

# Keeping Score: A New Approach to Geopolitical Forecasting

*February 2021*

## ABOUT PERRY WORLD HOUSE

Perry World House is a center for scholarly inquiry, teaching, research, international exchange, policy engagement, and public outreach on pressing global issues.

Perry World House's mission is to bring the academic knowledge of the University of Pennsylvania to bear on some of the world's most pressing global policy challenges and to foster international policy engagement within and beyond the Penn community.

Located in the heart of campus at 38th Street and Locust Walk, Perry World House draws on the expertise of Penn's 12 schools and numerous globally oriented research centers to educate the Penn community and prepare students to be well-informed, contributing global citizens. At the same time, Perry World House connects Penn with leading policy experts from around the world to develop and advance innovative policy proposals.

Through its rich programming, Perry World House facilitates critical conversations about global policy challenges and fosters interdisciplinary research on these topics. It presents workshops and colloquia, welcomes distinguished visitors, and produces content for global audiences and policy leaders, so that the knowledge developed at Penn can make an immediate impact around the world.

**Acknowledgements:** *Thanks to the experts named in Appendix: Interviews who provided valuable insight and feedback throughout the process.*

*All errors are the sole responsibility of the authors. It does not represent the positions, policies or opinions of Penn Global, Perry World House, or the University of Pennsylvania. Thanks to Open Philanthropy, whose support allowed for the drafting of this white paper.*

## REPORT AUTHORS

**Michael C. Horowitz**

*Richard Perry Professor and  
Director of Perry World House*

**Julia Ciocea**

*Research Fellow at  
Perry World House*

**Lauren Kahn**

*Research Fellow at  
Perry World House*

**Christian Ruhl**

*Global Order Program Manager  
at Perry World House*

<b>4</b>	<b>EXECUTIVE SUMMARY</b>	
<b>6</b>	<b>INTRODUCTION</b>	
<b>8</b>	<b>WHY CROWDSOURCED GEOPOLITICAL FORECASTING MATTERS</b>	
	DEFINING GEOPOLITICAL FORECASTING.....	9
	GENERATING FORECASTS TO HELP INFORM GOVERNMENT PROGRAMS AND POLICY.....	10
	KEEPING SCORE AS A FORM OF ACCOUNTABILITY AND CONTINUAL JOB TRAINING.....	10
	KEEPING SCORE TO EVALUATE ANALYTIC METHODS AND SOURCES.....	11
<b>12</b>	<b>EVALUATING U.S. GOVERNMENT EFFORTS AT GEOPOLITICAL FORECASTING</b>	
	BACKGROUND ON RECENT USG GEOPOLITICAL FORECASTING EFFORTS .....	12
	BACKSLIDING IN INTELLIGENCE COMMUNITY FORECASTING EFFORTS .....	16
	OTHER MODELS.....	18
<b>20</b>	<b>KEY ADOPTION QUESTIONS</b>	
	WHETHER TO MANDATE USE OF GEOPOLITICAL FORECASTING PLATFORMS .....	20
	PREDICTION MARKETS VS. FORECASTING AGGREGATION PLATFORMS.....	22
	FORECASTING PLATFORM CLASSIFICATION LEVEL .....	23
	TRACKING INDIVIDUAL PERFORMANCE AND LEADERBOARDS .....	24
	INTEGRATION WITH OTHER FORECASTING APPROACHES.....	25
	RESEARCH AND DEVELOPMENT VERSUS APPLICATION.....	26
	SUPERFORECASTERS .....	27
<b>28</b>	<b>PATHS FORWARD FOR USG GEOPOLITICAL FORECASTING</b>	
	THE INTELLIGENCE COMMUNITY .....	28
	DEFENSE DEPARTMENT.....	29
	THE WHITE HOUSE .....	29
	CONGRESS.....	30
	OTHER OPTIONS.....	30
<b>32</b>	<b>CONCLUSION</b>	
<b>33</b>	<b>APPENDIX: INTERVIEWS</b>	

## > EXECUTIVE SUMMARY

**Cassandra, the Trojan priestess of Greek mythology, was cursed; she could see the future, but no one—not even her father, King Priam—would believe her prophecies. Like King Priam, policymakers want to see the future, but they may not know that more accurate predictions are in sight. The following white paper argues that rigorous crowdsourced geopolitical forecasting methods may be one of the most promising ways to improve the nation’s ability to know the future—if policymakers heed the call to implement these methods in government.**

Such crowdsourced forecasting methods have the potential to significantly improve current analytic techniques in government. To develop assessments of the future, the U.S. intelligence community and other parts of the national security community rely on traditional qualitative analysis of sources, scenario-planning exercises, war games, and more recently, quantitative models, a trend that may grow with new advances in data analytics. From prediction polls to team competitions to prediction markets, experiments conducted by the U.S. intelligence community over the last decade demonstrate the value of crowdsourced forecasting methods as a complement to existing approaches. Experiments show that, for some kinds of questions, crowdsourced methods can offer substantial accuracy advantages compared with other approaches, while also generating additional organizational benefits by providing new means for accountability and continual job training for analysts.

Despite the success of these research and development programs, the U.S. government (USG) has struggled to transition them into programs that can consistently inform important intelligence products like National Intelligence Estimates and assist policymakers in processing a complex world on a daily basis. This white paper assesses existing crowdsourced forecasting efforts in the USG and their promise, exploring the bureaucratic, cultural, and political issues that have caused adoption to lag.

**The primary recommendation of this white paper is for the USG to keep score concerning its own geopolitical forecasts. As part of that effort, the USG should revive efforts to use crowdsourced methods to generate quantifiable, falsifiable, geopolitical forecasts.** The intelligence community is arguably the natural home of such an initiative, but other plausible locations include the Department of Defense, the Office of Science and Technology Policy, the National Security



Council staff, or even Congress. This white paper discusses the pros and cons of each option.

In addition to questions about where to place a geopolitical forecasting platform, it is critical to design any platform around the key sources of value, which generate better intelligence, and the ability to clearly communicate that value to policymakers. Thus, this white paper also makes a number of recommendations on the details of implementation strategy:

- Promote crowdsourced and quantitative geopolitical forecasting as a complement to existing forecasting methods, particularly given evidence suggesting they may actually prove more successful in combination.
- Require use of the outputs of a forecasting platform in relevant products, including analyst engagement when they disagree with the forecast.
- Design a communication system to effectively translate probabilities into useful information for policymakers in a way that conveys the information without generating false confidence.
- Create both classified and unclassified forecasting platforms. This will encourage broad participation, which will improve forecasting accuracy on most questions, while still presenting a classified platform for questions where access to classified information is necessary to even consider the question.
- Decide on the relative importance of identifying and cultivating superforecasters as a programmatic goal prior to launching a crowdsourced forecasting platform.

- Use prediction polls, aggregating them with algorithms and other proven methods, rather than prediction markets. Implement a flexible approach to leaderboards, tracking data on success over time for every forecaster, while creating options to frame it based on self-improvement, in-office “pods,” and other ways that create continuing incentives to participate rather than undermine motivation.
- Fund additional research and development designed to improve the accuracy of crowdsourced geopolitical forecasting methods and efforts to effectively communicate quantitative insights to policymakers.

Improving its geopolitical forecasting capacity could make a substantial difference in improving U.S. national security. As this report describes, proven crowdsourced methods could provide additional information to busy decision makers that help them make judgments with greater context than at present.

## ➤ INTRODUCTION:

**How well does the United States forecast geopolitical futures, and how could it improve? This is an enduring question with substantial and ongoing relevance. Almost 50 years ago, then-Secretary of State Henry Kissinger reportedly said that policymaking is about “making complicated bets about the future,” wishing that “intelligence would supply him with estimates of the relevant betting odds.”<sup>1</sup> Andrew Marshall, in the first year of his decades-long tenure as the director of the Office of Net Assessment in the Department of Defense (DoD) in 1973, described the benefits of quantitative forecasting as not only enhancing intelligence analysis itself, but improving the ability of analysts to clearly communicate their findings to policymakers.<sup>2</sup>**

Unfortunately for the U.S. government (USG), when it comes to the generation and delivery of geopolitical forecasts to policymakers, the state of play remains close to where it was in 1973. From the intelligence-sharing issues that failed to prevent the September 11, 2001, attacks to the inability of the USG to adequately predict and prepare for a pandemic event like COVID-19, events have demonstrated that geopolitical forecasting is an important tool that the USG has not mastered. Policymaking itself involves forecasting, as policymakers’ beliefs about the likely impact of particular plans, and the consequences of alternative courses of action, influence

their decision-making. Yet, the connection between forecasting and policy is often implicit, as policymakers frequently make decisions based on implied theories of cause and effect that they do not make explicit.

As deaths in the United States from COVID-19 continue to rise, it is critical to think about how the United States can improve its geopolitical forecasting efforts. The Biden administration has already taken an initial step by announcing the creation of a National Center for Epidemic Forecasting and Outbreak Analytics.<sup>3</sup> Traditionally, USG forecasting efforts involve scenario-planning exercises, war games, qualitative assessments of evidence, and related activities. This white paper assesses efforts over the last decade to introduce a complement to these methods in the form of crowdsourced geopolitical forecasting. Evidence from a decade of experimentation demonstrates that crowdsourced forecasting methods and other quantitative approaches, such as using algorithms to sort through large quantities of data, can aid analysts at the working level and policymakers in the national security community and beyond. Probabilistic forecasts, simply by forcing the explicit articulation of ideas that will otherwise be expressed more vaguely, serve to spur thinking and clarify insights, even if the probabilities themselves do not drive policy outcomes.

