcome at that given point in time. See the following equation.

$$C = b * \ln\left(e^{q1/b} + e^{q2/b}\right),$$

wherein

C=cost per share

b=market liquidity maintained by the market maker (number of shares)

q_1=quantity of shares of outcome 1

q_2=quantity of shares of outcome 2

---

[157] Robin Hanson, "Combinatorial Information Market Design," *Information System Frontiers* 5, no. 1 (2003): 107–119.

Then, traders individually submit how many shares of each outcome they want to buy or sell. The market makers then use the cost function to price each outcome and come up with a cost of the trade.[158] Dynamic pari-mutuel markets (DPMs) are an alternative market mechanism to market scoring rule-based market makers. In traditional pari-mutuel markets:

> The …pari-mutuel market… is operated in a manner where market traders purchase shares for a specific possible outcome. When the outcome is determined, the money collected is paid out to the winners in proportion to the number of winning shares that they hold. This technique protects market organizer from sustaining a loss under any circumstance.[159]

To quote the seminal work on DPMs by Pennock:

> A DPM acts as hybrid between a pari-mutuel market and a continuous double auction (CDA), inheriting some of the advantages of both. Like a pari-mutuel market, a DPM offers infinite buy-in liquidity and zero risk for the market institution; like a CDA, a DPM can continuously react to new information, dynamically incorporate information into prices, and allow traders to lock in gains or limit losses by selling prior to event resolution. The trader interface can be designed to mimic the familiar double auction format with bid-ask queues, though with an addition variable called the payoff per share. The DPM price function can be viewed as an automated market maker always offering to sell at some price, and moving the price appropriately according to demand. Since the mechanism is pari-mutuel (i.e., redistributive), it is guaranteed to pay out exactly the amount of money taken in.[160]

Thus, unlike CDAs with market scoring rules, market makers do not experience any risk under the DPM, as it maintains the liquidity of the market. Luckner highlights one shortcoming of prediction markets using a DPM market maker; that under DPM, no

---

[158] Robin Hanson, "On Market Maker Functions," *Journal of Prediction Markets* 1, no. 1 (2007): 3–15.

[159] Mark Peters, Anthony Man-Cho, and Ye Yinyu, "Pari-Mutuel Markets: Mechanisms and Performance," in *WINE 2007: Internet and Network Economics*, Lecture Notes in Computer Science Series, vol. 4858 (Heidelberg: Springer, 2007), 82–83.

[160] David M. Pennock, "A Dynamic Pari-mutuel Market for Hedging, Wagering, and Information Aggregation," in *Proceedings of the Fifth ACM Conference on Electronic Commerce (EC'04)* (New York: ACM, 2004), 172.

incentive exists to buy early.[161] The best strategy is to wait until the last moment to buy, which negates the continuous discovery of event probabilities at all points in time that is a hallmark of prediction markets using CDAs with market scoring rules.[162] The utility of market makers using DPM-based market makers is supported by Slamka, Skiera, and Spann when they found:

> That logarithmic scoring rules and the dynamic pari-mutuel market attain the highest forecasting accuracy, good robustness against parameter misspecification, the ability to incorporate new information into prices, and the lowest losses for operators.[163]

3. Design of Prediction Markets Clearing Houses

Table 7 lists design factors related to prediction market trading processes.[164] Order matching can be based on price-submission (sort and match by price and then sort by time of submission and give priority to the oldest orders) or price quantity (sort and match by price and then sort by quantity and give priority to the smallest orders). In theory, spending caps may be required in play money markets to prevent participants form manipulating the market.[165] However, most extant prediction markets do not enforce spending caps.

Table 7. Clearing House Design

Prediction Markets Design Factors: Clearing House		
Order	Order matching rules	Price and submission Price and quantity
	Order spending caps	Enforced

[161] Stefan Luckner, "Prediction Markets: Fundamentals, Key Design Elements and Applications," in *Proceedings on the 21st Bled Conference* (Bled, Slovenia: Association for Information Systems, 2008), 236–247.

[162] Luckner, 236–247.

[163] Slamka, Skiera, and Spann, "Prediction Market Performance," 180–181.

[164] Gaspoz, *Prediction Markets Supporting Technology*, 170–171.

[165] Michael Abramowicz, "Deliberative Information Markets for Small Groups," in *Information Markets: A New Way of Making Decisions*, ed. Robert Han and Paul Tetlock (Washington, DC: AEI Press, 2006), 101–125.

Prediction Markets Design Factors: Clearing House		
		No caps
	Order type	Market Limit
	Short selling	Allowed Not allowed
Asset	Asset type	Real money Play money
	Inflation	Play money only
	Borrowing	Margin purchases
	Endowment	Initial endowment (money and/or contracts) Weekly endowment (money and/or contracts)
Claim	Claim IPO	Fees and rewards Screening
	Initial Claims	Starting quotes Quantity
	Claim ontology	
	Claim type	Winner take all Conditional Index Spread
	Claim structure	Bundle Independent
Payoff	Settlement date	Public Random
	Settlement judge	
	Settlement price	“Truth” Proxy Volume weighted average price Final market price

CDA markets without market makers only use limit orders (buy or sell when the price reaches X; if X is not reached, the trade does not execute), whereas CDA markets with market makers can use market orders as well (execute the order at the present market price). Many have argued that, in theory, constraints on short selling (selling shares one has borrowed from the market maker at a price set today in anticipation that the market price will drop) may lead to speculative bubbles in prediction markets; the degree to which short selling improves the quality of prediction market predictions is unknown.[166] Similarly, the impact of margin (being able to borrow liquidity or other assets from the market maker) is unknown; however, in theory, this type of borrowing can result in poor predictors in play money markets "doubling down" by borrowing in support of poor predictions and thus exerting too much influence on the market.[167] Players can be given endowments of play money or stocks (predictions) to use in the market once during the start of the prediction market or on a weekly basis or some combination thereof. The literature does not adequately explore the impact of the timing, asset mix, and quantity of such endowments.

The claim IPO process is the process by which new assets (predictions) are created. Traders and the market makers can initiate IPOs. Individuals can decide which predictions should have IPOs by including items of interest to the market makers. The predictions can be screened by experts appointed by the market makers or made by the market itself. In the latter case, once enough players have placed orders for the asset (prediction), the IPO takes place. IPO prices can be fixed by the market makers or discovered by auctioning the asset on the market. One thing is certain; all IPOs should have clear claims. Badly worded claims may be unresolvable (i.e., impossible to prove), and as such, it is essential that claims are clear and understood by all participants in the market.

Claim payoffs can be winner take all (pays in full if the event occurs), conditional (pays off if the event occurs, and if another event occurs), indexed (pays $1 for each percentage of the event occurs; e.g., market share is X percent), or based on a spread (pays $2 if a threshold X is exceeded). Claims can be mutually exclusive (if A occurs, B does not occur) or bundled (in the case of indexed- or spread-based claims more than one

---

166 Gaspoz, *Prediction Markets Supporting Technology*, 93.

167 Gaspoz, 94.

claim is paid and that any claim that beats the spread or exceeds the index are all paid). According to Ozan:

> When finding the probability of the future event is the main objective, winner-takes-all contract scheme designates the most direct approach. If the... analysts are interested in determining the mean value of an outcome than index contracts can provide the optimal performance…Spread contracts are used when median values are needed to be uncovered.[168]

Settlement dates can be public or random and are either bound to a known end date (the date on which the result is known) or are on a date chosen by the market makers (random). The settlement judge is the authority who defines whether an event occurred. In the case of an election, it can be the news media or the relevant secretary of state. In the case of less well defined claims, it can be the newspaper of record, the relevant government agency, etc. As Gaspoz notes, "Claims could [payoff] regarding facts…proxies… in this case experts... [can] define the payoff for each claim… [or] the Volume Weighted Average Price…over the last five trading days" can be used to set payoffs.[169]

4. Design of Practices in Managing Prediction Markets Investors (Participants)

Traders are the investors in a prediction market. It is clear that the larger the community of traders, the more robust the prediction markets' predictive prowess.[170] Table 8 describes design factors affecting how these investors and their participation in prediction markets can be managed.

[168] Erol Ozan, *Optimization of Information Technology Risk Event Prediction Markets* (Greenville, NC: East Carolina University, 2013); Erol Ozan, "The Use of Prediction Markets in Information Technology Risk Management" (paper presented at American Society for Engineering Management Conference, Virginia Beach, VA, 2012), 2.

[169] Gaspoz, *Prediction Markets Supporting Technology*., 101.

[170] Joyce E. Berg, Forrest Nelson, and Thomas Rietz, *Results from a Dozen Years of Election Futures Market Research*, quoted in Gaspoz, *Prediction Markets Supporting Technology*, 98.

Table 8. Investor Management

Prediction Markets Design Factors: Investor Management		
Market	Market policy	Open market Closed market
	Market transparency	Display all information Restrict to some indicators
Investor	Investor anonymity	Userid Username No anonymity
	Investor unicity	Enforced Trust
	Investor selection	Quantity Diversity Informed versus non-informed Benefit from market outcome

Prediction markets can be closed (participants are from a given community or are selected on some basis or affiliation) or open (open to the public). Furthermore, prediction markets are usually transparent (all available information is available to all traders). Yang, Li, and van Heck examined the implications of prediction market transparency and found that:

> Improved information transparency (disclosure of different traders’ buy and sell orders) can lead to higher levels of traders’ dynamic interactions. Increases in traders’ participation activity and dynamic interactions lead to higher information aggregation efficiency and greater market predictive accuracy.[171]

However, available information can be restricted to achieve objectives, such as preventing the impacts of combinatorics, moral hazard, manipulation, hidden prices, and

[171] Sheng-yun Yang, Tung Li, and Eric van Heck, “Information Transparency in Prediction Markets,” *Decision Support Systems* 78 (2015): 67.

decision selection bias.[172] Investor anonymity is usually guaranteed by most public and private prediction markets to prevent the impacts of secret accounts, shared interests, etc.

In terms of investor selection, as already stated, the more investors the better when it comes to the accuracy of predictions made using prediction markets. This “more is better” approach also applies to investor diversity. As to informed versus uninformed traders, and involving those who benefit from market outcomes, the literature suggests that both informed and uniformed traders are critical to a give prediction market’s predictive accuracy, while the impact of the latter has not been adequately explored.

C. RELEVANT PRACTICES IN PREDICTION MARKETS DESIGN

Practices in prediction markets design relevant to the proposed methodology developed in this thesis are summarized in Table 9 and discussed in the subsequent sections. These practices are chosen because they both reflect best practices as found in the literature and are specifically relevant to the proposed methodology.

Table 9. Prediction Markets Design: Relevant Practices

Prediction Markets Design: Relevant Practices			
Incentive Mechanisms	Performance	Performance evaluation	Accuracy: best predictors
	Reward	Reward type	Non-Monetary
		Reward base	Tournament: performance based
	Involvement		24/7
Trading Process	Trading Mechanisms	Fees	No Fees
		Market maker	Dynamic pari-mutuel
		Trading time	24/7

[172] Robin Hanson, “Impolite Innovation: The Technology and Politics of ‘Terrorism Futures’ and Other Decision Markets,” in *Promoting the General Welfare, American Democracy and the Political Economy of Government Performance*, ed. Eric Patashnik and Alan Gerber, 151–173 (Washington, DC: Brookings Institution Press, 2006).

Prediction Markets Design: Relevant Practices			
Clearing House	Order	Order matching rules	Price and quantity
		Order spending caps	No caps
		Order type	Market
			Limit
		Short selling	Allowed
	Asset	Asset type	Play money
		Inflation	None
		Borrowing	Margin purchases
		Endowment	Initial endowment (money and/or contracts)
	Claim	Claim IPO	Screening
		Initial Claims	Starting quotes
		Claim type	Conditional
			Index
			Spread
		Claim structure	Independent
	Payoff	Settlement date	Public
		Settlement judge	Sponsor
		Settlement price	Final market price
	Market	Market policy	Closed market

Prediction Markets Design: Relevant Practices			
		Market transparency	Display all information
	Investor	Investor anonymity	Username
		Investor unicity	Enforced
		Investor selection	Diversity
			Informed versus non-informed

In terms of investment mechanisms, given that improved predictive accuracy is the overarching goal of this thesis methodology, evaluation of performance based on accuracy is the best possible option. Furthermore, due to practical and political reasons (participants earning monetary rewards if a "bad" outcome is predicted and comes to pass due to the possible action of the participant) coupled with the fact that the literature suggests no difference in accuracy if nonmonetary rewards are used, suggests that using non-monetary rewards is appropriate. Finally, tournament-based rewards earned by trading assets 24/7 can increase involvement (number of trades) due to the intrinsic value participants place on winning competitions and the increased availability and ease of participation (participants can participate whenever they have time available), respectively.

The trading process uses DPM market makers, no fees, and 24/7 trading hours. The choice of the DPM is based in part on the assertion by Slamka, Skiera, and Spann, who assert, "The dynamic pari-mutuel market attains the highest forecasting accuracy, good robustness against parameter misspecification, the ability to incorporate new information into prices, and the lowest losses for operators."[173] The need for fees is

[173] Slamka, Skiera, and Spann, "Prediction Market Performance," 160.

negated by the reality that most extant prediction markets do not charge trading fees and the dearth of information on the impact of fees in this context in the literature. This researcher chose 24/7 trading due to its likely positive impact on participant involvement and the fact that such a structure allows for continuous price (prediction) discovery.

The choice of clearinghouse characteristics is based on the fact that prediction markets are closely analogous to financial markets, and as is the case with financial markets, practices that encourage accurate price discovery and increase market efficiency (accuracy in the case of prediction markets) should result from the choices made. As a result, prediction markets clearinghouse practices that mimic those of financial markets (price and quantity-based order matching, no spending caps, allowing limit and market orders, allowing short selling, allowing margin purchases) can and should be implemented in the prediction market methodology this thesis proposes. The choice of play money as the yardstick by which participants' predictive portfolios' value is measured is driven by the fact that the literature suggests no appreciable difference in performance (accuracy) between prediction markets that use real money versus those that use play money.

This researcher chose some clearinghouse practices unique to prediction markets (settlement dates, settlement judges, asset characteristics, claim characteristics) for inclusion in the methodology based on how they drive increased forecast accuracy. For example, the allowance of IPOs and conditional, index and spread-based contracts allow the prediction market sponsors to tailor the assets traded to the events simulated (e.g., the sponsors may create an asset that reflects a conditional event-if A happens, then B happens), or the sponsors may reflect new contracts for events that have not been included in the prediction market to date via the IPO process. In addition, not only does the choice of making each claim in the prediction market independent of other claims simplifies market functioning, it increases market transparency, and as a consequence, market efficiency and thus prediction market accuracy. Similarly, for the market to have public settlement dates and have the prediction market sponsor be the settlement judge, increases prediction market accuracy as well. The use of the final market price as the settlement price ensures that market participants' portfolios are priced in a way correctly reflecting their predictive accuracy over time.

Finally, this researcher selected prediction market practices relating to investor management based on their consistency with the goal of outside participation driving the diversity of opinion. As discussed in the literature, it is likely that many of the recent failures of prediction markets to predict outcomes was likely due to insufficient diversity of participation and thus opinion among market participants. Investor (participant) anonymity is maintained to enable the free and transparent sharing of opinion while investor unicity is enforced to prevent one or a few investors from having multiple positions on the same event. Thus, the prediction market is not an exercise in hedging bets. The market is closed to allow the sponsors to control who participates, and as a result, to maintain or enhance the achievement of the goals of diversity in general, as well as maintain a mix of informed and uninformed participants. The latter allows the market sponsors to include opinion based on nonpublic sources, analogous to insider information in financial markets and to encourage strong form market efficiency, and as a result, increase predictive accuracy.

D. CONCLUSION

Design characteristics that drive the accuracy of prediction markets include incentive mechanisms, trading processes, clearing house design, and investor (participant) management processes. Although many possible variations of prediction markets design may be available, this chapter reviewed some of the most salient prediction markets characteristics and selected several design parameters for inclusion in the prediction markets design methodology that this thesis is developing. This researcher selected these parameters based on their likely impact on prediction market accuracy, as well as their impact on participant involvement and diversity (an indirect driver of prediction market accuracy). The choice of parameters also reflects a conscious attempt to model the design of existing financial markets. Indeed, given that the EMH in financial markets is the driver by analogy of prediction markets' accuracy, the degree to which the prediction market design mimics that of financial markets is also a driver of prediction market accuracy. Finally, this chapter included some examples of extant public and private prediction markets.

Next, the thesis discusses how crowd- and prediction markets-based approaches can address the cultural factors driving analytic and forecast accuracy (or inaccuracy) in intelligence analysis.

THIS PAGE INTENTIONALLY LEFT BLANK

V. IMPLICATIONS OF INTELLIGENCE COMMUNITY CULTURE FOR CROWDSOURCED AND PREDICTION MARKETS-BASED ANALYTIC METHODOLOGIES

A. INTRODUCTION

Any analytic methodology hoping to improve the predictive accuracy of the Intelligence Community analysts must both reflect and adapt to the underlying Intelligence Community culture. If this improvement does not occur, it is likely that any new or modified methodology either would be limited in its adoption, or more likely, be ignored by the analytic community at large. Furthermore, the processes and outcomes of intelligence analysis itself are strongly shaped by intelligence culture. Therefore, any proposed intelligence analytic methodology must address the challenges and limitations posed by intelligence culture as first discussed in Chapter II. This focus is especially true of crowd- and prediction markets-based methodologies, such as the one this thesis is developing. The fundamental and operational basis of how these types of methodologies work in practice addresses the cultural sources of intelligence community analytic and forecast inaccuracy. This chapter discusses how they do so in detail. However, these types of methodologies may nevertheless conflict with overarching intelligence community culture. For example, the Intelligence Community's tendency to resist information sharing between agencies is a source of forecast inaccuracy that the methodology must address.

This chapter considers how the crowd and prediction markets analytic and forecast methodology addresses salient Intelligence Community cultural factors leading to forecast inaccuracy. This discussion also provides context for the next chapter, which builds on the work presented in the thesis to this point and discusses the specifics of a proposed crowd-based prediction markets forecasting methodology driving improved intelligence community forecasting and analysis.

B. THE NATURE OF THE OVERARCHING INTELLIGENCE CULTURE AND ITS IMPLICATIONS FOR CROWDSOURCED AND PREDICTION MARKETS-BASED ANALYTIC METHODOLOGIES

The examination of what intelligence is and its implications for analysis and analysts, which Chapter II discusses, strongly suggests that the proposed analytic methodology must be able to cope with or indeed thrive on taking in, processing, and drawing conclusions from a large volume of information from disparate sources and in a wide variety of formats. Furthermore, the methodology must have mechanisms to address the fact that forecasts will likely be based on a significantly incomplete and vague set of facts whose accuracy is indeterminate. It also must be able to cope with the reality of limited available feedback in the short term to refine the analysis and associated forecast. Moreover, the methodology must explicitly address the fact that the availability of input information may suffer from the consequences of the "need to know" and limited distribution approaches to information sharing and dissemination, as well as the consequences of the "information is power" thought pattern at each agency.

Crowd and prediction markets-based analytic and forecasting methodologies address these issues head on. First, the nature of crowd-based prediction markets revolves around distilling all information available to all participants in the market into one thing, the price of the outcome or asset (or in other words, the instantaneous likelihood of the outcome occurring). An attribute of crowd-based prediction markets is that they can do this extraction without the need for intensive analysis or active management on the part of the market sponsor. The choices made by individual market participants as to how they price a given outcome or asset causes the market to make an instantaneous assessment of the aggregate likelihood of that event or outcome occurring through the "magic" of the market, namely the interaction of supply and demand. In addition, the estimate is likely to incorporate large volumes of information from disparate sources since the individual's participation in the marketplace makes trading decisions based on the information available to them. With a sufficiently diverse pool of active market participants, each with an endowment of information that may or may not be the same as other participants' endowments, the market price can capture the entire universe of available information on a given topic (a result of the EMH discussed in Chapter II).

Second, crowd-based prediction markets are uniquely suited to addressing situations wherein the available information on a given event or outcome is vague, limited, or has gaps. This situation occurs for three reasons. First, individual participants in the prediction market are endowed with different information sets, and thus, it is likely that information gaps faced by one participant may not be faced by other participants or each participant may have different information gaps. Given that the market's forecast is based on aggregate results of individual trading decisions, these gaps may be offset and result in the forecast by the market as a whole being based on a nearly complete set of information.

Third, if all participants in the market face the same information gap, all participants then apply their own mental models and heuristics to close the gap as trading decisions are made, which would result in the market generating an aggregate forecast reflecting the consensus of the market participants in how to close or address the global information gap. This phenomenon has important implications because if the market participants are not sufficiently diverse in information endowments and analytic prowess, the predictive accuracy of the market may be diminished. In the case of gaps that reflect unknown unknowns, the EMH suggests that prediction markets can capture the instincts or desires of the participants through their trading behavior, such that the likely impacts of unknown unknowns are included in the market forecasts.

Fourth, crowd-based prediction markets address the issue of limited feedback in traditional intelligence analysis due to their very nature. The evolution of the price of an asset and associated likelihood of an event as it evolves through time is an instantaneous assessment of the probability of that event at any given point in time until the contract expires (the event does or does not occur). Market participants can watch how the price of the asset is evolving and either use analysis or make educated guesses about why the particular pattern of changes in price is taking place and revise their trading decisions accordingly. The market price is an instantaneous measure of the probability of an event, and its evolution over time is in itself feedback that drives revisions of trading behavior by market participants that reflects new or additional information as it arrives.

Finally, issues related to the availability of input information suffering from the consequences of the "need to know" and limited distribution approaches to information sharing and dissemination, as well as the consequences of the "information is power" thought pattern at each agency, are addressed via the diverse nature of crowd-based prediction markets. If the crowdsponsor choses the crowd correctly, some participants are then "inside the wall" and have the need to know to access closely held sources of information while others do not. The prediction markets-based forecast therefore benefits from participants who have access to privileged information while at the same time incorporating the beliefs and analysis of those who do not have such access. The information set used to drive the market's forecasts thus becomes global and does not suffer from agency-based parochiality or the consequences of a need to know or a limited distribution of information. However, an agency overseeing the development of forecasts using the proposed methodology is still necessary. Under the proposed methodology, the prediction market will be sponsored by an agency or agencies, but the information driving the working of the market will be global and include a myriad of sources outside the sponsoring agency (or agencies),

C. THE NATURE OF THE INTELLIGENCE ANALYTIC CULTURE AND ITS IMPLICATIONS FOR CROWDSOURCED AND PREDICTION MARKETS-BASED ANALYTIC METHODOLOGIES

In terms of analytic process, a successful methodology addresses or otherwise works around the implications of a culture characterized by emphasizing tradecraft to the detriment of more scientific analytic methods, problems related to the role of experts and expertise, cognitive bias, and according to Hare and Collinson, "extreme time constraints; focus on current production; the rewards and incentives"[174] for analysts, norms and taboos, the impact of secrecy, and finally, the analyst's identity and training. Each of these challenges can be addressed by the appropriate methodological design of the crowd-based prediction market and are discussed in detail in the next sections.

174 Hare and Collinson, "Organisational Culture and Intelligence Analysis," 217–218.

1. Implications of the Emphasis on Tradecraft and Expertise

The proposed methodology can address the emphasis on tradecraft, and as a consequence, the lack of transparent, scientific-like analytic processes as follows. If the methodology used to improve predictive accuracy is based on using analytic results as an input, as is the case with crowd-based prediction markets, the underlying method used to generate inputs, be it tradecraft or more scientific methods, becomes irrelevant. This lack of relevance results because the key to the success of the prediction market-based forecasting effort is the input itself, not how it is generated. Market participants can and likely will use different and multiple methods to drive their trading decisions. All market participants may use Intelligence Community tradecraft-based analysis to drive their trading behavior or may use other analytic techniques of varying degrees of scientific rigor to support their input decisions. In either case, the crowd-based prediction market will aggregate results and generate a consensus forecast. The added benefit is that the market forecast will incorporate the results of the application of open-source analytic techniques, which in some ways, may be of equal or greater analytic power than traditional intelligence analytic tradecraft as well.

The role of experts is more problematic. Crowd-based prediction markets address this issue by anonymizing the source of the input data. Given that market participants can only see the aggregate behavior of the forecast embodied in the asset's market price, it is not possible to parse out the contribution of individual participants in the markets whether they are experts or not. Furthermore, the very diversity of prediction markets participants ensures that the analytic sclerosis associated with experts and expertise does not have an outsize impact; market participants have their own mental models and heuristics and all participants' inputs are weighted equally in the market's forecast. Crowd-based prediction markets also offer the opportunity to select market participants based on each individual participant's proven forecasting prowess by using a "warmup exercise" that will be open to all and in which participation in a prediction market tests and validates each participant's predictive skills. Then, the subset truly demonstrated to be superforecasters due to their predictive prowess in the warmup is tasked with making the actual predictions that are inputs to the "real" thing. Finally, psychological and general knowledge tests can be used to screen for those who, according to Tetlock and Gardner are:

> Better at inductive reasoning, pattern detection, cognitive flexibility, and open-mindedness…. [and have a] greater understanding of geopolitics, training in probabilistic reasoning, and opportunities to succeed in cognitively enriched team environments….and [who] viewed forecasting as a skill that required deliberate practice, sustained effort, and constant monitoring of current affairs.[175]

Thus, identifying and selecting those with personal characteristics is strongly correlated with forecasting prowess for participation in the actual prediction market.

2. Implications of Cognitive Biases

Cognitive biases can be addressed by applying the crowd-based prediction markets methodology via training, the choice of participants (utilizing experts or not), and by having the methodology rely on the nature of probability and statistics to address biases. Prior to participation, prediction market participants can undergo a short period of training in which they receive education in basic probability and statistics and become familiar with typical cognitive biases and how to avoid them. The literature reveals this comparatively cursory level of training to increase market superforecasters' predictive prowess. Superforecasters who have undergone this level of training often have forecast accuracies that often exceed those of professional analysts by 30 percent or more.[176]

To address the issues with experts and expertise, the proposed methodology may include a diverse pool of prediction markets participants that includes, but is not dominated by, those the Intelligence Community regards as experts. Indeed, the performance of the experts as opposed to lay forecasters and superforecasters in the prediction market may allow the development of weighting schemes whereby the forecasts of market participants with superior forecast accuracy have a greater weight in the market. Participant diversity minimizes the impact of cognitive biases in a prediction markets-based forecasting methodology. A sufficiently diverse prediction market will

[175] Tetlock and Gardner, *Superforecasting*, 181.