Avoiding false confidence regarding probability assessments is crucial, which requires effectively communicating forecasts in ways that make clear the odds, the reasoning behind those odds, the track record of the approach, and the limits of the approach.

To move the debate forward, this white paper describes the advantages and limits of different crowdsourced forecasting approaches, and some of the bureaucratic obstacles that have made effective implementation of crowdsourced forecasting challenging. The paper then shifts to describing new paths forward for crowdsourced geopolitical forecasting efforts in the USG, laying out several different organizational strategies, along with the upsides and downsides of each. The primary recommendation of this white paper is for the USG to

1 Andrew Marshall, “Memorandum for Brent Scowcroft: Probabilities in Intelligence Analysis,” LOC-HAK-538-3-5-6, November 9, 1973.

2 Ibid.

3 The White House, National Security Directive on United States Global Leadership to Strengthen the International COVID-19 Response and to Advance Global Health Security and Biological Preparedness, January 21, 2021, <https://www.whitehouse.gov/briefing-room/statements-releases/2021/01/21/national-security-directive-united-states-global-leadership-to-strengthen-the-international-covid-19-response-and-to-advance-global-health-security-and-biological-preparedness/>.

keep score on its own forecasts concerning world events. Keeping score matters. It generates analytical clarity, improves accountability, and enhances accuracy. As part of keeping score, the United States should revive efforts to use crowdsourced methods to generate quantifiable, falsifiable, geopolitical forecasts.

A key question is where to host such a forecasting initiative. The options include: (1) a revised attempt at crowdsourced geopolitical forecasting based in the intelligence community, the location of key initiatives launched during the Obama administration, either through a new geopolitical forecasting office or as part of the National Intelligence Council; (2) basing future geopolitical forecasting efforts in the DoD; (3) launching a geopolitical forecasting initiative in the White House, either as part of the National Security Council staff or in the Office of Science and Technology Policy; and (4) creating a forecasting platform in Congress, perhaps as part of the Congressional Research Service. This white paper also discusses some other, less likely alternatives, such as placing a forecasting platform in the Treasury Department or the State Department.

The natural home of a crowdsourced geopolitical forecasting platform would be the intelligence community. However, a forecasting initiative could launch and thrive in a number of locations with enough political support.

All options will require engaging Congress, as well as making key decisions about the classification level of any forecasting platform, determining the methodology for eliciting forecasts (primarily, using a prediction market versus a forecast aggregation platform), delineating the requirements for using forecasts, and deciding how to set up leaderboards in ways that generate accountability but also encourage participation. Importantly, there are likely genuine tradeoffs between maximizing the use of forecasting platforms for accountability and job training, versus implementing a bureaucratically feasible approach that could generate long-term success. There are also different choices if the goal of the platform is to identify superforecasters and then have those superforecasters become professional forecasters within the bureaucracy, versus having larger-scale participation over time from a broad pool.<sup>4</sup>

Finally, it is also critical to recognize the limits of quantitative and crowdsourced geopolitical forecasting methods. Crowdsourced forecasting is not a magic bullet but an additional tool in the tool kit that can significantly improve forecasting accuracy in some

areas. These methods are most likely to work in combination with other, more traditional forecasting methods. For example, scenario-planning exercises could cue more specific, quantifiable, forecasting questions. Approaching these methods as complements, rather than substitutes, for traditional futures analysis will also increase the chance of adoption.

Given those issues, this white paper makes a number of recommendations on the details of implementation strategy:

- Promote crowdsourced and quantitative geopolitical forecasting as a complement to existing forecasting methods, particularly given evidence suggesting they may actually prove more successful in combination.
- Require the use of the outputs of a forecasting platform in relevant products, including analyst engagement when they disagree with the forecast.
- Design a communication system to effectively translate probabilities into useful information for policymakers in a way that conveys the information without generating false confidence.
- Create both classified and unclassified forecasting platforms. This will encourage broad participation, which will improve forecasting accuracy on most questions, while still presenting a classified platform for questions where access to classified information is necessary to even consider the question.
- Decide on the relative importance of identifying and cultivating superforecasters as a programmatic goal prior to launching a crowdsourced forecasting platform.
- Use prediction polls, aggregating them with algorithms and other proven methods, rather than prediction markets. Implement a flexible approach to leaderboards, tracking data on success over time for every forecaster, while creating options to frame it based on self-improvement, in-office “pods,” and other ways that create continuing incentives to participate rather than undermine motivation.
- Fund additional research and development designed to improve the accuracy of crowdsourced geopolitical forecasting methods and efforts to effectively communicate quantitative insights to policymakers.

---

<sup>4</sup> It would be possible to use such a forecasting system to not just have accountability for forecasts, but to tie it to job performance. A challenge with this approach is that, as explained below, most analysts are not hired for their explicit forecasting skills. To then make it a part of how they are evaluated for performance would be difficult, though it might be a possible outcome down the road if the recommendations in this white paper are adopted.



# ➤ WHY CROWDSOURCED GEOPOLITICAL FORECASTING MATTERS

All policymaking involves theory and prediction. The very notion of believing that an action will generate a particular consequence for a particular reason suggests a forecast about action and consequence, and an underlying view or theory of how the world works. Accurately understanding the world is therefore critical to policy success. There are a number of different methods that governments use to understand what is likely to happen in the future, and how policy choices can influence that future. These approaches include traditional qualitative analysis (looking at the evidence and making a judgment based on an interpretation of the evidence), scenario planning, war games, statistical models, and more.

Improved forecasting methods could help the USG better understand the probabilities of future geopolitical developments and allocate resources to reflect the probabilities of these threats. The failure to prepare adequately for the COVID-19 pandemic has only underscored the need for improved probabilistic forecasting and risk assessment. The pandemic generated new proposals for forecasting and early warning systems on global health, to help predict and hopefully prepare for the next pandemic.<sup>5</sup>

Renewed interest in forecasting should go beyond global health, however. Improved forecasting methods could help the United States better understand a wide range of questions, such as: What is the probability that there will be a lethal confrontation between the national military or law enforcement forces of Iran and Saudi Arabia before July 1, 2021? Will a date be set for a referendum on Scotland's status within the United Kingdom before January 1, 2022?<sup>6</sup> Question-clustering approaches, as described [here](#), can even help disentangle broad questions, such as “What factors would signal greater Chinese revisionist moves in the South China Sea?”

A key element of any improved forecasting approach, however, is that it is falsifiable, meaning it can be proved true or false through externally available means. Forecasting approaches that do not generate falsifiable insights may be useful exercises for planning and thinking but ultimately serve somewhat different ends.<sup>7</sup> Compare the question of “Will Iran launch an intercontinental ballistic missile (ICBM) before June 1, 2022?” with the question of “Will Iran be more aggressive in 2022?” The ICBM question is falsifiable—

5 Andrew S. Natsios, “Predicting the Next Pandemic: The United State Needs an Early Warning System for Infectious Diseases,” *Foreign Affairs*, July 14, 2020, <https://www.foreignaffairs.com/articles/united-states/2020-07-14/predicting-next-pandemic>.

6 With different time horizons, these were all live questions on Good Judgment Open (<https://www.gjopen.com/>), an online forecasting challenge, as of October 20, 2020.

7 After all, if a forecasting approach does not generate a falsifiable prediction, then advocates of that system can always make arguments to avoid accountability.



either Iran will test an ICBM or not. The question of aggression is more nebulous, since reasonable people might disagree about how to define aggression.

Fortunately, crowdsourced geopolitical forecasting, particularly on sets of discrete, falsifiable questions, can provide early warning of risks and a level of accuracy that can improve U.S. national security policy. Empirical evidence from a decade of experiments and related efforts by the intelligence community beginning around 2010 illustrates that the wisdom of crowds is not just real but applicable to national security problems. Moreover, creating open forecasting platforms can help identify top forecasters within an organization who can then be deployed to focus more on forecasting tasks.<sup>8</sup> Research on these methods, partly based on Intelligence Advanced Research Projects Activity (IARPA) forecasting tournaments, shows that the process of identifying and teaming up top forecasters can offer the benefits of aggregating forecasts and layering in algorithms even without large forecaster pools.<sup>9</sup> To illustrate, the top forecasters in IARPA tournaments, using only open-source information, performed 30 percent better than career intelligence analysts when both groups forecasted on the same questions.<sup>10</sup>

These methods complement existing futures analysis and offer new sources of information. For example, traditional futures analysis, through scenario planning

or war games, can generate testable propositions that are then assessed through quantifiable forecasting questions. Information from qualitative analysis or machine-learning approaches can be fed to forecasters to give them the capacity to improve the calibration of their quantitative forecasts. Additionally, quantitative forecasts from crowds or teams could be inputs in machine-learning models or in scenario planning.