[176] Ungar et al., *The Good Judgment Project*, 38–41.

contain participants whose cognitive biases are not all the same or even moderately correlated. As a result, the impact of the biases should offset each other and result in the market's aggregate forecast converging to a value minimizing the impacts of cognitive biases. Finally, combining forecasts from different prediction markets or from the prediction market and traditional intelligence community analysis should lead to significantly improved forecast accuracy. This aim will be achieved by having the methodology use multiple forecasts from different sources as inputs to average out the impact of cognitive biases as long as the input forecasts are independent and uncorrelated.

3. Implications of Time Constraints and the Tyranny of Production

The methodological design of the crowd-based prediction market can address the impact of severe time constraints. Prediction markets produce results reflecting the latest information at a given point in time, and any one-time snapshot of its outputs will be just that, a snapshot in time. The continuous nature of the prediction market's analytic results over time means that the analysts can simply look up the current state of a prediction as often as they like, which eliminates the impact of time pressure that results with one-time products or with periodic products prepared on short timelines. This approach also addresses the tyranny of the need to produce and update products constantly in real time because the forecasts embodied in the market evolve constantly as new information arrives.

One potential criticism of the prediction markets-based approach is that the tasks are seemingly binary; an event happens or does not happen, and the price of that event or asset in a prediction market is an instantaneous estimate of the likelihood of that event at a given point in time. This approach seemingly implies that the complexity of analytic tasks that can be performed by a prediction market is limited to simple binary tasks, which is not true for the following reasons. First, complex tasks can be modular and nested, in that they can be decomposed into individual tasks each reflecting binary choice that aggregates up into an overall forecast. Second, the asset can be designed to reflect conditionality; if event A happens then event B happens with a certain probability. Third, tasks can be set up to assess the likelihood that a threshold will be exceeded. For example, what is the probability of a North Korean missile with over a 3,000 nautical mile range? Generally speaking, appropriate task design can address the issue of task complexity when using a crowd-based prediction markets forecasting methodology.

4. Implications of Rewards and Incentives

A crowd-based prediction markets methodology also addresses the issue of rewards and incentives. Explicit, rigorous, backward-looking (quantitative not qualitative) analysis of the accuracy and relevance of forecasts in a public manner is allowed. Market participants can earn non-monetary rewards (mission badges, points, game rankings etc.) as a direct consequence of their forecasting success. In this way, tying each market participant's incentive structure to each forecast is encouraged, which leads to analytic accuracy and relevance. Chapter VI expands on how the thesis methodology operationalizes this concept.

5. Implications of Taboos and Secrecy

Given the information and analysis Chapter II presents, a complete methodological solution to the issue of the taboo against changing, reversing, or otherwise straying from the agency position, regardless of new or even contradictory information, is unlikely. A cultural change on the part of policy makers and agency managers and leadership addressing the fact that a well-reasoned and supported change in position actually increases predictive accuracy is necessary to overcome this obstacle.

The proposed methodology addresses the issue of secrecy and the primacy of secret or covert information in analysis because prediction markets inputs are analytic results, and the source of those inputs is irrelevant to the market outcome. The market does not care what drives an individual participant's trading decisions; just that the decisions are made and acted upon. Thus, those with access to secret information and analysis can use that information and analysis to guide their trading behavior, while others can rely on open source or even commercial information as an input to drive theirs. The market then aggregates information from all sources available to market participants, whether based on closely held information or not, and generates a consensus estimate.

The forecast aggregation feature of prediction markets may also allow the use of weighting schemes whereby the forecasts made by market participants with a quantitatively demonstrated track record of forecast success can be given greater weight in the market. Thus, if those participants with access to secret information and analysis do indeed demonstrate a better track record of success, their inputs can then be more heavily weighted as the market aggregates its forecast. This approach has the disadvantage of potentially causing bias in the forecast, as the forecasts of experts with access to secret information are given greater emphasis. However, the efficacy of such weighting schemes can be tested as the Intelligence Community implements the methodology, and if successful, the community can replicate the weight schemes. Such a precedent already exists in the financial markets. By their nature, high volume (in dollar value) traders have an outsize impact in determining the course of the market, as smaller investors attempt to follow the market leaders, and as algorithms that base their trading behavior on trading patterns in the market as a whole do their work.

6. Implications of Analysts' Identity and Training

Issues related to the intelligence analysts' identity are also hard to address using a crowd-based prediction markets methodology. The issue of identity is exacerbated by the fact that generalists (some of whom are outside the Intelligence Community) probably have an active role in operationalizing the methodology. Furthermore, the possibility that non-experts can outperform the experts within the Intelligence Community (see the section on the Good Judgment Project in Chapter II) will exacerbate this problem. An appeal to the agency mission of providing the most accurate information possible may address some of these issues.

The issue of training is similar to that of identity in its implications. Since the best forecasters benefit from formal training in probability and statistics, and in how to address the potential for cognitive biases,[177] limited formal training may be one way to address this issue. Furthermore, if the methodology uses analytic results as an input, the underlying differences in training, and as a result analytic approaches, becomes irrelevant

[177] Mellers et al., "The Psychology of Intelligence Analysis," 6–12.

as the methodology is a black box that takes the results of analysis as its input and produces predictions as an output without having to consider the issues caused by different training standards and practices.

D. ADDRESSING THE INTELLIGENCE CONSUMERS' LIKELY CONCERNS

Consumers of intelligence products derived from the results of crowd-based prediction markets very likely would have concerns about how the products were developed and their reliability. These concerns would likely arise from the following:

- Although sponsored by an agency or agencies, in some sense, no one and everyone develops the prediction markets-based estimates; the prediction market's estimate is not traceable to specific analysts or to specific sources or methods. Obviously an issue, the consumer of the intelligence forecast would likely require the estimate to a source or sources be traced for purposes of accountability.

- The information set prediction markets participants use to develop the estimate is in some sense unknown.

- Limited tools are available to generate confidence measures for the estimates.

- The estimate may radically conflict with existing agency positions.

Each of these issues can be addressed by educating the consumer. Consumers have to be convinced of the utility, efficacy, and applicability of prediction markets-based estimates to have them sponsor the prediction markets as described in Chapters VI and VII. First, prediction markets are just one technique. Analysts should create a mosaic using the results of different techniques weighted by a track record to develop the products being presented to the consumer. Consideration of the impact of combining independent forecasts in Chapter II of this thesis demonstrates the robustness of this approach. Second, given that a prediction market can be designed to aggregate all available information on a topic, including open source and non-public information, it is likely that the prediction market estimate is a better representation of the global consensus view on a given question than analysis performed by a single or a few analysts using a limited set of analytic techniques. The track record in terms of prediction market accuracy in other contexts supports this view as well. Furthermore, the lack of traceability to specific individuals, sources, or methods is actually a positive in that prediction markets address the biases inherent in basing the analytic process on limited information sets or a few analysts and analytic techniques. The confidence level of the prediction markets-based estimate can be approximated based on volatility over time of the prediction market estimate itself or by comparing it to the historical track record. Finally, in the cases in which the prediction market results differ from long-held agency positions, such disagreement can be a warning that the global consensus on a topic is different from the agency viewpoint and drive further analysis using non-prediction markets-based techniques. In sum, prediction markets-based estimates are yet another tool in the analysts toolkit; albeit one with great potential.

E. CONCLUSION

This chapter has discussed how crowd-based prediction markets methodologies address and potentially solve the problems posed by many of the culturally driven sources of intelligence community forecast and analytic inaccuracy Chapter II first identifies. The next chapter discusses the specifics of a crowd-based prediction markets forecasting methodology that drives improved intelligence community forecasting and analysis that builds on the work of Chapters III and IV. Note that the methodology proposed in the next chapter generates probabilistic forecasts that are then incorporated into usable intelligence community products, such as narrative reports, numerical charts and tables, spreadsheets and graphs, spot advisory “flash” reports, and status boards.

THIS PAGE INTENTIONALLY LEFT BLANK

VI. A METHODOLOGY FOR IMPROVING FORECAST ACCURACY

A. INTRODUCTION

This chapter describes a proposed methodology for improving intelligence community forecast accuracy. The work in the previous chapters centered on the implications of combining forecasts, the importance of Intelligence Community culture for analytic methodologies, and on crowdsourcing and predictions markets; all of which are inputs for the proposed methodology. In particular, this chapter discusses the characteristics of the proposed crowd, the proposed structure of the forecasting effort, the proposed incentive structure, the proposed task design, the proposed prediction market design and associated structural parameters underlying the forecasting effort, and the key characteristics of the proposed platform used to implement the prediction market. The goal of this chapter is to use all these critical concepts to design a methodology—a crowd-sourced forecasting tournament—that can be used by the U.S. Intelligence Community to improve its forecast accuracy. The first section provides an overview of the tournament and a real-world example of the application of the tournament in an intelligence context, while later sections describe each major aspect of the methodology previously mentioned in detail.

B. A CROWD-SOURCING TOURNAMENT

The existing research on crowdsourcing and prediction markets strongly suggests that the Intelligence Community can improve its forecast accuracy by sponsoring a crowdsourcing tournament in which a specially recruited group of individuals (the crowd) compete to make forecasts about future events of strategic significance.

During the tournament, the crowd uses their assessments about the intelligence question posed by the tournament sponsor as inputs to their trades on a prediction market on an asset embodying the sponsor's question (see Chapter IV). The likelihood of a given outcome (or answer to the question posed) at any point in time until whether or not the event has happened can be assessed is given by the instantaneous price of the contract embodying that event in the prediction market. The structure of the proposed tournament and the underlying prediction market this chapter discusses is adjusted to fit the nature and culture of intelligence agencies and personnel (discussed in Chapters II and V).

The utility of this approach to forecasting can be illustrated using an example. Suppose someone wants to understand the current state of the North Korean nuclear program. Given recent events, potential questions to the crowd may include:[178]

- Will President Kim meet with President Trump in the next three months?
- If a summit between the presidents occurs, what are the implications of the summit for North Korea's nuclear program?
    - claim of complete denuclearization by North Korea
    - phased denuclearization by North Korea
    - admission of independent verifiers to verify denuclearization by North Korea
    - summit failure; no agreement reached
- Will the North Koreans resume testing nuclear weapons?
    - if the summit fails
    - breakout after making an agreement at the summit
- When will they resume testing nuclear weapons?
- If they resume testing, what size will the test be?

In each case, the crowdsourced prediction market provides a likelihood of these events coming to pass. Note that this set of questions does not address the why's of the various courses of action listed. To elicit these responses, the questions must be modified to provide options as to the underlying reasons for the event. Take summit failure, for example. The summit can fail for many reasons, and the plausibility of alternative

[178] This example was drafted before the Trump-Kim summit in Singapore in June 2018.

reasons for summit failure can be tested using the prediction market. One example to consider in this context is:

- The summit between the two presidents fails because:
    - President Kim withdraws from negotiations:
        - China pressures Kim not to take the deal on the offer.
        - Meaning of denuclearization to both sides is different.
        - Insufficient economic incentives are given to Kim.
    - President Trump withdraws:
        - North Korean terms are unacceptable.
        - Requires time-phased denuclearization.
        - United States asked to exit the Korean Peninsula.

In this case, the crowd-based prediction market assesses both the overall odds of summit failure, as well as gives rank-ordered reasons as to why the summit has failed. Choices different from those in the question design can be captured. In addition, another set of questions can be posed in a similar vein that probes for the evidence supporting each assessment made.

The methodology this thesis describes is both distinct from and an extension of other efforts in this area, such as the Good Judgment Project due to the following.

- The methodology uses a warm-up tournament to select the best forecasters from the crowd both within and outside the Intelligence Community, and possibly, subsequently limits participation to these superforecasters.

- The methodology uses iterative tournaments with periodic redesign. After the initial warm-up tournaments, sponsors redesign subsequent tournaments as needed and iterated on a six-month cycle.

- Sponsors adjust the choice of crowd in each iteration of the tournament based on actual forecasting success.

- The methodology tests what drives superior forecasting (question design, choice of crowd, teaming, training, experience, education, etc.) and upon identification of key drivers for forecasting success, iterates prediction market parameters as a whole to take advantage of the testing results.

- The methodology uses questions (prediction market asset or contract) designed to elicit not only the likelihood of an event but also both a rank-ordered list of underlying reasons for that event, as well as an assessment of the evidence supporting the reasons given.

- The methodology balances the need for simplicity against the need for complex question design by using a survey like frontend to elicit participants' assessments about the questions posed.

- The methodology makes extensive use of gamification.

- The methodology allows for teaming.

- The methodology uses commercial-off-the-shelf platforms to support and manage the prediction market.

This chapter proposes that the Intelligence Community use rolling six-month tournaments with an initial (practice) warm-up tournament followed by subsequent iterations of the actual tournament itself. The warm-up tournament serves both to familiarize participants with the structure and operation of the prediction market and to identify the best forecasters to be selected to participate in subsequent tournament rounds.

During each round of the tournament, participants trade contracts or assets embodying questions about real-world outcomes of events on a prediction market. Participant's performance (relative forecast accuracy) in each round may in part determine whether the participant is invited back to participate in subsequent rounds. During each round, statistical testing is used to identify likely drivers of realized forecast accuracy and that information is used to inform redesigns of crowd choice and of both the question and prediction market design to improve forecast accuracy further. The added advantage is that the relative merits of each component of the methodology can be understood fairly quickly and real-time redesigns done to address the inevitable frictions seen in the real-world possible.

Finally, the relative merits of the methodology in improving forecast accuracy becomes clear very quickly. When measured over two rounds beyond the initial warm-up round, if the iterative redesigns are not significantly improving forecast accuracy compared to traditional open or closed source forecasting methods, then the effort is quickly halted before it uses undue resources.

C. CHOOSING THE CROWD

Since the proposed methodology employs crowd-sourcing, the choice of who is in the crowd is critical to forecast accuracy. For the methodology to be effective, the crowd must be diverse, comparatively large, and well-motivated. The crowd should include "non-experts, experts, informal members, customers, business partners, etc.," and it must vary in "size, composition, uniformity, and level of expertise."[179] Also, given that tasks are designed for individuals and for teams, choices in terms of how teams are formed are also critical.

Diversity in the crowd can be addressed by targeting multiple markets segments for the crowdsourcing effort envisaged under the methodology. Obvious communities to be targeted as part of implementing the methodology include intelligence analysts, business intelligence analysts, journalists, think tank staff, and political science and international relations students and faculty. Less obvious target communities include students and faculty in the social science community (e.g., economics, anthropology, language), engineers, health care professionals, epidemiologists, weather forecasters, etc. The lay public should also be included.

The methodology should encourage teaming, and participants should be allowed to participate in implementing the methodology both as teams and as individuals. This team participation can involve participants either forming their own teams, or by

[179] Chiu, Liang, and Turban. "What Can Crowdsourcing Do?" 43.

expressing a willingness to participate on teams formed by the group implementing the methodology. It is critical that the crowdsourcerss sustain the outreach effort as the forecasting tournament proceeds. Sufficient diversity and largeness of numbers throughout each forecast cycle and the continued engagement of superforecasters, identified as the tournament proceeds, is critical to successful forecast elicitation.

Part of the outreach to crowdsourcees envisaged in this methodology's implementation requires gathering detailed data on the potential crowd using survey tools for such issues as geographic location, areas of interest, gender, level of education, type of education, subject matter knowledge, degree of analytic expertise, degree of formal training in probability and statistics, years on the job, etc. Other areas the crowdsourcers should consider include self-perceived competence, comfort with collaboration, motivation to volunteer, degree of intrinsic and extrinsic motivation, drivers of mental satisfaction, degree of self-perceived self-esteem, level of development of personal skills, degree of comfort with knowledge sharing, and degree of love of community. During the warm-up tournament, the crowdsourcers use statistical testing to identify the demographic and psychographic attributes of forecasters associated with forecast accuracy using the completed surveys. This testing may be used to inform the selection of participants in subsequent rounds of the forecast tournament. This testing also enables the implementers of the methodology to assess the diversity of the potential crowd and make adjustments in terms of increased outreach to given market segments or task reformulation or elimination as needed. Ideally, participants exhibit sufficient diversity for each task, while at the same time, the group is large enough to elicit meaningful results.

Under the methodology, members of the crowd can choose which tasks they participate in, both individually, and as part of a team. Indeed, if it turns out that an insufficiently diverse crowd "opts in" to a given task, the crowdsourcers may either drop the task from consideration or re-formulate the task and task design to make it appealing to a more diverse crowd.

D. TOURNAMENT STRUCTURE

Six-month forecast cycles form the overarching structure of the forecast elicitation methodology. At the beginning of each forecast cycle, individuals and teams are endowed with a stock of play money with which to buy into contracts associated with tasks (each task or task set has a traded contract associated with its outcome). Earnings from previous forecast cycles (the portfolio value at the end of a given cycle) are added to this endowment to allow for forecasters not successful in making forecasts in a given round to participate in subsequent rounds, as well as those who have been successful (have a positive portfolio value at the end of a forecast cycle). This result is important for three reasons. First, uninformed or unsuccessful traders add information to the prediction market regardless of their success or failure. Second, competition is engendered, as well as perhaps a willingness to keep participating even after an initial failure. Third, a track record of success leads to greater rewards.

Note that prior to the start of the tournament proper, individuals and teams can participate in a six-month warmup forecast cycle. The crowdsourcers use the performance in this cycle to identify and select potential superforecasters and to create teams of superforecasters to play in the forecast tournament proper. The warmup tournament also allows the crowdsourcers to work through alternative task designs, how to opt in or opt out of tasks (they only have to buy or not buy or short a given contract associated with a task), and identify any task design related issues prior to starting the tournament proper. At this time, participants also are given the chance to take basic training in probability and statistics and to complete the psychographic questionnaire, which is used to assess adequate diversity on each task. In addition, participants become familiar with the tournament platform, how to trade on their predictions, and the reward structure and game elements. The methodology assumes a certain amount of financial acumen regarding how financial markets work in terms of buying and selling, market and limit orders, and shorting and buying on margin, but the crowdsources provide an online self-paced training and instructor-led webinars on these topics.

E. INCENTIVE STRUCTURE

The proposed methodology gives paying individuals play money for forecasting success as a measure and metric of an individual or team's success. The choice of play money as an incentive is deliberate as the literature has not demonstrated a difference in outcomes that depends on whether real or play money is used. The perception that the Intelligence Community is paying participants to gamble and all the consequences the realization of that reality entails is avoided. At the end of each forecast period (every six months), the crowdsourcers assess value of an individual or team's play money forecast portfolio and the individual (or team) can redeem it for nominal real prizes. The five best forecasters, as measured by portfolio value, also receive pro-rated shares (with a 5–4-3-2-1 prorating scheme) of a play money for X prize. The forecasting rounds (six-month tournaments) are also gamified to the extent possible. The gamification takes the form of an individual and team play money portfolio value leaderboard. Individuals and teams earn play money, not only based on the value of their forecast portfolio, but also on their degree of participation in each round and on the complexity of the task they undertake (more complex forecasting tasks pay a variable bonus depending on complexity for forecasts with one standard deviation of the reality).

A final game element consists of different missions that crowdsourcees can undertake. For example, envision a North Korea mission area, for which crowdsourcees can earn mission badges for participating frequently and effectively in a given number of forecasting activities related to North Korea. Alternatively, a mission area for each area of subject matter expertise can be created. Moreover, crowdsourcees can undertake multiple missions with play money bonuses as awards for mission completion.

Leaders in the Intelligence Community may initially object to the notion of "intelligence forecasting as a game," as the stakes are so high in terms of how consequential Intelligence Community forecasts are. As referred to in several sources, the counterargument to this objection is based on the demonstrated ability of prediction markets to elicit highly accurate forecasts in other contexts and on the reality that the success or failure of the methodology in improving forecast accuracy is both unambiguous and quickly clear. Furthermore, an argument can be made that the gamification of tasks as proposed in this methodology has cracked hitherto intractable problems like optimal protein folding in the biosciences and the identification of certain astronomical objects in astronomy.[180]

F. TASK DESIGN

Task design is critical to the success or failure of the methodology. Proper task design allows independently determined multiple forecasts to be created and forecasts to be combined to improve accuracy in play. Furthermore, it is then possible to determine what works and what does not in terms of task design as the results from the crowdsourcing effort come in; thereby, the operationalized task design can be revised to engender the best forecast results.

Crowdsourcers may design tasks to be atomic (that is indivisible) and may involve varying degrees of complexity both in terms of the required results and the complexity of analysis necessary to formulate those results. Tasks, such as estimating the probability of a nuclear weapons test by North Korea (for example), are on the surface atomic; the crowdsourcees only have to estimate the likelihood based on their judgment. However, if the task is broken down into subtasks, each of which requires an independent estimate as an output (and as an input into the next higher order task), then the formerly atomic task becomes much more complex. For example, if instead of asking what the probability of a nuclear test is within a given timeframe, the task instead estimates the conditional probability of a nuclear test's kilo-tonnage, exceeding a given threshold if the test occurs, and if the test occurs and exceeds a given threshold, that test is conducted using a missile, then the formerly atomic task has become granular.

Thus, tasks in the proposed methodology consist of participants estimating "one shot" probabilities of individual, unique events, estimating conditional probabilities for event sequences, estimating when an outcome exceeds a threshold, and estimating the

---

[180] "Solve Puzzles for Science," Fold-it, accessed June 12, 2018, https://fold.it/portal/; Alan Boyle, "Gamers Solve Molecular Puzzle that Baffled Scientists," *NBC News*, November 2, 2015, https://www.nbcnews.com/science/science-news/gamers-solve-molecular-puzzle-baffled-scientists-f6C10402813; Matias Celasco, Juan Ignacio Yanez, and Roberto Gamen, "Galaxy Conqueror: Astronomy, Citizens, and Gamification," in *2016 XI Latin American Conference on Learning Objects and Technology (LACLO)* (San Carlos, Costa Rica: IEEE, 2016), doi: 10.1109/LACLO.2016.7751798.

likelihood and extent of forecast errors. The payoffs of individuals' tasks are suited to a winner-takes-all approach in this case. If the task involves determining the mean value of an outcome (e.g., how many kilotons equivalent a given nuclear test is or the median value for the amount of counterfeit currency manufactured by North Korea is in circulation), the crowdsourcers then incorporate the indices and spreads, respectively, into the task design (in other words, the payoff for the successful completion of the task is measured against the extent to which the predicted value exceeds an index or the size of the spread between the predicted value and the expected value, respectively).

Crowdsourcers can also design tasks to be allocated to both individuals and teams. Although the warmup tournaments proposed as part of the methodology identify superforecasters within each crowd segment, it may also be possible for teams of superforecasters to perform even better than individual superforecasters alone. To this end, in the case of individual tasks, no coordination or communication with others is required, as opposed to team tasks where teams coordinate and post a team consensus answer to estimation tasks. Teams are better suited to address complex analytic tasks (ones that involve several conditional events or ones that involve estimating spreads or multiple simultaneous thresholds), both due to the level of effort complex analytic tasks require and the complexity of analysis they require to generate inputs. On the other hand, crowdsourcers can design individual tasks to be as atomic as possible, which allow individuals to address credibly the task without an excessive use of resources, such as time and analytic effort. Two potential useful outcomes can result. First, in the case of tasks for individuals, all participants can apply their own expertise and knowledgebase to making the estimate independently of others in the effort. Thus, multiple independent estimates can be created and combined, which then results in increased forecast accuracy on average. Second, in the case of team tasks, depending on the choice of team composition, information that may be unavailable to the many may be incorporated into the team estimates, as each team member may have access to different knowledge and experience bases (e.g., classified information and the analytic process used in intelligence analysis), and the team estimate itself may be less likely to suffer from cognitive biases due to the diversity of team members' interactions, experience, and inputs.

Finally, crowdsourcers can design the degree of structuring of tasks to vary to allow both individuals and teams to test existing analytic approaches. In the case of structured approaches, the task is broken down into a set of discrete subtasks, the results of which serve as inputs to higher order subtasks, which in turn, serve as inputs to the final result. As far as possible, the structuring of the tasks and subtasks reflect analytic best practices of the Intelligence Community. Crowdsourcers address unstructured tasks by having the participants assess the top line question, without having to go through a series of structured subtasks first that then tests the impact of analytic design on forecast accuracy.

G. PREDICTION MARKET STRUCTURAL PARAMETERS

The incentive structure under this methodology requires the implementation of structures related to performance measurement, rewards, and involvements. Crowdsourcers evaluate performance based on the value of an individual or team's play money portfolio. Portfolio value consists of the value of assets being traded or expired options (over 80 percent), play money awards based on effort (number of trades, training completed) (five percent), and play money awards based on mission completion and based on task complexity (15 percent). Hopefully, the methodology will then ensure sufficient incentive to motivate engagement, that the level of work required for participation is manageable, and that adequate reciprocity occurs between the crowdsourcers and crowdsourcees.

The trading process used under the methodology is as follows. First, fees for trading and IPOs, and an expiration are not imposed. As discussed in Chapter IV, the impact of implementing trading fees (primarily to prevent bubbles) is not well understood, and the fact that most extant and past prediction markets do not impose such fees suggests that the methodology need not operationalize this aspect of prediction market design. Second, to encourage participation, crowdsourcees are allowed to trade 24/7, both enabling ease of participation and also encouraging prompt price discovery. Third, due to liquidity concerns (thin trading may be a possibility on some tasks), a DPM trading mechanism should be used. Finally, the crowdsourcers or crowdsourcees can generate new prediction tasks through the IPO process. In the case of crowdsourcers, they simply add another contract to the market traded at some IPO price that can be determined by a variety of methods, including expert judgment, consensus forecasts, and the like. In the case of crowdsources, new forecasts are accommodated by having players propose an IPO. If players place sufficient orders for the asset (prediction) as defined by the crowdsourcers, the IPO takes place. IPO prices in this case are fixed by the market maker or discovered by auctioning the asset on the market. Table 10 lists the clearinghouse processes under the methodology.