Keeping score does, of course, raise important questions concerning false confidence in numerical probabilities and the way numerical forecasts can look like a “black box,” generating risk in their misuse by policymakers. The risk of false confidence is not a reason to avoid probabilistic forecasting, however, but a reason to develop strong communication and education strategies so analysts and policymakers can correctly interpret forecasts, understand their utility, and recognize their limits.

## DEFINING GEOPOLITICAL FORECASTING

In finance, medicine, and national security, experts routinely make predictions about the future, but social science research shows that subject matter experts are often wrong about their predictions, yet confident in those same predictions.<sup>11</sup> Forecasting complicated geopolitical events will always be challenging, but probabilistic methods can deliver forecasts not unlike weather forecasts, conceptually and with empirically

8 Barbara Mellers, Eric Stone, Terry Murray, et al., “Identifying and Cultivating Superforecasters as a Method of Improving Probabilistic Predictions,” *Perspectives on Psychological Science*, 10(3): 267-281, 2015, <https://journals.sagepub.com/doi/abs/10.1177/1745691615577794#articleCitationDownloadContainer>; Philip E. Tetlock and Dan Gardner, *Superforecasting: The Art and Science of Prediction* (New York: Broadway Books, 2015).

9 Note that these results appear to hold true for smaller sample sizes and limited times to identify forecasters. See: I. Katsagounos, D.D. Thomakos, and K. Litsiouetal, “Superforecasting Reality Check: Evidence from a Small Pool of Experts and Expedited Identification,” *European Journal of Operational Research*, 289(1): 107-117, February 16, 2021, <https://doi.org/10.1016/j.ejor.2020.06.042>. See also: Mellers et al., “Identifying and Cultivating Superforecasters as a Method of Improving Probabilistic Predictions.”

10 Measured using Brier scores. See: Tetlock and Gardner, *Superforecasting*. See also: Seth Goldstein, Rob Hartman, Ethan Comstock, and Thalia Shamash Baumgarten, “Assessing the Accuracy of Geopolitical Forecasts from the U.S. Intelligence Community’s Prediction Market,” Case Number 15-3175, The MITRE Corporation, 2015, <https://goodjudgment.com/wp-content/uploads/2020/11/Goldstein-et-al-GJP-vs-ICPM.pdf>; J. Peter Scoblic and Philip E. Tetlock, “A Better Crystal Ball: The Right Way to Think About the Future,” *Foreign Affairs*, 99(6), November-December 2020, <https://www.foreignaffairs.com/articles/united-states/2020-10-13/better-crystal-ball>.

11 Mellers et al., “Identifying and Cultivating Superforecasters as a Method of Improving Probabilistic Predictions.”

testable levels of accuracy across types of questions.<sup>12</sup> Geopolitical forecasting refers to the process of generating falsifiable predictions about future geopolitical events expressed as numerical probabilities and probability ranges, and scoring or evaluating these predictions. Quantitative geopolitical forecasting involves the explicit use of numerical probabilities or probability ranges in making predictions.

The falsifiability, quantifiability, and the use of scoring for accuracy and accountability distinguish these kinds of forecasts from more common predictions in the government, such as the intelligence community's *Global Trends* reports.<sup>13</sup>

### GENERATING FORECASTS TO HELP INFORM GOVERNMENT PROGRAMS AND POLICY

The USG allocated \$81.7 billion in fiscal year 2019 for the intelligence budget, according to the Office of the Director of National Intelligence.<sup>14</sup> Much of the intelligence budget focuses directly on gathering and analyzing information to better understand the future.

More calibrated forecasts would help the USG prioritize policies and better allocate resources. So-called “failures of imagination” are often really failures of prioritization or resource allocation, which could be corrected using clear probabilistic forecasts. For instance, the threat of a pandemic spreading from Asia to the United States featured in intelligence reports throughout the last 20 years. Then-Senator Barack Obama even penned an op-ed on pandemic threats in *The New York Times* in 2005, but relative risks and changes in those risks are not made explicit in traditional intelligence analysis.<sup>15</sup> Expressing risks as quantitative probabilities—like gambling odds—could aid policymakers in recognizing the rate of change in such large-scale risks. For example, if the probability of a global pandemic in the next year hovered near 1 percent for years, and then one day jumped up to 5 percent, this could signal a need for policy review of surveillance systems, supply stockpiles, and other preventive and preparatory efforts.

Better strategic foresight could also help the USG spend money more efficiently. A forecasting platform could serve as a means for asking questions about the likely impact of government policies and for early warning on programmatic risks, in addition to providing geopolitical futures.

### KEEPING SCORE AS A FORM OF ACCOUNTABILITY AND CONTINUAL JOB TRAINING

The process of keeping score—evaluating how well one's forecasts match actual outcomes—can also serve to improve accountability and train analysts to recognize how cognitive biases may impact their assessments. Forecasting accuracy represents one of the ways to understand what makes an analyst successful. The goal is to generate the best intelligence possible for the USG, of which one element is forecasting.

Quantified scorecards help to elucidate the actual track record of analysts and forecasters, making it easier for the forecasters to self-optimize, for their supervisors to correct biases, and for other branches of government (and potentially the press and the American public, as secrecy regimes allow) to appropriately evaluate government action.

Keeping score as part of geopolitical forecasting efforts may therefore help to address the problem of “accountability ping-pong,” whereby political blame for false positives (predicted events that do not occur) or false negatives (unpredicted events that do occur) leads to superficial adjustments and oscillations.<sup>16</sup> Treating forecasters not as oracles but as performers with “batting averages” tracked on scorecards would help the USG establish baselines about the probability of different types of events and reduce bias in its analysts.<sup>17</sup> For instance, an intelligence analyst whose 70 percent confidence predictions are wrong 30 percent of the time should not be punished for being wrong some of the time, but rewarded for being well-calibrated—her confidence in her forecasts is a good predictor of how frequently those forecasts come true.

12 Caitlin Rivers and Dylan George, “How to Forecast Outbreaks and Pandemics: America Needs the Contagion Equivalent of the National Weather Service,” *Foreign Affairs*, June 29, 2020, <https://www.foreignaffairs.com/articles/united-states/2020-06-29/how-forecast-outbreaks-and-pandemics>.

13 Michael C. Horowitz and Philip E. Tetlock, “Trending Upward: How the Intelligence Community Can Better See into the Future,” *Foreign Policy*, September 7, 2012, <https://foreignpolicy.com/2012/09/07/trending-upward/>.

14 Office of the Director of National Intelligence, “U.S. Intelligence Community Budget,” last updated March 18, 2019, <https://www.dni.gov/index.php/what-we-do/ic-budget>.

15 Barack Obama and Richard Lugar, “Grounding a Pandemic,” *The New York Times*, June 6, 2005, <https://www.nytimes.com/2005/06/06/opinion/grounding-a-pandemic.html>.

16 Phillip E. Tetlock and Barbara A. Mellers, “Intelligent Management of Intelligence Agencies: Beyond accountability ping-pong,” *American Psychologist*, 66(6): 542-554, September 2011, <https://doi.org/10.1037/a0024285>.

17 David R. Mandel, Alan Barnes, and Karen Richards, *A Quantitative Assessment of the Quality of Strategic Intelligence Forecasts*, Defence R&D Canada, Technical Report, DRDC Toronto TR 2013-036, March 2014, [https://cradpdf.drdc-rdc.gc.ca/PDFS/unc142/p538628\\_A1b.pdf](https://cradpdf.drdc-rdc.gc.ca/PDFS/unc142/p538628_A1b.pdf).

The quantitative aspect of such “batting averages” has advantages over relying exclusively on traditional qualitative feedback and training due to increased clarity, continuity, comparability, and objectivity. The clarity of quantitative accountability measures—such as a single easily understood score—would simplify feedback and accountability measures. Relatedly, the continuity of this kind of quantitative feedback would allow forecasters to understand and improve their performance in real time.

In addition to the clarity and continuity of quantitative scoring, the comparability of such scores both over time and across individuals or teams increases objectivity and decreases opportunities for favoritism. Measurement and quantification represent building blocks of objective understanding or at least of decreasing space for biases, value judgments, and related factors relevant for objective accountability and feedback; in the words of Lord Kelvin, “When you cannot express it in numbers, your knowledge is of a meagre and unsatisfactory kind.”<sup>18</sup>

This kind of assessment would also help agencies identify and cultivate top-performing forecasters, or “superforecasters.” Top performers in geopolitical forecasting tournaments tend to excel at “inductive reasoning, pattern detection, cognitive flexibility, and open-mindedness.”<sup>19</sup> Keeping score would help U.S. national security organizations select, train, and promote individuals and teams more likely to succeed at geopolitical forecasting. This would increase efficiency and create a more meritocratic workplace.

Finally, keeping score would help in continual job training, useful even for non-forecasting tasks, by making it easier to identify over- and under-confidence, as well as other cognitive biases. One study of probabilistic forecasts in Canadian strategic intelligence reports, for instance, found that under-confidence represented a major source of inaccuracy.<sup>20</sup> Identifying those biases in individuals and teams could enable training and professional development to trigger improvement. In the words of the National Resource Council’s 2011 report *Intelligence Analysis for Tomorrow*, “Continuous learning and sustainable improvement require systematic procedures to assess performance and provide feedback.”<sup>21</sup>

## KEEPING SCORE TO EVALUATE ANALYTIC METHODS AND SOURCES

In addition to tracking individual performance, the ability to keep score in aggregated geopolitical forecasts also opens new avenues for evaluating analytic methods and intelligence sources. If forecasters or teams of forecasters use different methods or sources, leaders could examine resulting differences in forecast accuracy to improve overall accuracy. For example, scorekeeping could help to show whether Analysis of Competing Hypotheses approaches contribute to accuracy or add unnecessary work for analysts. On sources, forecasting platforms could help decision makers understand differences in accuracy of forecasts based on, for example, human intelligence versus signals intelligence. These are important questions that the intelligence community could answer more thoroughly using data generated by forecasting aggregation platforms.

---

18 On quantification and objectivity (and Kelvin), see: Julian Reiss and Jan Sprenger, “Scientific Objectivity,” *The Stanford Encyclopedia of Philosophy*, <https://plato.stanford.edu/entries/scientific-objectivity/#MeasQuan>.

19 Barbara Mellers, Eric Stone, Pavel Atanasov, et al., “The Psychology of Intelligence Analysis: Drivers of Prediction Accuracy in World Politics,” *Journal of Experimental Psychology: Applied*, 21(1): 1-14, 2015, <https://www.apa.org/pubs/journals/releases/xap-0000040.pdf>.

20 Mandel et al., *A Quantitative Assessment of the Quality of Strategic Intelligence Forecasts*.

21 National Resource Council, *Intelligence Analysis for Tomorrow: Advances from the Behavior and Social Sciences* (Washington, DC: National Academies Press, 2011), 23, <https://www.nap.edu/catalog/13040/intelligence-analysis-for-tomorrow-advances-from-the-behavioral-and-social>.

# ➤ EVALUATING U.S. GOVERNMENT EFFORTS AT GEOPOLITICAL FORECASTING

Given the advantages of falsifiable forecasting outlined in the previous section, why have these methods failed to take off within the national security community and the USG more broadly? Efforts from the Obama administration to introduce crowdsourced geopolitical forecasting in the USG have stalled in recent years. The decline mirrors the 1970s attempt to introduce quantifiable expressions of uncertainty and earlier work on words of estimative probability by Sherman Kent, the “father of the modern intelligence profession,” who once quipped, “I’d rather be a bookie than a goddamn poet.”<sup>22</sup> In the 1970s, Defense Intelligence Agency experimentation and intelligence community training courses suggested real value to quantifiable geopolitical forecasting, but the effort collapsed, as James Marchio writes, with “the departure of officials who had pushed strongly for greater precision in

conveying analytic uncertainty.”<sup>23</sup> Put more simply, 1970s efforts faltered because they were not institutionalized enough to survive leadership turnover. What has happened today?

## BACKGROUND ON RECENT USG GEOPOLITICAL FORECASTING EFFORTS

Since the early 2000s, the U.S. intelligence community has launched a number of crowdsourced forecasting efforts. These efforts followed incidents, from the fall of the Berlin Wall to the September 11, 2001, attacks, that generated criticism of the intelligence community, fairly or not, for a failure to anticipate events around the world. At the same time, the intelligence community also faced critiques, again fairly or not, for documents like the 2002 National Intelligence Estimate, which supported claims that Iraq had weapons of mass destruction.<sup>24</sup>

In 2004, Congress passed the Intelligence Reform and Terrorism Prevention Act,<sup>25</sup> which emphasizes developing better coordinated interagency work and information sharing between the 15 different intelligence agencies as well as improved analytical methods. In response to scientific advances, the intelligence community and the DoD also initiated experiments designed to adapt prediction markets and crowdsourcing to the needs of the intelligence

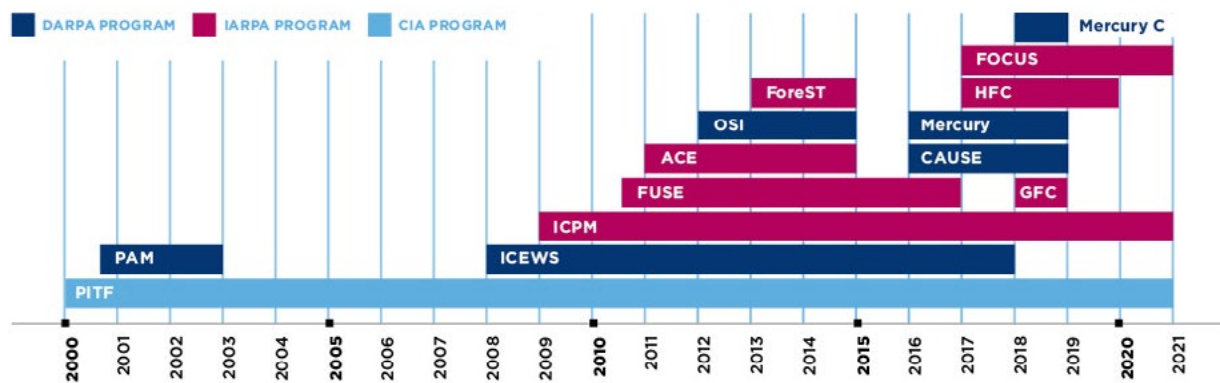
22 J. Peter Scoblic, “Beacon and Warning: Sherman Kent, Scientific Hubris, and the CIA’s Office of National Estimates,” *Texas National Security Review*, 1(4), 2018, <https://tnsr.org/2018/08/beacon-and-warning-sherman-kent-scientific-hubris-and-the-cias-office-of-national-estimates/>.

23 James Marchio, “‘If the Weatherman Can...’: The Intelligence Community’s Struggle to Express Analytic Uncertainty in the 1970s,” *Studies in Intelligence*, 58(4): 31-42, December 2014. <https://www.cia.gov/library/center-for-the-study-of-intelligence/csi-publications/csi-studies/studies/vol-58-no-4/pdfs/If%20the%20Weatherman%20Can.pdf>.

24 Congressional Record, E1545-E1546, July 21, 2003, [https://fas.org/irp/congress/2003\\_cr/h072103.html](https://fas.org/irp/congress/2003_cr/h072103.html).

25 Intelligence Reform and Terrorism Prevention Act of 2004, Public Law 108-458, December 17, 2004, [https://www.dni.gov/files/NCTC/documents/RelatedContent\\_documents/Intelligence\\_Reform\\_Act.pdf](https://www.dni.gov/files/NCTC/documents/RelatedContent_documents/Intelligence_Reform_Act.pdf).

**Figure 1: Timeline of Intelligence Community Forecasting Efforts**



CE: Aggregative Contingent Estimation  
 CAUSE: Cyber-attack Automated Unconventional Sensor Environment  
 FOCUS: Forecasting Counterfactuals in Uncontrolled Settings  
 ForeST: Forecasting Science and Technology  
 FUSE: Foresight and Understanding from Scientific Exposition  
 GFC: Geopolitical Forecasting Challenge  
 HFC: Hybrid Forecasting Competition

ICEWS: Integrated Conflict Early Warning System  
 ICPM: Intelligence Community Prediction Market  
 Mercury  
 Mercury C: Mercury Challenge  
 OSI: Open Source Indicators  
 PAM: Policy Analysis Market  
 PITF: Political Instability Task Force (Note: began in 1994)

community. These experiments included the Defense Advanced Research Projects Agency’s (DARPA) short-lived Futures Markets Applied to Prediction/Policy Analysis Market, and several programs at IARPA, including Aggregative Contingent Estimation, Forecasting Science and Technology program, Open-Source Indicators, Hybrid Forecasting Competition, Geopolitical Forecasting Challenge, Forecasting

Counterfactuals in Uncontrolled Settings (FOCUS), as well as the Intelligence Community Prediction Market.<sup>26</sup>

Other programs on quantitative forecasting focused on statistical models and machine learning, including DARPA’s Integrated Conflict Early Warning System (ICEWS), and IARPA’s Foresight and Understanding from Scientific Exposition (FUSE), Open-Source

26 For an overview of the listed projects, see: Policy Analysis Market, “A Market in the Future of the Middle East: PAM Concept Overview,” 2003, [https://ratical.org/ratville/CAH/linkscopy/PAM/pam\\_concept.htm](https://ratical.org/ratville/CAH/linkscopy/PAM/pam_concept.htm); Office of the Director of National Intelligence, “Aggregative Contingent Estimation (ACE),” Intelligence Advanced Research Projects Activity, <https://www.iarpa.gov/index.php/research-programs/ace/baa>; Leah McGrath Goodman, “The EMBERS Project Can Predict the Future With Twitter,” Intelligence Advanced Research Projects Activity, Office of the Director of National Intelligence, March 07, 2015, <https://www.iarpa.gov/index.php/newsroom/iarpa-in-the-news/2015/461-the-embers-project-can-predict-the-future-with-twitter>; Office of the Director of National Intelligence, “Forecasting Science & Technology (ForeST),” Intelligence Advanced Research Projects Activity, <https://www.iarpa.gov/index.php/research-programs/forest>; Office of the Director of National Intelligence, “Foresight and Understanding from Scientific Exposition (F.U.S.E.),” Intelligence Advanced Research Projects Activity, <https://www.iarpa.gov/index.php/research-programs/fuse/baa>; Office of the Director of National Intelligence, “Prize Challenges: Geopolitical Forecasting Challenge,” Intelligence Advanced Research Projects Activity, <https://www.iarpa.gov/index.php/working-with-iarpa/prize-challenges/1070-geopolitical-forecasting-challenge>; Office of the Director of National Intelligence, “Hybrid Forecasting Competition (HFC),” Intelligence Advanced Research Projects Activity, <https://www.iarpa.gov/index.php/research-programs/hfc>; and Office of the Director of National Intelligence, “Forecasting Counterfactuals in Uncontrolled Settings (FOCUS),” Intelligence Advanced Research Projects Activity, <https://www.iarpa.gov/index.php/research-programs/focus/focus-baa>.