Table 10. Methodology Clearinghouse Processes[181]

Clearing House Processes		
Order	Order matching rules	Price and quantity
	Order spending caps	No caps
	Order type	Market Limit
	Short selling	Allowed
Asset	Asset type	Play money
	Borrowing	Margin purchases
	Endowment	Initial endowment (Money and/or Contracts)
Claim	Claim IPO	Rewards Screening
	Initial claims	Quantity

181 Adapted from Gaspoz, *Prediction Markets Supporting Technology*, 83–85.

Clearing House Processes		
	Claim ontology	
	Claim type	Winner take all Conditional Index Spread
	Claim structure	Independent
Payoff	Settlement date	Public
	Settlement judge	
	Settlement price	“Truth”

As far as possible, the clearinghouse process the methodology uses should attempt to mimic that of real-world financial markets for several reasons. First, participants' existing knowledge and how they work is leveraged. The analogy to real-world financial markets should make process related issues like buying, selling, buying on margin, and shorting easier to explain to participants. Second, when participants buy through buying on margin and shorting, it increases the efficiency of the market in theory to allow traders to trade on information that cannot be traded on using a simple buy or sell order. Similarly, the allowance of market and limit orders lets traders act efficiently on active market trends, whether or not they are actively on the trading platform. Third, by providing endowments each cycle, even comparatively unsuccessful traders can continue to trade across tournament cycles; remember that even unsuccessful traders add information to the market. Fourth, by using a DPM-based market maker, the crowdsourcers can ensure liquidity in comparatively thin markets with no risk. Fifth, by allowing winner-take-all, conditional, index-based, and spread-based contracts to be traded in the market, the crowdsourcers can accommodate a diversity of task designs easily. Sixth, the market having a public settlement date and an agreed settlement judge ensures that both the market and the settlement process are transparent and enjoy a clear understanding of forecast stating what success is. Finally, having "truth" as the settlement price (the contract has a value of $1 or $0 at expiry, given the success of the forecast as determined by the settlement judge) allows a transparent way to determine the impact of forecast success or failure on a player's portfolio.

The trading process under the proposed methodology, as shown in Table 11, is characterized by no trading fees, an open order book, and the use of a DPM market maker. Trading fees in the methodology are not used because the effect of trading fees on prediction markets has not been studied in detail in the literature and because most extant prediction markets do not charge any of the types of trading fees.[182] The researcher chose an open order book with 24/7 trading for reasons of transparency and to encourage informed, active trading. Finally, the researcher chose DPM market makers because they maintain the liquidity of the market, even in thin markets, and because they "attain the highest forecasting accuracy, good robustness against parameter misspecification, the ability to incorporate new information into prices, and the lowest losses for operators."[183]

Table 11. Methodology Trading Process[184]

Trading Process		
Fees		None
Trading Mechanisms	Order Book	Open Order Book
	Market maker	Dynamic pari-mutuel

[182] Gaspoz, *Prediction Markets Supporting Technology*, 88.

[183] Slamka, Skiera, and Spann, "Prediction Market Performance," 169.

[184] Adapted from Hosseini et al., "On the Configuration of Crowdsourcing," 27–45.

Trading Process		
	Trading time	24/7

A coherent set of rules for trader management is essential to the forecasting success of the methodology. First, by definition, the market is a closed market (open by invitation only) because it is crucial to gather and engage crowds with well understood demographics, skill sets, expertise, and psychographic profiles. In addition, some level of training is necessary so that the engaged crowd is motivated and appropriately incentivized. Second, as far as possible, the market should be fully transparent. The argument is that the market should be consistent with semi-strong form efficiency, and since opportunities may be limited, by definition, to trade on insider information. Third, in the methodology, a degree of anonymity is allowed. Participants are known by username, so that the gamified aspects of the market work effectively. However, participants' actual identities are masked to prevent external concerns from inhibiting behavior. Investor unicity (not having multiple accounts) is also essential as the proper measurement of forecasting success requires it and because a lack of unicity may allow manipulation or misbehavior by participants. Finally, the management of the crowd selection process encourages diversity and largeness on parameters, such as demographic, psychographic, expertise, information access, experience, and education measures. Table 12 addresses trader management issues.

Table 12. Trader Management Characteristics[185]

Trader Management		
Market	Market policy	Closed market
	Market transparency	Display all information
Investor	Investor anonymity	Username
	Investor unicity	Enforced
	Investor selection	Quantity
		Diversity
		Informed versus non-informed

H. PREDICTION MARKET PLATFORM CHARACTERISTICS

Multiple commercial software platforms are available to use for implementing the methodology, including CrowdWorx, GNOSIS, Augur, and Inkling. Table 13 lists all these tools capable of performing the functions required of any platform implementing the methodology.

[185] Adapted from Gaspoz, *Prediction Markets Supporting Technology*, 83–85.

Table 13. Trading Platform Characteristics[186]

Facilities	Short Descriptions
1. Crowd-related Interactions	Facilities in the platform that relate to the crowd
1.1. Provide enrolment	Means to enroll the individuals
1.2. Provide authentication	Means to authenticate the individuals
1.4. Provide task assignment	Means to assign tasks to the right individuals
1.5. Provide assistance	Means to help the individuals during the performing of the task
1.6. Provide result submission	Means to help the individuals to send their results
1.7. Coordinate crowd	Means to coordinate performers in a certain task
1.8. Supervise crowd	Means to supervise individuals during their performance
1.9. Provide feedback loops	Means to give feedback to individuals about their performance and about the results
2. Crowdsourcer-related Interactions	Facilities in the platform that relate to the crowdsourcer
2.1. Provide task broadcast	Means to broadcast the task to the right individuals
2.2. Provide assistance	Means to help the crowdsourcers for announcing the task
2.3. Provide time negotiation	Means to help crowdsourcers negotiate time requirements with the individuals
2.7. Provide result verification	Means to verify whether submitted results meet the needs of crowdsourcers
2.8. Provide feedback loops	Means to give feedback to crowdsourcers about individuals' performances
3. Task-related Facilities	Facilities in the platform that relate to the task
3.1. Aggregate results	Means to collect and unify submitted results
3.2. Hide results from others	Means to hide individuals' results from each other for privacy reasons
3.3. Store history of completed tasks	Means to keep a history of the completed tasks and related information (such as who completed them, the spent time, etc.)
3.4. Provide quality threshold	Means to guarantee the required quality of results

186 Adapted from Hosseini et al., "On the Configuration of Crowdsourcing," 27–45.

Facilities	Short Descriptions
3.5. Provide quantity threshold	Means to guarantee the required number of responses
4. Platform-related Facilities	Facilities in the platform that relate to the platform itself
4.1. Online environment	Means to keep the platform online and accessible to individuals
4.2. Manage platform misuse	Means to report if there are instances of platform misuse
4.3. Provide ease of use	Means to keep the platform simple to use
4.4. Provide attraction	Means to keep the platform attractive to use

I. CONCLUSION

This chapter has presented a description of a proposed methodology to improve intelligence analysis. The methodology is derived from work in previous chapters of this thesis, including work on (1) the implications of combining forecasts for accuracy, (2) a discussion of intelligence community culture in general, (3) a discussion of crowdsourcing and crowdsourced effort design, (4) a discussion of predictions markets and prediction markets design, and (5) a specific discussion of the implications of intelligence community culture on analytic methodologies. The chapter also discussed key aspects of the methodology, from the choice of crowd, the structure of the prediction market tournaments, participant incentive design, task design, prediction markets implementation and key prediction markets platform parameters. Although the details of how this proposed methodology is implemented and tested are in some sense part of the methodology itself, that aspect of the methodology is discussed in the next chapter.

VII. IMPLEMENTING AND TESTING THE METHODOLOGY AND DIRECTIONS FOR FURTHER RESEARCH

A. INTRODUCTION

For the proposed methodology in this thesis to have relevance for its target audience, Intelligence Community analysts and analytical management, someone must implement and test it, as well as demonstrate the extent of its ability to improve the quality of intelligence analysis. A full-blown test may not be justified at this point; rather, this thesis recommends an incremental approach starting with implementing and testing the methodology on a small-scale pilot. First and foremost, the utility of the methodology needs to be demonstrated to the analytic branches of the intelligence to get buy-in to the methodology as a whole, as well as to engender participation in the methodology's prediction market from individuals and teams drawn from the analytic community. Second, the results of the pilot, if positive, can be used to justify the Intelligence Community spending the financial, analytic-time based, administrative-time based, and other resources to implement the methodology. Third, the pilot should allow various aspects of the methodology, from outreach to task design, to be tested and tweaked to ensure that the methodology as implemented does indeed result in the analytic improvements being sought. Therefore, this chapter discusses practical aspects of setting up and implementing the pilot, practical aspects of scaling up the pilot to full-scale, and how both the pilot and the full-scale methodology is evaluated.

B. IMPLEMENTING THE PILOT

Implementation of the pilot must address (1) project sponsorship, (2) project outreach, (3) provision of participant training, (4) incentive implementation, (5) task design process, (6) task limitations, (7) prediction market software selection and configuration, and (8) pilot program operation and administration, including warmup tournaments, selection of superforecasters, teams and teaming, task success determination, and provision and timing of feedback. Each of these implementation factors is discussed further.

1. Project Pilot Sponsorship

IARPA has sponsored several prediction markets-based attempts to produce high quality forecasts in the past through its ACE program in the past. The pilot differs from these prior efforts in three ways: (1) how it uses teams and teaming, (2) tasks and task design, and (3) the level of support provided to participants in the pilot in terms of training in probability and statistics, cognitive biases, and the workings of financial markets and financial instruments and by the analogy prediction markets assets, contracts, and instruments is far greater than in these earlier prediction markets-based forecasting methodologies and tests. All these differences are associated with increased forecasting accuracy in the literature. This researcher hopes that given the strongly positive results of the ACE associated Good Judgment Project, IARPA should be willing to provide the resources to explore methodologies like the one proposed in this thesis further.

Alternatively, or if IARPA resources only cover part of the resource needs required for the pilot, the pilot should use a partnership approach. In this case, a supremely credible academic sponsor (a Tetlock or a Mellars in stature in the field) should take on the pilot and approach other leaders across the relevant academic and professional communities to help address participant outreach, participant incentivization, and participant training needs. It should also partner with one of the prediction markets software vendors to allow access to their platforms for use in the pilot. The sponsor should also create and manage teams of graduate or undergraduate students to address tasks and task design, administration of the prediction markets platform, and pilot program test and evaluation. In this way, the pilot will be run and tested on a shoestring budget.

2. Project Pilot Outreach

As the previous sections of the thesis discuss, getting the right crowd to participate in the pilot is crucial to its potential success. Indeed, the pilot sponsor must to implement a large, successful, scalable outreach effort. Additionally, sponsor outreach efforts must calibrate to deliver certain crowd characteristics, such as crowd diversity, crowd motivation, large-crowd size, crowd scalability, crowd participation, and crowd responsiveness to incentives on offer.

From an internal to the Intelligence Community perspective, the level of management commitment to the pilot will probably drive participation rates, the degree to which the pilot accommodates the cultural factors Chapter IV addresses, and on the perceived value of the pilot. The proper choice of sponsor (e.g., IARPA or a suitable academic leader) may help drive management buy in to the pilot and thus the degree to which management motivates individual analysts to participate. Appropriate accommodation with and exploitation of Intelligence Community cultural factors can also aid internal participation. A key issue is that the pilot will be testing more than just forecasting the simple probability of an event. Rather, more nuanced forecasts based on conditional events, mean value-based events and threshold-based events are part and parcel of the task designs to elicit forecasts. Note that the value of the pilot to intelligence analysts is to provide an alternative way to address nuanced questions independently and thus also provide useful information to their private (inside the Intelligence Community) forecasts.

From an external to the Intelligence Community perspective, sponsor outreach effort must address crowd diversity and crowd largeness by focusing outreach efforts on the benefits to the participants in the pilot including:

- possibility of contributing to intelligence and the Intelligence Community
- emphasis on the gamified aspects of participation in the effort
- use of challenge aspects of “beating the professional”
- offer of the possibility of interacting with or indeed teaming with Intelligence Community professionals
- training in finance and financial instruments participants will receive that can be applied in their own lives
- offer of the possibility of winning the nominal prizes on offer

The creation of an easily navigable, attractive, and informative outreach website (separate from the pilot prediction market website) should aid outreach efforts. The sponsor should use both pull and push techniques to drive traffic to the outreach website and to get potential participants to register and subsequently participate in the pilot. The sponsor should place attractive advertising (similar to those used for SETI@HOME) in the online and print-based professional journals of each target community, trade journals, and general interest publications. Additionally, the sponsor should encourage preeminent bloggers in each relevant community to promote the pilot in their blogs.