Indicators (OSI), Cyber-attack Automated Unconventional Sensor Environment (CAUSE), Mercury, and the Mercury Challenge. A precursor to these programs, and an inspiration in particular for both the ICEWS and OSI projects, the CIA's Political Instability Task Force was formed in 1994 to "assess and explain the vulnerability of states around the world to political instability and state failure."<sup>27</sup> Since then, the Political Instability Task Force has continued to calculate probabilistic forecasts for up to 167 countries.<sup>28</sup>

### ***Policy Analysis Market***

The Policy Analysis Market emerged in late 2001, one of the first projects to do so after DARPA issued a call for proposals under the title of "Electronic Market-Based Decision Support," later renamed "Future Markets Applied to Prediction" (FutureMAP). The program objective stated that FutureMAP would concentrate on market-based techniques for predicting future events.<sup>29</sup> The winning proposal featured a Policy Analysis Market that, according to the project's archived website, would serve as a "market in the future of the Middle East" (in particular, it was to focus on "the economic, civil, and military futures of Egypt, Jordan, Iran, Iraq, Israel, Saudi Arabia, Syria, and Turkey") and "provide insight into the interactions among Middle Eastern and U.S. interests and policy decisions."<sup>30</sup> After receiving harsh public criticism that the Policy Analysis Market was essentially a betting parlor on assassinations and terrorist attacks, the program was canceled in 2003 before the market opened for trading.

### ***Integrated Crisis Early Warning System***

Launched in 2008 as a DARPA program, the ICEWS data and model are supported by U.S. Strategic

Command and operated by Lockheed Martin.<sup>31</sup> The object of the program was to "develop a comprehensive, integrated, automated, generalizable, and validated system to monitor, assess, and forecast national, subnational, and international crises in a way that supports decisions on how to allocate resources to mitigate them." It did so by seeking to answer three essential questions: (1) what countries are likely to become more or less unstable? (2) what are the factors driving that instability? and (3) what combination of efforts are likely to have the greatest chance of mitigating the instability? Lockheed argued that the program has a number of benefits, including a forecast accuracy of 80 percent, providing near real-time data, trend visualization, and analysis, as well as a "mixed methods approach to instability events of interest (EOI) forecasting," and open-source sentiment analysis, all presented as an interactive, web-based portal.<sup>32</sup> The program was considered a success among the DoD's Human and Sociocultural Behavior Capability projects.<sup>33</sup> The main ICEWS program ended in 2018.<sup>34</sup>

### ***Intelligence Community Prediction Market***

The Intelligence Community Prediction Market (ICPM) launched in 2010 on a top secret, classified network.<sup>35</sup> Government employees and contractors with appropriate clearances could participate. The ICPM was a prediction market in which participants used play money to buy and sell shares in the outcomes of geopolitical forecasting questions. The ICPM aimed to support decision makers and analysts and to aid in identifying the best forecasters in the intelligence community, as well as to provide "a test for future study."<sup>36</sup>

Although the self-selected forecasters represented a small share of the intelligence community, one analysis suggests

27 Political Instability Task Force, "Internal Wars and Failures of Governance, 1955-2005," George Mason University (archived from the original), last updated October 3, 2006, <https://web.archive.org/web/20061208000556/http://globalpolicy.gmu.edu/pitf/>.

28 Cameron Evers, "The CIA Has a Team of Clairvoyants," *The Week*, July 14, 2016, <https://theweek.com/articles/635515/cia-team-clairvoyants>; Jack A. Goldstone, Robert H. Bates, David L. Epstein, et al., "A Global Model for Forecasting Political Instability," *American Journal of Political Science*, 54(1): 190-208, January 2010, <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1540-5907.2009.00426.x>.

29 Robert E. Looney, "DARPA's Policy Analysis Market for Intelligence: Outside the Box or Off the Wall?" *International Journal of Intelligence and Counter Intelligence*, 17(3): 405-419, 2004, <http://hdl.handle.net/10945/42204>.

30 Policy Analysis Market, "A Market in the Future of the Middle East," 2003, [https://ratical.org/ratville/CAH/linkscopy/PAM/pam\\_home.htm](https://ratical.org/ratville/CAH/linkscopy/PAM/pam_home.htm).

31 Sean P. O'Brien, "Crisis Early Warning and Decision Support: Contemporary Approaches and Thoughts on Future Research," *International Studies Review*, 12(1): 87-104, March 2010, <https://www.jstor.org/stable/40730711?seq=1>; Harvard Dataverse, "Integrated Crisis Early Warning System (ICEWS) Dataverse," Harvard University, <https://dataverse.harvard.edu/dataverse/icews>.

32 Lockheed Martin Corporation, "Integrated Crisis Early Warning System (ICEWS)," <https://www.lockheedmartin.com/en-us/capabilities/research-labs/advanced-technology-labs/icews.html#:~:text=The%20Integrated%20Crisis%20Early%20Warning,allocate%20resources%20to%20mitigate%20crisis>.

33 John Boiney and David Foster, "Progress and Promise: Research and Engineering for Human Sociocultural Behavior Capability in the U.S. Department of Defense," The MITRE Corporation: 25, March 15, 2013, <https://apps.dtic.mil/dtic/tr/fulltext/u2/a587451.pdf>.

34 James Starz, Mark Hoffman, Joe Roberts, et al., "Supporting Situation Understanding (Past, Present, and Implications on the Future) in the STRATCOM ISPAN Program of Record," in: Dylan D. Schmorrow and Denise M. Nicholson (eds.), *Advances in Design for Cross-Cultural Activities Part I* (Boca Raton, FL: CRC Press, July 6, 2012), <https://www.taylorfrancis.com/books/e/9780429108716/chapters/10.1201%2Fb12316-23>.

35 Jonathan McHenry, "Three IARPA Forecasting Efforts: ICPM, HFC, and the Geopolitical Forecasting Challenge," PowerPoint Presentation, Federal Foresight Community of Interest 18th Quarterly Meeting, January 26, 2018, <https://www.ffcoi.org/wp-content/uploads/2019/03/Three-IARPA-Forecasting-Efforts-ICPM-HFC-and-the-Geopolitical-Forecasting-Challenge-Jan-2018.pdf#page=5>.

36 Alexander Halman, "Before and Beyond Anticipatory Intelligence: Assessing the Potential for Crowdsourcing and Intelligence Studies," *Journal of Strategic Security*, 8(3): 15-24, 2015, <https://www.jstor.org/stable/26465241>.

that “ICPM has the largest dataset on the accuracy of analytic judgments in the history of the IC, including > 190,000 predictions made by > 4,300 users on a large array of geopolitical questions.”<sup>37</sup> One study of the ICPM suggests that its predictions were significantly more accurate than those made in comparable intelligence reports by professional intelligence analysts, though others have questioned that analysis.<sup>38</sup>

One challenge is that the ICPM was designed to predict events, not to explain why they were or were not going to occur. The lack of an explanation of the causal logic behind a forecast made it difficult for analysts to trust the forecasts, for analysts to communicate them to policymakers, and, in turn, for policymakers to trust them or use them.

### **Open-Source Indicators**

OSI was an IARPA program for anticipatory intelligence that automatically generated forecasts using open-source data. OSI forecasts started in April 2012 in coordination with a group of research universities and industry contractors. It transitioned in 2015, and a version of it focused on instability forecasts is running within the intelligence community.<sup>39</sup> The automated predictions of OSI systems focused largely on population-level events, including but not limited to civil unrest, election outcomes, or disease outbreaks in Latin America, later expanded to the Middle East and North Africa.<sup>40</sup> The system was fed a wide range of open-source data, from tweets to economic indicators, and OSI forecasts were scored using human “Gold Standard Reports” at the MITRE Corporation.<sup>41</sup> Although the

basis of the program’s sentiment analysis has been criticized due to its reliance on an American English lexicon, OSI reportedly produced several successful predictions of major events, including the 2013 Brazilian Spring and early 2014 protests in Venezuela.<sup>42</sup>

### **Aggregative Contingent Estimation**

The Aggregative Contingent Estimation (ACE) program represents the most significant test, and most substantial success, for crowdsourced forecasting in the U.S. intelligence community. The goal of the ACE program was to “dramatically enhance the accuracy, precision, and timeliness of forecasts for a broad range of event types, through the development of advanced techniques that elicit, weight, and combine the judgments of many intelligence analysts.”<sup>43</sup> After the first year of the project, the results demonstrated that the statistical aggregation models used by ACE “consistently [beat] the baseline forecasts (the unweighted average forecasts).”<sup>44</sup> Over the course of the program, the winning research team, the Good Judgment Project, used several different methods, including algorithms layered on top of aggregated forecasts and teams of forecasters with superior accuracy rates compared with the ICPM on the same questions.<sup>45</sup>

The Forecasting Science and Technology (ForeST) program directly leveraged the “programmable and technical achievements of IARPA’s ACE and FUSE programs.”<sup>46</sup> ForeST focused specifically on questions of science and technology (S&T).<sup>47</sup> The program generated SciCast, one of the largest S&T forecasting tournaments in the world, but it went on hiatus when the project lost funding in 2015.<sup>48</sup>

37 McHenry, “Three IARPA Forecasting Efforts: ICPM, HFC, and the Geopolitical Forecasting Challenge.”

38 Bradley J. Stastny and Paul E. Lehner, “Comparative Evaluation of the Forecast Accuracy of Analysis Reports and a Prediction Market,” *Judgment and Decision Making*, 13(2): 202-211, March 2018, <http://journal.sjdm.org/17/17815/jdm17815.html>. In contrast, see: David R. Mandel, “Too Soon to Tell if the U.S. Intelligence Community Prediction Market is More Accurate than Intelligence Reports: Commentary on Stastny and Lehner (2018),” *Judgment and Decision Making*, 14(3): 288-292, May 2019, <http://journal.sjdm.org/19/190417/jdm190417.html>.

39 Leah McGrath Goodman, “The EMBERS Project Can Predict the Future with Twitter,” *Newsweek*, March 7, 2015, <https://www.newsweek.com/2015/03/20/embers-project-can-predict-future-twitter-312063.html>.

40 Andy Doyle, Graham Katz, Kristen Summers, et al., “Forecasting Significant Societal Events Using the EMBERS Streaming Predictive Analytics System,” *Big Data*, 2(4): 185-195, December 2014, <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4276118/>.

41 McGrath Goodman, “The EMBERS Project Can Predict the Future with Twitter.”

42 Heather M. Roff, “Forecasting and Predictive Analytics: A Critical Look at the Basic Building Blocks of a Predictive Model,” *Brookings Tech Stream*, The Brookings Institution, September 11, 2020, <https://www.brookings.edu/techstream/forecasting-and-predictive-analytics-a-critical-look-at-the-basic-building-blocks-of-a-predictive-model/#fn2>; Sathappan Muthiah, Patrick Butler, Rupinder Paul Khandpur, et al., “EMBERS at 4 years: Experiences operating an Open Source Indicators Forecasting System,” *KDD '16: ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016, <https://arxiv.org/pdf/1604.00033.pdf>.

43 Jason Matheny, “Aggregative Contingent Estimation (ACE),” PowerPoint Presentation, Proposer’s Day, May 19, 2010, [https://www.iarpa.gov/images/files/programs/ace/ACE\\_Proposers\\_Day\\_Brief.pdf](https://www.iarpa.gov/images/files/programs/ace/ACE_Proposers_Day_Brief.pdf); Office of the Director of National Intelligence, “Aggregative Contingent Estimation (ACE).”

44 Dirk B. Warnaar, Edgar C. Merkle, Mark Steyvers, et al., “The Aggregative Contingent Estimation System: Selecting, Rewarding, and Training Experts in a Wisdom of Crowds Approach to Forecasting,” Association for the Advancement of Artificial Intelligence, 2012, <https://www.aaai.org/ocs/index.php/SSS/SSS12/paper/viewFile/4290/4699>.

45 Seth Goldstein, Rob Hartman, Ethan Comstock, Thalia Shamash Baumgarten, “Assessing the Accuracy of Geopolitical Forecasts from the U.S. Intelligence Community’s Prediction Market,” <https://goodjudgment.com/wp-content/uploads/2020/11/Goldstein-et-al-GJP-vs-ICPM.pdf>.

46 Office of the Director of National Intelligence, “Forecasting Science & Technology (ForeST);” “Foresight and Understanding from Scientific Exposition (FUSE).”

47 Matheny, “Forecasting Science & Technology (ForeST).”

48 Jessie Jury and Community Contributor, “SciCast Launch! Crowdsourced Forecasting Project from George Mason University,” *Chicago Tribune*, December 16, 2013, <https://www.chicagotribune.com/suburbs/community/chi-ugc-article-sci-cast-launch-crowdsourced-forecasting-proj-2013-12-16-story.html>; *The Official SciCast Blog*, “So Long, and Thanks for All the Fish!,” George Mason University, June 08, 2015, <https://web.archive.org/web/20160304114756/http://blog.sci-cast.org/2015/06/08/so-long-and-thanks-for-all-the-fish/>.



Other Forecasting Projects (Geopolitical Forecasting Challenge, Hybrid Forecasting Competition, CAUSE, Mercury, Mercury Challenge, and FOCUS).

The promising results of the ACE program catalyzed broader interest in probabilistic forecasting and prediction markets. In response, IARPA launched three new forecasting projects in 2017 and 2018: the Geopolitical Forecasting Challenge (GFC), Hybrid Forecasting Competition (HFC), and FOCUS.

The GFC was a forecasting competition initiated by IARPA in 2018 that invited “solvers from around the world to develop innovative solutions and methods for integrating crowdsourced forecasts and other data into accurate, timely forecasts on worldwide issues.”<sup>49</sup> The competition dovetailed with IARPA’s larger competition-based geopolitical forecasting program, the HFC (also known as SAGE, Synergistic Anticipation of Geopolitical Events).<sup>50</sup> HFC sought to build on the ACE program by developing and testing a combination of human and machine geopolitical forecasting systems.<sup>51</sup> HFC terminated in 2020. CAUSE forecasted cyberattacks from a variety of unclassified data.<sup>52</sup> Mercury analyzed classified signals intelligence to forecast military mobilization and disease outbreaks.<sup>53</sup> Mercury Challenge was an unclassified tournament to forecast the same events.<sup>54</sup> FOCUS, similarly building on ACE, works to develop better counterfactual forecasting systems and explanations for forecasts.<sup>55</sup>

## BACKSLIDING IN INTELLIGENCE COMMUNITY FORECASTING EFFORTS

Whether the projects simply came to a natural end or lost bureaucratic support, according to publicly available information, few are still active. Given the demonstrated need for probabilistic forecasting methods by the USG, and the positive reception and outcome of many of these projects, why has adoption lagged as these initiatives shift

from research and development to potential programs?

Broadly, the USG generally faces challenges in converting promising research and development efforts into programs of record. A Government Accountability Office (GAO) report in 2015, focused on the DoD, described a “valley of death” that engulfs advances in S&T before they get in the hands of the warfighter.<sup>56</sup>

There are several factors related to the valley of death in defense acquisitions that also could undermine the adoption of crowdsourced geopolitical forecasting methods by the USG. First, sometimes programs are successes in S&T, but do not have a clear end user, delaying technologies ready for broader national security adoption. Just within the intelligence community, for example, one could make arguments for creating a separate forecasting office in the Office of the Director of National Intelligence or placing a forecasting platform within the National Intelligence Council. Without a clear end user eager to pull technology from S&T into a specific program, it is easy for mature, useful, technology to hit adoption roadblocks.

Second, S&T programs generally focus on developing specific technical advances or breakthroughs, like stealth or new types of sensors. S&T based in social and behavioral science falls outside the rubric of the “normal” type of S&T conversion into a program. As social and behavioral science initiatives, S&T efforts involving geopolitical forecasting thus create challenges for a bureaucracy not used to thinking about advances in analytic techniques as a technological advance. Related to these issues is finding dedicated financial support for intelligence community forecasting efforts. In the private sector, advertisements and subscriptions can help fund crowdsourced forecasting platforms, but these models are not easily transferable to the government. Investment for non-hardware efforts such as behavioral science therefore remains a challenge.

49 Warnaar, Merkle, Steyvers, et al., “The Aggregative Contingent Estimation System: Selecting, Rewarding, and Training Experts in a Wisdom of Crowds Approach to Forecasting”; Office of the Director of National Intelligence, “Geopolitical Forecasting (GF) Challenge.”

50 Office of the Director of National Intelligence, “Hybrid Forecasting Competition (HFC)”; Fred Morstatter, Aram Galstyan, Gleb Satyukov, et al., “SAGE: A Hybrid Geopolitical Event Forecasting System,” Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence (IJCAI-19), 6557-59, August 2019, <https://www.ijcai.org/Proceedings/2019/0955.pdf>; ScienceBlog, “It’s Not Magic, It’s Science: Predicting The Future,” February 24, 2020, <https://scienceblog.com/514351/its-not-magic-its-science-predicting-the-future/>.

51 Office of the Director of National Intelligence, “IARPA Launched Hybrid Forecasting Competition To Improve Predictions Through Human-Machine Integration,” Intelligence Advanced Research Projects Activity, August 18, 2017, <https://www.dni.gov/index.php/newsroom/press-releases/item/1785-iarpa-launches-hybrid-forecasting-competition-to-improve-predictions-through-human-machin>.

52 Office of the Director of National Intelligence, “Cyber-attack Automated Unconventional Sensor Environment (CAUSE),” Intelligence Advanced Research Projects Activity, <https://www.iarpa.gov/index.php/research-programs/cause/cause-baa>.