Potential participants should complete a short, but nonetheless useful survey, as part of the registration process to allow the pilot sponsors to target their efforts to ensure participant segment size and diversity across segments are being adequately addressed and achieved. The sponsors can use these efforts to address scalability concerns to adjust approaches to demographic, psychographic, expertise, and experience segments with low participation rates, and emphasize what works well to the detriment of ineffective approaches. Given that participants should have a choice as to which predictions to make (which contracts to buy) and that this information about choices should be available to the organizers, the crowdsources should redesign tasks with low participation rates redesigned to make them more attractive, or crowdsources can adjust their incentive structure for the same purpose.

The warmup tournament phase of the pilot will allow the organizers to assess the drivers of crowd participation and crowd responsiveness to incentives on offer. Participants will be asked to complete periodic short questionnaires focusing on these drivers to thus allow organizers to determine what works and what does not in terms of driving participation and the effectiveness of incentives and make adjustments accordingly.

3. Project Pilot Participant Incentivization

Participants internal to the Intelligence Community can be incentivized by:

- management support and management buy in

- perceived value of the approach to estimating and forecasting
- payoffs in terms of social capital associated with teaming within, across, and external to the Intelligence Community
- ability of participants to go beyond simple tasks to more complicated estimating tasks
- gamified experience as manifested by leaderboards, missions etc.
- nominal prizes, such as patches, coins etc., and real money awards from $20 to $50 for "beating the crowd"

Additionally, participants external to the Intelligence Community can be incentivized to participate by:

- possibility of contributing to intelligence and the Intelligence Community
- emphasis on the gamified as aspects of participation in the effort
- use of the challenge aspects of "beating the professional"
- offer of the possibility of interacting with, or indeed, teaming with Intelligence Community professionals
- training in finance and financial instruments participants will receive that can be applied in their own lives
- offer of the possibility of winning the nominal prizes on offer

The nominal prizes should include a variety of patches associated with task or mission completion, nominal prizes ranging from $20 to $50 for overall forecast accuracy, etc.

4. Project Pilot Participant Training

It is clear from the literature that training participants in probability and statistics and cognitive biases results in improved forecast performance. The pilot should provide online, live, and self-paced training in these areas, as well as in the workings of financial markets and financial instruments. The latter should train participants in how prediction markets are analogous to financial markets, and as a result, on how prediction market assets, contracts, and instruments work in detail. Self-paced, web-based, and instructor-led training can be developed by simply repurposing (with slight terminology changes) some of the large body self-paced and instructor-led content in introductory finance in academia. Another option is to partner with a provider, such as the Khan Academy, to develop and stream dedicated content. Training in the workings of the software platform itself has already been developed by most providers of such tools and can be tweaked fairly easily to reflect how participants should interact with the software platform and with other participants and the organizers of the estimation effort.

5. Task Design and Task Limitations

Tasks in the pilot should consist of participants estimating "one shot" probabilities of individual, unique events, estimating conditional probabilities for event sequences, estimating when an outcome exceeds a threshold, estimating a mean outcome, and estimating the likelihood and extent of forecast errors. These tasks should be limited by design during the pilot to geopolitical forecasts associated with future North Korean actions, statistics, economic performance, and the like.

The pilot sponsor should design some of these tasks to be decomposable in to subtasks, each of which may or may not be serial in design or amenable to being performed in parallel. Other tasks may involve nuance. For example, the task related an asset traded on the market should require not only an estimate of a conditional probability but also an estimate of the forecast error associated with that conditional probability combined in a public and transparent way to determine asset value.

The organizers should split tasks in the pilot into individual tasks and team tasks. Team and individual task payoffs should include an assessment of task complexity that is a multiplier for the payoff for a given task. The value of this multiplier is set by the pilot organizers prior to the task being posted for bids.

Both team and individual tasks have limitations in that sponsors may and should make adjustments for task complexity, task urgency, and the like, but these adjustments may likely decrease the transparency of the market due to the necessarily complex resulting payoff schemes. Sponsors should explore these limitations during the warmup phase of the pilot and adjust the task design as needed to ensure a fully functioning market that drives forecast accuracy.

6. Prediction Market Software Selection and Configuration

The pilot should use commercially available prediction market software configured to reflect the prediction market design and platform characteristics discussed in Chapter VI. Depending on the sponsor of the project, the software may be obtained on commercial terms or as part of a partnership agreement. The organizers must first select an appropriate software platform, configure it to be consistent with methodological and platform facility considerations, and operate the platform over the life of the pilot. In selecting the software platform for the pilot, the organizers may be limited in their choices to venders willing to partner with them for the pilot, or organizers may be able to purchase the software on commercial terms. In either case, the organizer's team for the pilot must become expert at how to configure, administer, and operate the software. Furthermore, the organizer's team must also manage the interface between the outreach website and the software platform.

7. Pilot Program Operation and Administration

Pilot program operation and administration activities include running the warmup tournaments, selection of superforecasters, teams and teaming, task success determination, and provision and timing of feedback. Prediction markets software platform should manage the purely administrative tasks, such as those associated with participant registration, participant account management, participant contact management, administrative aspects of participant task selection and tracking, forecast entry, task payoff accounting, portfolio accounting, feedback provision and tracking, game related list generation and tracking, award generation and provision, etc. These tasks are not discussed in detail in this thesis.

a. Warmup Tournament

At the conclusion of the initial outreach effort (a sufficiently large, diverse, and motivated crowd chosen to participate), the organizers should start the education phase of the pilot. The organizers need to provide access to in person and online training appropriate for the pilot and discussed previously. Once a suitable fraction of potential participants has completed the training on offer, the pilot project proper should begin with a warmup tournament. During this phase, three to five forecasting tasks of each type (estimating "one shot" probabilities, participants estimating conditional probabilities for event sequences, estimating when an outcome exceeds a threshold, estimating a mean outcome, and estimating the likelihood and extent of forecast errors) associated with future North Korean actions, statistics, economic performance, and the like should be on offer from which potential participants can choose. Participants should be able to choose (bid on) some, all, or none of the tasks on offer. Sponsors should use bidding information, coupled with data gathered via survey when participants first register for the pilot, to ensure that tasks have sufficiently large, diverse, and well trained crowds working on them. If not, organizers should adjust both task design and outreach efforts to ensure that the crowd for each task has the appropriate characteristics. At this point, participants should be allowed to form teams (or if they have stated they are amenable to working in teams, organizers should place them in teams). Participants should be allowed to participate in the warmup tournament as individuals, self-formed teams, or organizer-formed teams. The warmup tournament will run for six months, at which point sponsors should evaluate individual and team performance.

b. Teams and Teaming

Sponsors should test characteristics related to teams and teaming during the warmup tournament. As mentioned previously, participants should be allowed to participate in the warmup tournament as individuals, self-formed teams, or organizer-formed teams. Sponsors should assess the performance of teams during the warmup tournament on an ongoing basis to allow the real-time capture of any relationships between team composition, characteristics, participation rates, task choice, and forecast accuracy. Then, sponsors should feed this information back into the task design and outreach processes to ensure that the crowd of teams is characterized by adequate diversity, backgrounds, expertise, and largeness. One aspect of teams and teaming that sponsors should explicitly examine during the warmup phase is the impact of the addition or deletion on Intelligence Community members to the teams on performance. The sponsors also need to examine the performance of teams with a membership consisting of Intelligence Community members.

c. Selection of Superforecasters

During the warmup of the tournament, sponsors should select individual and team superforecasters. Sponsors should deem individuals and teams in the top decile of forecast performance during the warmup tournament superforecasters and their forecasts should be included in the prediction market by first allowing them to participate in a superforecasters *only* prediction market, and with appropriate adjustment, also to trade in the overall prediction market. In theory, superior forecasts in both markets are allowed with the added bonus that the overall market may capture information outside the market consensus.

d. Task Success Determination

Sponsors should design all tasks during the pilot so that forecast success, failure, or margin is clear at the end of each market period. Proper task design should ensure that even complex, nuanced tasks posed to the market have outcomes that sponsors can assess beyond a doubt.

e. Provision and Timing of Feedback

In some sense, the market price of the contract under trade should provide feedback. However, given that forecast performance improves with feedback, it behooves the organizers to provide participants with feedback beyond that simply of asset portfolio value. Given that contract settlement only occurs at the end of each forecast period, sponsors should not provide feedback during the pilot. However, sponsors should provide feedback from the pilot and from each subsequent round of forecasting to participants by given them a document reflecting the forecast performance of the market broken down by participant characteristic, task characteristics (design), task complexity, and the evolution of asset value over time (which can be compared to the value of an individual or team's assets and portfolios over time, which should also be available).

f. *Evaluating the Pilot of the Methodology*

Sponsors should initially evaluate the pilot of the methodology at the conclusion of the warmup tournament, and subsequently, at the end of each forecast period. The sponsors should evaluate several items of interest, including:

- individual and team forecast accuracy within and across tasks using Brier scores (where mutually exclusive discrete forecasts are involved) and mean squared error or mean absolute error (for tasks where forecasts are not mutually exclusive)
- side-by-side comparison of forecast error using the methodology versus forecast errors generated using traditional intelligence analytic processes
- impact of individual demographic, psychographic, experience and expertise factors on team performance
- impact of task design and task complexity on forecast accuracy
- impact of task design and task complexity on individual and team task selection
- impact of training on forecast accuracy
- whether or not superforecaster performance remains stable from round to round; if not the case, the sponsor should shut down the pilot, as the major premise of the methodology—that superforecasters exist and consistently overperform in terms of forecast accuracy—will have been falsified

Beyond this test, sponsors should use the results of the evaluation effort to improve the task design, especially for complex tasks, and to guide outreach efforts to attract individuals with profiles reflecting demographic, psychographic, experience and expertise factors most associated with forecast accuracy. Sponsors should only scale up the pilot if it confirms the expectation of increased forecast accuracy when compared to traditional methods, and if it is able to attract a large enough, diverse enough crowd to participate in the project.

C. SCALING UP THE PILOT

If the pilot is successful, sponsors can then scale up the pilot to address tasks beyond those related to North Korea. The best (most successful) task designs from the pilot can serve as examples for how these new tasks will be structured. Sponsors should enhance outreach efforts to make the participating crowd even bigger, more diverse, and one reflective of desired areas of experience and expertise. Sponsors should implement learning from the pilot regarding teams and teaming, especially those related to team structure and composition for increased forecast accuracy. Additionally, sponsors should test alternative forecast horizons (three months instead of six months or one year versus six months) in the scaled up application of the methodology. Sponsors should provide feedback from forecast session to forecast session to all participants and the impact of feedback on forecast accuracy tested.

D. CONCLUSION AND DIRECTIONS FOR FURTHER RESEARCH

This thesis has proposed a methodology for applying crowd-based analytic methodologies to the problem of intelligence analysis while accounting for and taking advantage of the unique characteristics of the intelligence analysis process and the Intelligence Community culture itself. The crowd-based techniques utilized in developing the methodology include using combined forecasts based on prediction markets-based technique and crowdsourcing techniques to improve forecast accuracy. The thesis' particular contribution focuses on understanding the unique characteristics of the Intelligence Community culture and work processes as a basis for applying crowd-based methodology to improve predictions of real-world events.

This thesis is just a starting point; the methodology should be subject to several rounds of peer review and revision before implementation, even in pilot form. Once this review and revision occurs, sponsors can implement the pilot and test the reality of the methodology creating consistently more accurate forecasts than traditional methods. If the pilot is successful, the methodology becomes one more tool in the intelligence analysts' quiver. At the end of the day, if the degree of success or failure of the methodology is knowable once, at a minimum, the pilot runs. If successful, analysts can then use the methodology both for intelligence analysis and for any field in which forecasts are subject to significant uncertainty.

APPENDIX

Table 14 provides a summary of the mean error reductions by combining forecasts across 30 studies that Armstrong reviewed.