53 Office of the Director of National Intelligence, “Mercury,” Intelligence Advanced Research Projects Activity, <https://www.iarpa.gov/index.php/research-programs/mercury>.

54 Office of the Director of National Intelligence, “Mercury Challenge,” Intelligence Advanced Research Projects Activity, <https://www.iarpa.gov/challenges/mercury.html>.

55 Office of the Director of National Intelligence, “Forecasting Counterfactuals in Uncontrolled Settings (FOCUS).”

56 U.S. Government Accountability Office, *Defense Advanced Research Projects Agency: Key Factors Drive Transition of Technologies, But Better Training and Data Dissemination Can Increase Success* (Washington, DC: U.S. Government Accountability Office, November 2015), 3-5, <https://www.gao.gov/assets/680/673746.pdf>.

Third, sometimes programs advance past the S&T stage too early, before they are mature enough for easy adoption by end users. As the GAO wrote: “This chasm, often referred to by department insiders as “the valley of death,” exists because the acquisition community often requires a higher level of technology maturity than the S&T community is willing to fund and develop.”<sup>57</sup>

Additionally, geopolitical forecasting initiatives in the intelligence community revealed organizational incentives problems that complicated building a broad, sustainable user pool and support within the bureaucracy. Focusing just on the ICPM, though similar issues could apply to other projects, possible issues included:

- absence of material incentives;
- challenge to status hierarchies;
- threat of accountability consequences;
- intransigence of bureaucratic politics; and
- lack of forecasting core competency.

### ***Material Incentives***

Open-source accounts of the ICPM show that participation was voluntary. Participants did not receive financial incentives or work-related incentives (such as integration of performance in ways that might lead to promotions).<sup>58</sup> ICPM participants used “play money”—points that could be wagered but not exchanged for real money or goods—which in aggregate created prices that indicated the market’s predictions.<sup>59</sup> Though some research suggests that there is little difference between real money and play money incentives for prediction market accuracy, these studies compared already-active markets.<sup>60</sup>

One could solve the incentives problem by allowing forecasters to use real money. However, there are also substantial downsides to government prediction markets that put real money on the line. As described above, in 2003, discussions of national security prediction markets led to concern about foreign actors attempting to exploit the market to make money, or even drive probabilities in ways that could distort U.S. foreign policy.<sup>61</sup>

Money aside, participants that experienced substantial success did not receive any direct commendation or credit for their performance in ways that would make promotions or pay raises more likely.

### ***Status Hierarchies***

Forecasting platforms, whether prediction polls or markets, create new sources of information for analysts and policymakers. In any bureaucracy, innovations that generate new sources of information and potentially challenge the veracity of existing sources of information create organizational challenges. They threaten traditional hierarchies and challenge notions of meritocracy. The economist Tyler Cowen stipulates that prediction markets fail to take off in large organizations in part because “prediction markets threaten the hierarchical control of top managers.”<sup>62</sup> For example, forecasting tournaments may show, in some cases, that a low-level analyst with little experience is better calibrated at forecasting tasks than senior experts. Experts in the middle of the bureaucracy may therefore have incentives to downplay, ignore, or block the use of new methods that threaten internal organizational perceptions of their expertise.

### ***Accountability Consequences***

Relatedly, though the impact was slightly small due to limited adoption, systems like the ICPM create risks for institutional morale. First, established employees may view the potential disruption wrought by a mechanism that outperforms many traditional analysts with a sense of impending doom, as a factory worker might view a new assembly robot on the line with similar suspicion. Falsifiable scoring, while generating accountability, also potentially reveals that most people are not great forecasters. In other words, “Prediction markets make a big chunk of the bettors into ‘losers.’”<sup>63</sup>

57 The case of the ICPM is somewhat different, though, because it did not grow out of a formal S&T program.

58 McHenry, “Three IARPA Forecasting Efforts: ICPM, HFC, and the Geopolitical Forecasting Challenge.”

59 Ibid.

60 Emile Servan-Schreiber, Justin Wolfers, David M. Pennock, Brian Galebach, “Prediction Markets: Does Money Matter?” *Electronic Markets*, 14(3), 2004, <https://www.tandfonline.com/doi/abs/10.1080/1019678042000245254>.

61 Carl Hulse, “Threats and Responses: Plans and Criticisms; Pentagon Prepares A Futures Market On Terror Attacks,” *The New York Times*, July 29, 2003, <https://www.nytimes.com/2003/07/29/us/threats-responses-plans-criticisms-pentagon-prepares-futures-market-terror.html>.

62 Tyler Cowen, “Why Don’t More Businesses Use Prediction Markets,” *Marginal Revolution*, March 15, 2016, [https://marginalrevolution.com/marginalrevolution/2006/03/why\\_dont\\_busine.html](https://marginalrevolution.com/marginalrevolution/2006/03/why_dont_busine.html).

63 Cowen, “Why Don’t More Businesses Use Prediction Markets?”

## Bureaucratic Politics

Classic organizational and bureaucratic politics issues generated cultural dynamics, especially between different subgroups in the intelligence community.<sup>64</sup> Culturally, organizations built on secrecy and classified information may view the sharing of predictions on an open (even if classified) market with suspicion. This is particularly true when the forecast comes without the causal logic behind it, which can make the forecast seem like a black box. Moreover, like most large bureaucracies, compartmentalization features heavily within the intelligence community, and the intelligence community contains interagency conflict and other bureaucratic politics. Legal restraints put in place for much of the intelligence community's early history prevented some interagency cooperation, but even in cases where these legal barriers did not apply, information-sharing challenges still characterize some intelligence collection and operations.<sup>65</sup> Continuing to streamline information-sharing challenges as well as fostering a spirit of interagency cooperation will improve on the success that the intelligence community achieved after post-9/11 reforms.<sup>66</sup>

## Core-Competency Issues

A final reason that may have hobbled the success of the ICPM and the adoption of other forecasting technologies is that most analysts do not view forecasting as a core part of their job. Although making predictions about the future is one of many intelligence analysts' responsibilities, they are rarely the only or even the main part of that job. Thus, a China analyst may view her core responsibility as understanding the current political dynamics in China, not making predictions about the future of China's role in the world. This common "not my job" response may be one reason to build an institutional unit that does view forecasting as its core mission, as discussed in *Paths Forward*.

## OTHER MODELS

Forecasting platform models outside the USG may offer lessons for reinvigorated U.S. efforts. In the nongovernmental sector, the Center for Security and Emerging Technology (CSET) at Georgetown launched Foretell,<sup>67</sup> a technology security landscape forecasting platform, in the summer of 2020, recruiting graduate students, academic researchers, and professionals. Foretell is a partnership between CSET and Cultivate Labs, which IARPA used in many forecasting efforts.<sup>68</sup>

CSET published an issue brief describing Foretell's methodology in October 2020. In part, Foretell is designed to address a shortcoming of quantitative forecasting methods in general and crowdsourced methods in particular: The specificity and falsifiability requirements that enable crowdsourced forecasting make it difficult to ask big-picture questions, like "Is China becoming more aggressive as its military power grows?" or "What is the likely trajectory of U.S.-China relations?"<sup>69</sup> Foretell selects metrics that, in aggregate, can help inform historical "trend departures" that help answer big-picture questions like "What will the tech security landscape look like in 2025?" by showing which of multiple preselected scenarios appears most likely.<sup>70</sup> These efforts, and the core concept of combining scenario planning and probabilistic judgments, build on work by the Good Judgment Project to design Bayesian question clusters during the ACE program.<sup>71</sup> It is also consistent with recommendations by Tetlock and Harvard's J. Peter Scoblic.<sup>72</sup>

CSET's platform breaks down big questions about global politics into falsifiable propositions in several steps: (1) "decompose scenarios into predictors;" (2) "identify metrics for the predictors;" (3) "collect historical and forecast data" (this is where the crowd-sourced forecasting comes in); (4) "estimate trend departure;" and (5) aggregate these trend departures.<sup>73</sup>

64 Michael A. Turner, "CIA-FBI Non-Cooperation: Cultural Trait or Bureaucratic Inertia?" *International Journal of Intelligence and CounterIntelligence*, 8(3), 1995, <https://www-tandfonline-com.proxy.library.upenn.edu/doi/abs/10.1080/08850609508435284?needAccess=true&journalCode=ujic20>.

65 Richard A. Best Jr., "Intelligence Information: Need-to-Know vs. Need-to-Share," *Congressional Research Service*, June 6, 2011, <https://fas.org/sgp/crs/intel/R41848.pdf>.

66 Ibid.

67 Center for Security and Emerging Technology, "CSET Foretell Blog," Georgetown University, CSET, <https://www.cset-foretell.com/blog>.

68 Office of the Director of National Intelligence, "Geopolitical Forecasting (GF) Challenge," Intelligence Advanced Research Projects Activity, <https://www.iarpa.gov/challenges/gfchallenge.html>.