Table 14. Error Reduction Resulting from Combining Forecasts[187]

Exhibit 1
Error Reductions from Combining Ex Ante Forecasts

Study	Methods	Components	Criterion	Data	Situation	Validation Forecasts	Forecast Horizon	Percent error reduction
Levine (1960)	intentions	2	MAPE	annual	capital expenditures	6	1	18.0
Okun (1960)	“	2	“	“	housing starts	6	1	7.0
Landefeld & Seskin (1986)	“	2	MAE	“	plant & equipment	11	1	20.0
Armstrong et al. (2000)	“	4	RAE	“	consumer products	65	varied	5.5
Winkler & Poses (1993)	expert	4	Brier	cross-section	survival of patients	231	varied	12.2
Thorndike (1938)	“	4 to 6	% wrong	“	knowledge questions	30	varied	6.6
Makridakis et al. (1993)	“	5	MAPE	monthly	economic time series	322	1 thru 14	19.0
Richards & Fraser (1977)	“	5	“	annual	company earnings	213	1	8.1
Batchelor & Dua (1995)	“	10	MSE	“	macroeconomic	40	1	16.4
Kaplan et al. (1950)	“	26	% wrong	cross-section	technology events	16	varied	13.0
Zarnowitz (1984)	“	79	RMSE	quarterly	macroeconomic	288	1	10.0
Sanders & Ritzman (1989)	extrapolation	3	MAPE	daily	public warehouse	260	1	15.1
Makridakis & Winkler (1983)	“	5	“	monthly	economic time series	617	18	24.2
Makridakis et al. (1993)	“	5	“	“	“	322	1 thru 14	4.3
Lobo (1992)	“	5	“	quarterly	company earnings	6,560	1 thru 4	13.6
Schnaars (1986)	“	7	“	annual	consumer products	1,412	1 thru 5	20.0
Landefeld & Seskin (1986)	econometric	2	MAE	annual	plant & equipment	7	1	21.0
Clemen & Winkler (1986)	“	4	MAD	quarterly	GNP (real & nominal)	45	1 thru 4	3.4
Shamseldin et al. (1997)	“	5	MAPE	annual	rainfall runoff	22	1	9.4
Lobo (1992)	expert/extrap	2	MAPE		company earnings	6,560	1 thru 4	11.0
Lawrence et al. (1986)	“	3	“	annual monthly	economic time series	1,224	1 thru 18	10.7
Sanders & Ritzman (1989)	“	3	“	daily	public warehouse	260	1	15.5
Lobo & Nair (1990)	“	4	“	annual	company earnings	768	1	6.4
Landefeld & Seskin (1986)	intentions/econ	2	MAE	annual	plant & equipment	11	1	11.5
Vandome (1963)	extrap/econ	2	MAPE	quarterly	macroeconomic	20	1	10.1
Armstrong (1985)	“	2	“	annual	photo sales by country	17	6	4.2
Weinberg (1986)	expert/econ	2	“	cross-section	performing arts	15	varied	12.5
Bessler & Brandt (1981)	exprt/extrap/econ	3	“	quarterly	cattle & chicken prices	48	1	13.6
Fildes (1991)	“	3	MAE	annual	construction	72	1 & 2	8.0
Brandt & Bessler (1983)	“	6	MAPE	quarterly	hog prices	24	1	23.5
							Unweighted average	**12.5**

187 Adapted from Armstrong, “Combining Forecasts,” 417–439.

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF REFERENCES

Abramowicz, Michael. “Deliberative Information Markets for Small Groups.” In *Information Markets: A New Way of Making Decisions*, edited by Robert Han and Paul Tetlock, 101–125. Washington, DC: AEI Press, 2006.

Aid, Matthew M. “Sins of Omission and Commission: Strategic Cultural Factors and U.S. Intelligence Failures during the Cold War.” *Intelligence and National Security* 26, no. 4 (2011): 478–494. doi: 10.1080/02684527.2011.580602.

Alpert, Marc, and Howard Raiffa. “A Progress Report on the Training of Probability Assessors.” In *Judgment under Uncertainty: Heuristics and Biases*, edited by David Kahneman, Paul Slovic, and Amos Tversky, 294–305. New York: Cambridge University Press, 1982.

Armstrong, J. Scott. “Combining Forecasts.” In *Principles of Forecasting: A Handbook for Researchers and Practitioners*, edited by J. Scott Armstrong, 417–439. Norwell, MA: Kluwer Academic Publishing, 2001. http://repository.upenn.edu/marketing_papers/34.

Arneson, Sveinung, and Ole Bergford. “Prediction Markets versus Polls: An Examination of Accuracy for the 2008 and 2012 Elections.” *Journal of Prediction Markets* 8, no. 3 (2014): 24–33.

Arrow, Kenneth J., Robert Forsythe, Michael Gorham, Robert Hahn, Robin Hanson, John O. Ledyard, Saul Levmore, and Robert Litan et al. “The Promise of Prediction Markets.” *Science* 320 (2008): 877–878.

Bell, Tom W. “Private Prediction Markets and the Law.” *Journal of Prediction Markets* 3, no. 1 (2009): 89–110.

Berg, Joyce E., Forrest D. Nelson, and Thomas A. Rietz. “Prediction Market Accuracy in the Long Run.” *International Journal of Forecasting* 24, no. 2 (2008): 285–300.

Berg, Joyce E., George R. Neumann, and Thomas A. Reitz. “Searching for Google’s Value: Using Prediction Markets to Forecast Market Capitalization Prior to an Initial Public Offering.” *Management Science* 55, no. 3 (2009): 348–361.

Bisogno, Raymond. “Problem Solving in Homeland Security and Creating Policy Conditions for Enhanced Civic Engagement: An Examination of Crowdsourcing Models.” Master’s thesis, Naval Postgraduate School, 2017.

Boardman, Chase. "Organizational Culture Challenges to Intelligence Community Communication and Interaction." Master's thesis, Joint Forces Staff College, 2006.

Boyle, Alan. "Gamers Solve Molecular Puzzle that Baffled Scientists." *NBC News*, November 2, 2015. https://www.nbcnews.com/science/science-news/gamers-solve-molecular-puzzle-baffled-scientists-f6C10402813.

Brown, Barbara G., and Allen. H. Murphy. "Improving Forecasting Performance by Combining Forecasts: The Example of Road-surface Temperature Forecasts." *Meteorological Applications* 3, no. 3 (1996): 257–265. doi: 10.1002/met.506 0030307.

Buckly, Patrick. "Harnessing the Wisdom of Crowds: Decision Spaces for Prediction Markets." *Business Horizons* 59, no. 1 (2016): 85–84.

Burch, James. "The Domestic Intelligence Gap: Progress since 9/11?" *Homeland Security Affairs* 4 (2008). https://www.hsaj.org/articles/129.

Cartwright, Susan, and Cary L. Cooper. "The Role of Culture Compatibility in Successful Organizational Marriage." *The Academy of Management Executive (1993–2005)* 7, no. 2 (May 1993): 57–70.

Celasco, Matias, Juan Ignacio Yanez, and Roberto Gamen. "Galaxy Conqueror: Astronomy, Citizens, and Gamification." In *2016 XI Latin American Conference on Learning Objects and Technology (LACLO)*. San Carlos, Costa Rica: IEEE, 2016. doi 10.1109/LACLO.2016.7751798.

Central Intelligence Agency. *Intelligence Community and Policymaker Integration: A Study in Intelligence Anthology*. Washington, DC: Central Intelligence Agency, 2014. https://www.cia.gov/library/center-for-the-study-of-intelligence/csi-publications/books-and-monographs/intelligence-community-and-policymaker-integration/IC%20and%20Policymaker%20Integration-A%20Studies%20in%20Intelligence%20Anthology.pdf.

Chen, Kay-Yut, and Charles R. Plott. *Prediction Markets and Information Aggregation Mechanisms: Experiments and Applications*. Pasadena, CA: California Institute of Technology, 1998.

Chittilappilly, Anand Inasu, Lei Chen, and Sihem Amer-Yahia. "Survey of General-Purpose Crowdsourcing Techniques." *IEEE Transactions on Knowledge and Data Engineering* 28, no. 9 (2016): 2246–2266.

Chiu, Chao-Min, Ting-Peng Liang, and Efraim Turban. “What Can Crowdsourcing Do for Decision Support?” *Decision Support Systems* 65 (September 2014): 40–49.

Clemen, Robert T. “Combining Forecasts: A Review and Annotated Bibliography.” *International Journal of Forecasting* 5 (1989): 559–583.

Clintin P. Davis-Stober, David V. Budescu, Jason Dana, and Stephen Broomell. “When Is a Crowd Wise?” *Decision* 1, no. 2 (2014): 79–101.

Commission on the Intelligence Capabilities of the United States Regarding Weapons of Mass Destruction. *Report to the President*. Washington, DC: Commission on the Intelligence Capabilities of the United States Regarding Weapons of Mass Destruction, 2005.

Cowgill, Bo, Justin Wolfers, and Eric Zitzewitz. “Using Prediction Markets to Track Information Flows: Evidence from Google.” In *Auctions, Market Mechanisms and Their Applications: First International ICST Conference, AMMA*, vol. 14, edited by Sanmay Das, Michael Ostrovsky, David Pennock, and Boeslaw Szymanski. Boston, MA: Springer, 2009. https://www.stat.berkeley.edu/~aldous/157/Papers/GooglePredictionMarketPaper.pdf.

Davis, Danny M. “Designing a Viable Prediction Market to Forecast Defense Acquisition Cost and Schedule Outcomes.” *Defence and Peace Economics* 22, no. 3 (2011): 351–366. doi: 10.1080/10242694.2010.491680.

Deck, Cary, Lin Shengle, and David Porter. “Affecting Policy by Manipulating Prediction Markets: Experimental Evidence.” *Journal of Economic Behavior and Organization* 85 (2013): 48–62.

Decker, Carolin, Isabelle. M. Welpe, and Bernd H. Ankenbrand. “How to Motivate People to Put Their Money Where Their Mouth Is: What Makes Employees Participate in Electronic Prediction Markets.” *Technological Forecasting and Social* 78, no. 6 (2011): 1002–1015.

Dhami, Manpreet K., David R. Mandel, Barbara A. Mellers, and Philip E. Tetlock. “Improving Intelligence Analysis with Decision Science.” *Perspectives in Psychological Science* 10, no. 6 (2015): 753–757.

Doan, An Hai, Raghu Ramakrishnan, and Alon Y. Halevy. “Crowdsourcing Systems on the World-wide Web.” *Communications of the ACM* 54, no. 4 (2011): 86–96.

Fama, Eugene F. “Efficient Capital Markets: A Review of Theory and Empirical Work.” *Journal of Finance, Papers and Proceedings of the Twenty Eighth Annual Meeting of the American Finance Association* 25, no. 2 (1969): 383–417.

Fold-it. “Solve Puzzles for Science.” Accessed June 12, 2018. https://fold.it/portal/.

Gaspoz, Cederic. *Prediction Markets Supporting Technology Assessment* n.p., Printed in the World, 2011.

Geiger, David, Stefan Seedorf, Thimo Schulze, Robert C. Nickerson, and Martin Schader. “Managing the Crowd: Towards a Taxonomy of Crowdsourcing Processes.” In *Proceedings of the Seventeenth Americas Conference on Information Systems*. Detroit, MI: Association for Information Systems, 2011. https://pdfs.semanticscholar.org/d134/065587b5276bec1b0e93695edd673d0bfc10.pdf.

George, Roger Zane. “Beyond Analytic Tradecraft.” *International Journal of Intelligence and CounterIntelligence* 23, no. 2 (2010): 296–306. doi: 10.1080/08850600903566124.

Graefe, Andreas, and J. Scott Armstrong. “Comparing Face-to-Face Meetings, Nominal Groups, Delphi and Prediction Markets on an Estimation Task.” *International Journal of Forecasting* 27, no. 1 (2011): 183–195. http://dx.doi.org/10.1016/j.ijforecast.2010.05.004.

Graefe, Andreas, J. Scott Armstrong, Randall J. Jones Jr., and Alfred G. Cuzáne. “Combining Forecasts: An Application to Elections.” *International Journal of Forecasting* 30, no. 1 (2014): 43–54. https://doi.org/10.1016/j.ijforecast.2013.02.005.

Gruca, Thomas S., and Joyce. E. Berg. “Public Information Bias and Prediction Market Accuracy.” *Journal of Prediction Markets* 1, no. 3 (2007): 219–231.

Hamari, Juho, Jonna Koivisto, and Harri Sarsa. “Does Gamification Work? A Literature Review of Empirical Studies on Gamification.” In *Proceedings of the 47th Hawaii International Conference on System Sciences—HICSS*, 3025–3034. Waikoloa, HI: IEEE, 2014. doi: 10.1109/HICSS.2014.377.

Hamrah, Satgin S. “The Role of Culture in Intelligence Reform.” *Journal of Strategic Security* 6, no. 3 (Fall 2013): 160–171, Supplement, *Ninth Annual IAFIE Conference: Expanding the Frontiers of Intelligence*.

Hanson, Robin. “Combinatorial Information Market Design.” *Information System Frontiers* 5, no. 1 (2003): 107–119.

———. “Impolite Innovation: The Technology and Politics of ‘Terrorism Futures’ and Other Decision Markets.” in *Promoting the General Welfare, American Democracy and the Political Economy of Government Performance*, edited by Eric Patashnik and Alan Gerber, 151–173. Washington, DC: Brookings Institution Press, 2006.

———. “On Market Maker Functions.” *Journal of Prediction Markets* 1, no. 1 (2007): 3–15.

Hare, Nicholas P., and Paul Collinson. “Organisational Culture and Intelligence Analysis: A Perspective from Senior Managers in the Defence Intelligence Assessments Staff.” *Public Policy and Administration* 28, no. 2 (2013): 214–229.

Hayek, Friedrich A. “The Use of Knowledge in Society.” *American Economic Review* XXXV, no. 4 (1945): 519–530. http://www.econlib.org/library/Essays/hykKnw1.html.

Ho, Anson T. Y., Phillip M. Polgreen, and Thomas Prendergast. “Prediction Market for Disease Surveillance, a Case Study of Influenza Activity.” *Journal of Prediction Markets* 10, no. 1 (2016): 68–82.

Ho, Tung H., and Kay. Y. Chen. “New Product Blockbusters: The Magic and Science of Prediction Markets.” *California Management Review* 50, no. 1 (2007): 144–158.

Hopman, Jay. “Using Forecasting Markets to Manage Demand Risks.” *Intel Technology Journal* 11, no. 2 (2007): 126–136.

Hosseini, Mahmood, Alimohammad Shahri, Keith Phalp, and Raian Ali. “Recommendations on Adapting Crowdsourcing to Problem Types.” In *IEEE 9th Conference on Research Challenges in Information Science*. Athens, Greece: IEEE RCIS, 2015.

Hosseini, Mahmood, Keith Phalp, Jacqui Taylor, and Raian Ali. “On the Configuration of Crowdsourcing Projects.” *International Journal of Information System Modeling and Design* 6, no. 3 (July 2015): 27–45.

Hubbard, Douglas W. *How to Measure Anything: Finding the Value of Intangibles*. 3rd ed. Hoboken, NJ: John Wiley and Sons, 2014.

Hunt, Tam. “How I Became a Superforecaster.” *Slate*. Last updated November 19, 2015. http://www.slate.com/articles/technology/future_tense/2015/11/good_judgment_project_how_i_became_a_superforecaster_for_the_intelligence.html.

Intelligence Advanced Research Projects Agency. “Aggregative Contingent Estimation (ACE).” Accessed April 29, 2018. https://www.iarpa.gov/index.php/research-programs/ace/baa.

Johnston, Rob. *Analytic Culture in the U.S. Intelligence Community: An Ethnographic Study*. Washington, DC: Central Intelligence Agency, 2005.

Kajdasz, James E., Jason A. Burdick, Matthew R. Christ, and David Lange. “An Alternative Analysis Technique: Examining the IC Prediction Market.” *Studies in Intelligence* 3, no. 58 (2014): 22–37.

Kent, Sherman. *Strategic Intelligence for American World Policy*. Princeton, NJ: Princeton University Press, 2015.

Kominers, Scott Duke. “Prediction Markets Didn’t Call Trump’s Win, Either.” *Bloomberg View*, November 15, 2016. https://www.bloomberg.com/view/articles/2016-11-15/prediction-markets-didn-t-call-trump-s-win-either.

Lewis, Jeffrey. “FSA Overruns Al Kibar.” *Arms Control Wonk* (blog), February 25, 2011. http://www.armscontrolwonk.com/archive/206309/fsa-overruns-al-kibar/.

Li, Eldon Y., Tung Chen-Yuan, and Shu-Hsun Chang. “User Adoption of Wisdom of Crowd: Usage and Performance of Prediction Market System.” *International Journal of Electronic Business* 12, no. 2 (2015): 185–214.

Lin, Hung-Wen, Chen Yuan Tung, and Jason Yeh. “Multivariate Methods in Assessing the Accuracy of Prediction Markets Ex Ante Based on the Highest Price Criterion.” *The Journal of Prediction Markets* 7, no. 3 (2013): 29–44.

Liu, Helen K. “Crowdsourcing Government: Lessons from Multiple Disciplines.” *Public Administration Review* 77, no. 5 (2017): 656–667.

Lorenz, Jan, Heiko Rauhut, Frank Schweitzer, and Dirk Helbing. “How Social Influence Can Undermine the Wisdom of Crowd Effect.” *Proceedings of the National Academy of Sciences* 108, no. 22 (2001): 9020–9025.

Lowenthal, Mark M. *Intelligence: From Secrets to Policy*, 3rd ed. Washington, DC: CQ Press, 2006.

Luckner, Stefan. “How to Pay Traders in Information Markets: Results from a Field Experiment.” *Journal of Prediction Markets* 1, no. 2 (2007): 147–156.

———. “Prediction Markets: Fundamentals, Key Design Elements and Applications.” In *Proceedings on the 21st Bled Conference*, 236–247. Bled, Slovenia: Association for Information Systems, 2008.

Luz, Nino, Nuno Silva, and Paulo Novais. “A Survey of Task Oriented Crowdsourcing.” *Artificial Intelligence Review* 44, no. 2 (2015): 187–213.

Maras, Marie-Helen. “Overcoming the Intelligence-sharing Paradox: Improving Information Sharing through Change in Organizational Culture.” *Comparative Strategy* 6, no. 3 (2017): 187–197. doi: 10.1080/01495933.2017.1338477.

Mason, Winter, and Duncan J. Watts. “Financial Incentives and the Performance of Crowds.” *ACM SigKDD Explorations Newsletter* 11, no. 2 (2010): 100–108. doi: 10.1145/1809400.1809422.

May, Ernest R., Roy Godson, and Gary James Schmitt, ed. *U.S. Intelligence at the Crossroads: Agendas for Reform*. Washington, DC: Brassey’s, 1995.

McBride, Marissa F., and Mark A. Burgman. “What Is Expert Knowledge, How Is Such Knowledge Gathered, and How Do We Use It to Address Questions in Landscape Ecology?” In *Expert Knowledge and Its Application in Landscape Ecology*, edited by Ajith H. Perera, C. Ashton Drew, and Chris J. Johnson, 11–39. New York: Springer, 2012.

McHugh, Patrick, and Aaron Jackson. “Prediction Market Accuracy: The Impact of Size, Incentives, Context, and Interpretation.” *Journal of Prediction Markets* 6, no. 2 (2012): 22–46.

Mellers, Barbara, Eric Stone, Pavel Atanasov, Nick Rohrbaugh, S. Emlen Metz, Lyle Ungar, and Michael Bishop et al. “The Psychology of Intelligence Analysis: Drivers of Prediction Accuracy in World Politics.” *Journal of Experimental Psychology: Applied* 21, no. 1 (2015): 1–14. doi: 10.1037/xap0000040.

Moncton, Nathan B. “U.S. Using Canadian Games to Improve Its Intel.” *The Times*, July 3, 2017.

Morschheuser, Benedikt, Juho Hamari, Jonna Koivisto, and Alexander Maedche. “Gamified Crowdsourcing: Conceptualization, Literature Review, and Future Agenda.” *International Journal of Human-Computer Studies* 106 (October 2017): 26–43.

Mouton, Troy Michael. “Organizational Culture’s Contributions to Security Failures within the United States Intelligence Community.” Master’s thesis, Louisiana State University, 2002. http://digitalcommons.lsu.edu/gradschool_theses/1121.

Nakatsu, Robbie T., Elissa B. Grossman, and Charalambos L. Iacovu. “A Taxonomy of Crowdsourcing Based on Task Complexity.” *Journal of Information Science* 60, no. 6 (2014): 823–834.

National Commission on Terrorist Attacks upon the United States. *Final Report of the National Commission on Terrorist Attacks upon the United States*. New York: W. W. Norton, 2004.

Noeth, Markus, Colin F. Camerer, Charles R. Plott, and Martin Webber. “Information Aggregation in Experimental Asset Markets: Traps and Misaligned Beliefs.” Working paper 1060, California Institute of Technology, Pasadena, CA, 1999.

Office of the Director of National Intelligence. *Analytic Standards*. Intelligence Community Directive 203. Washington, DC: Office of the Director of National Intelligence, 2015.

Ozan, Erol. *Optimization of Information Technology Risk Event Prediction Markets*. Greenville, NC: East Carolina University, 2013.

———. “The Use of Prediction Markets in Information Technology Risk Management.” Presented at American Society for Engineering Management Conference, Virginia Beach, VA, 2012.

Pennock, David M. “A Dynamic Pari-mutuel Market for Hedging, Wagering, and Information Aggregation.” In *Proceedings of the Fifth ACM Conference on Electronic Commerce (EC’04)*, 170–179. New York: ACM, 2004.

Peters, Mark, Anthony Man-Cho, and Ye Yinyu. “Pari-Mutuel Markets: Mechanisms and Performance.” In *WINE 2007: Internet and Network Economics*. Lecture Notes in Computer Science Series, vol. 4858. 82–95. Heidelberg: Springer, 2007.

Phythian, Mark. “Cultures of National Intelligence.” In *Routledge Companion to Intelligence Studies*, edited by Robert Dover, Michael S. Goodman, and Claudia Hillebrand, 33–41. Abingdon, United Kingdom: Routledge, 2013.

Phythian, Mark, and Peter Gill. *Intelligence in an Insecure World*. Cambridge: Polity Press, 2012.

Rajakovich, David, and Vladimir Vladimirov. “Prediction Markets as a Medical Forecasting Tool: Demand for Hospital Service.” *Journal of Prediction Markets* 3, no. 2 (2009):78–106.

Rothschild, David. “Forecasting Elections Comparing Prediction Markets, Polls, and Their Biases.” *Public Opinion Quarterly* 73, no. 5 (2009): 895–916.

Seaborn, Katie, and Deborah I. Fels. “Gamification in Theory and Action: A Survey.” *International. Journal of Human Computer Studies* 74 (February 2015): 14–31. http://dx.doi.org/10.1016/j.ijhcs.2014.09.006.

Schenk, Erik, and Claude Guittard. "Towards a Characterization of Crowdsourcing Practices." *Journal of Innovation Economics and Management* 7, no. 1 (2011): 93–107.

Servan-Schreiber, Emile, Justin Wolfers, David M. Pennock, and Brian Galebach. "Prediction Markets: Does Money Matter?" *Electronic Markets* 14, no. 3 (2004): 243–251. doi: 10.1080/1019678042000245254.

Simon, Herbert A. *Decision-making and Problem Solving, Research Briefings 1986: Report of the Research Briefing Panel on Decision-making and Problem Solving*. Washington, DC: National Academy Press, 1986.

Simmons, Joseph P., Leif D. Nelson, Jeff Galak, and Shane Frederick. "Intuitive Biases in Choice versus Estimation: Implications for the Wisdom of Crowds." *Journal of Consumer Research* 38, no. 1 (June 2011): 1–15.

Sjöberg, Lennart. "Are All Crowds Equally Wise? A Comparison of Political Election Forecasts by Experts and the Public." *Journal of Forecasting* 28, no. 1 (2009): 1–18.

Slamka, Christian, Bernd Skiera, and Martin Spann. "Prediction Market Performance and Market Liquidity: A Comparison of Automated Market Makers." *IEEE Transactions on Engineering Management* 60, no. 1 (2013): 169–185.

Speigel, Alix. "So You Think You Are Smarter Than a CIA Agent." *NPR*, April 2, 2014. https://www.npr.org/sections/parallels/2014/04/02/297839429/-so-you-think-youre-smarter-than-a-cia-agent.

Stottlemyre, Steven A. "HUMINT, OSINT, or Something New? Defining Crowdsourced Intelligence." *International Journal of Intelligence and CounterIntelligence* 28, no. 3 (2015): 578–589. doi: 10.1080/08850607.2015.992760.

Sunstein, Cass R. *Infotopia: How Many Minds Produce Knowledge*. Oxford: Oxford University Press, 2006.

Surowiecki, James. *The Wisdom of Crowds*. New York: Random House, 2005.

Tapia, Andrea H., Nicolas LaLone, and Hyun-Woo Kim. "Run Amok: Group Crowd Participation in Identifying the Bomb and Bomber from the Boston Marathon Bombing." In *Proceedings of the 11th International ISCRAM Conference*, 265–274. Rio de Janeiro, Brazil: Information Systems for Crisis Response and Management, 2014.

Teschner, Florian, and Christof Weinhardt. “A Macroeconomic Forecasting Market.” *Journal of Business Economics* 85 (2015): 293–317. doi: 10.1007/s11573-014-0741-5.

Tetlock, Phillip, and Dan Gardner. *Superforecasting: The Art and Science of Prediction*. New York: Penguin Random House, 2016.

Turner, Michael A. “A Distinctive U.S. Intelligence Identity.” *International Journal of Intelligence and Counter Intelligence* 17 (2004): 42–61.

Tyakoff, Alex. “Counter Terrorism and Systems Dynamics: Modeling Organizational Learning in Postmodern Terrorist Group.” In *Terrorism and Global Insecurity: A Multidisciplinary Perspective*, edited by Klint Alexander, 179–192. Chicago, IL: Linton Atlantic, 2009.

Ungar, Lyle, Barb Mellors, Jon Baron, Phil Tetlock, Jaime Ramos, and Sam Swift. *The Good Judgment Project: A Large Scale Test of Different Methods of Combining Expert Predictions,* AAAI Technical Report FS-12-06. Palo Alto, CA: Association for the Advancement of Artificial Intelligence, 2012.

Wadhaw, Tarun. “Lessons from Crowdsourcing the Bostom Bombing Investigation.” *Forbes,* April 22, 2013. http://www.forbes.com/sites/tarunwadhwa/2013/04/22/lessons-from-crowdsourcing-the-bostonmarathon-bombings-investigation/#1416d38312b5.

Williams, L. Vaughn, and James J. Read. “Forecasting Elections.” *Journal of Forecasting* 35, no. 4 (2016): 308–328. doi: 10.1002/for.2377.

Winkler, Robert I. “Probabilistic Prediction: Some Experimental Results.” *Journal of the American Statistical Association* 66, no. 336 (1971): 675–685.

Wolfers, Justin, and Eric Zitzewitz. “Prediction Markets.” *Journal of Economic Perspectives* 18, no. 2 (2004): 107–126.

Yang, Sheng-yun, Tung Li, and Eric van Heck. “Information Transparency in Prediction Markets.” *Decision Support Systems* 78 (2015): 67–79.

INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center
   Ft. Belvoir, Virginia

2. Dudley Knox Library
   Naval Postgraduate School
   Monterey, California