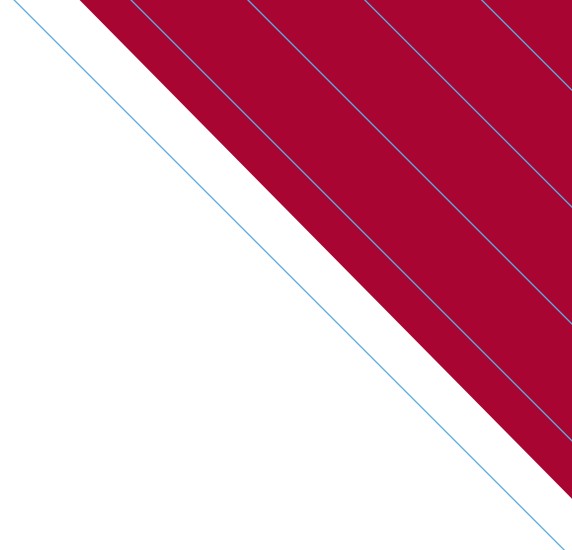
69 Michael Page, Catherine Aiken, and Dewey Murdick, "Future Indices: How Crowd Forecasting Can Inform the Big Picture," Georgetown University, Center for Security and Emerging Technology, October 19, 2020, <https://cset.georgetown.edu/wp-content/uploads/CSET-Future-Indices.pdf>.

70 Ibid., 5-8.

71 Good Judgment, Inc. offers these kind of question clusters as part of their client services. For example, see: Good Judgment, Inc., "Beyond Focus Groups: Actionable Superforecasts of Major Market Trends," Good Judgment, Inc., <https://goodjudgment.com/resources/case-studies/case-study-forecasting-the-future-of-food/>.

72 J. Peter Scoblic and Phillip E. Tetlock, "A Better Crystal Ball: The Right Way to Think About the Future," *Foreign Affairs*, November/December 2020, <https://www.foreignaffairs.com/articles/united-states/2020-10-13/better-crystal-ball>; Philip E. Tetlock, *Full-Inference-Cycle Tournaments: The Quality of Our Questions Matters as Much as the Accuracy of Our Answers*, Prepared for IARPA CASES Contract 2006-1148509-000 Prime Task Order 011, Subcontract S900230BAH, August 30, 2017, [https://www.dropbox.com/s/svl2dlqiztubfj/2017%20FICT\\_Report\\_FINAL%20pdf.pdf?dl=0](https://www.dropbox.com/s/svl2dlqiztubfj/2017%20FICT_Report_FINAL%20pdf.pdf?dl=0).

73 Ibid., 8-13.



Moreover, other countries have made progress in establishing programs to support crowdsourced geopolitical forecasting. In late 2019, the United Kingdom began a major forecasting program. The tournament (known by its project name COSMIC BAZAAR in the UK government) draws on forecasters from across the government, not just the intelligence community or a particular national security agency. Because the program uses open-source data and information, it is easier for a broader group of employees to participate, even though the questions focus on issues surrounding geopolitics and geopolitical risk. The program also links to the priorities of the broader Integrated Security and Defense Review by the Johnson government, which seeks to set the national security strategy of the United Kingdom in a way that leads to success over the next generation.

Though COSMIC BAZAAR is still experimental, the available information suggests it could represent a promising role model. A key question is whether the program has enough momentum to continue over time. If the program ends before advancing past experimentation, despite early promise, it would be additional evidence of the critical need to embed forecasting programs quickly within the national security bureaucracy in a way that gives civil servants a stake in their success. This helps protect the programs from changes in the political winds.

# > KEY ADOPTION QUESTIONS

Creating a more effective approach to geopolitical forecasting in the USG is not just a question of building initial support for adoption and ensuring that intelligence products include explicit numerical probabilities. It will also require addressing specific scientific and organizational challenges to maximize effectiveness.

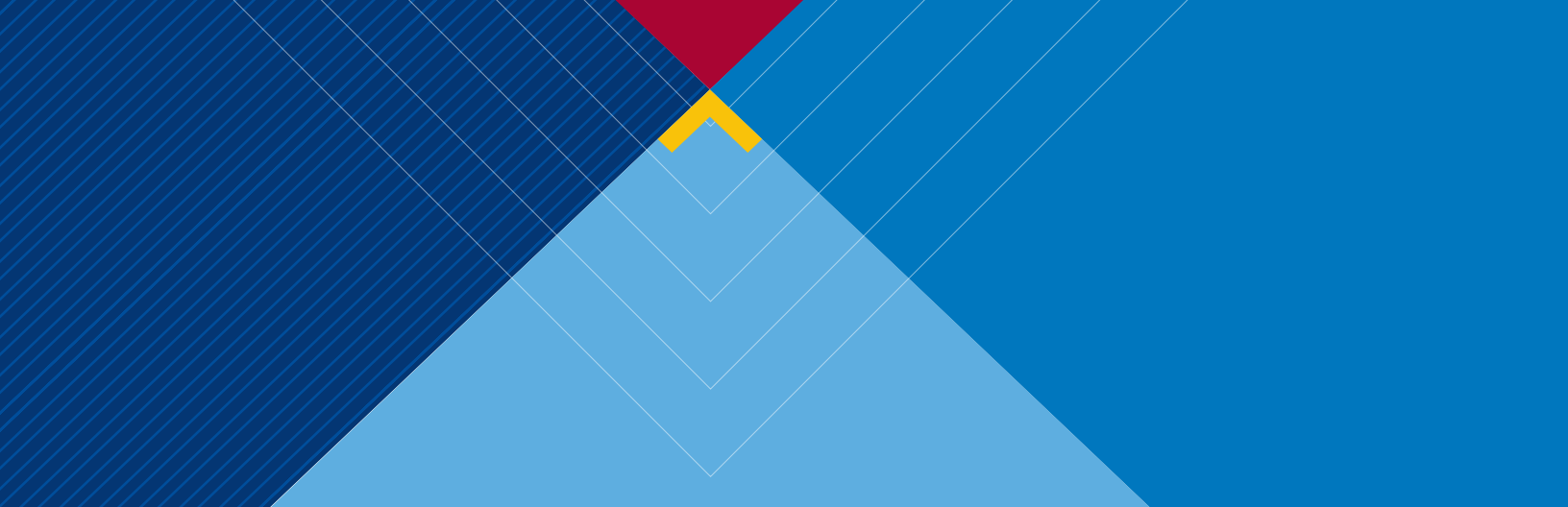
## WHETHER TO MANDATE USE OF GEOPOLITICAL FORECASTING PLATFORMS

At present, there are no mandates surrounding the use of crowdsourced or quantitative forecasts in intelligence products, even when there are forecasting questions directly relevant to the subject of a report. It is up to analysts whether or not to use these reports, and analysts (and their supervisors, and the rest of the intelligence community) may have incentives, as outlined above, to not amplify the forecasts for reasons linked to organizational politics, even if they agree with those forecasts. If analysts disagree with the forecasts, they can just ignore them.

To facilitate adoption and ensure the integration of crowdsourced forecasting methods into day-to-day intelligence operations, **the government could mandate that intelligence analysts engage with forecasts in their work by referencing any relevant crowdsourced forecasts and explaining their agreement or disagreement with the crowd forecast.** For this recommendation to succeed, it will have to be combined with a communication strategy to explain the causal logic behind the forecasts and those probabilities, making them easier to use for analysts and more persuasive for policymakers.

While there are some challenges to such a mandate, such as the subjectivity involved with respect to what constitutes a “relevant” forecast, a mandate would not be unfamiliar to the intelligence community. For example, Intelligence Community Directive 203 currently requires analysts to adhere to structured analytic techniques and standards in an effort to avoid bias and create cohesive intelligence products.<sup>74</sup> Similarly, such requirements to use forecasts would help integrate these products into the work of intelligence agencies and inform intelligence products.

<sup>74</sup> Office of the Director of National Intelligence, *Intelligence Community Directive 203: Analytic Standards*, Intelligence Community Directive [ICD-203], 2007/2015, <https://fas.org/irp/dni/icd/icd-203.pdf>.



These requirements could be combined with training in cognitive de-biasing and probability assessments, helping ensure that analysts have used relevant tools when making forecasts. Additionally, the training could help assist in ensuring that the best practices from forecasting influence the way that analysts generate other products as well.<sup>75</sup>

Crowdsourced forecasting methods are best employed on clear, falsifiable, narrow questions. A linchpin in the success of revitalizing crowdsourced forecasting efforts in the intelligence community, therefore, is ensuring questions both fit the above criteria and are also decision-relevant to analysts. It is important that question generation is customer-driven and responsive to particular policies and issues under consideration. Including analysts in this process, and determining which type of questions are most useful to them, such as if they are more inclined to conditional forecasting, would help with buy-in. This would further ease some of the bureaucratic challenges that might arise in response to such a mandate. Using these methods to complement traditional intelligence analysis would help diversify the methodological tools that analysts are using to generate reports, which may make them more informative and persuasive to policymakers.

The complementarity of traditional intelligence analysis with crowdsourced forecasts could be enhanced further by requiring analysts to assign probability ranges to their judgments. This would finally satisfy Kissinger's request for "betting odds" (and other calls for quantifying so-called "words of estimative probability"

such as "almost certainly") and would clarify any analyst's disagreement with a crowdsourced forecast. It may also contribute to the education and calibration of policymakers and other consumers of intelligence analyses. Intelligence consumers who are used to reading traditional reports may learn to associate familiar qualitative expressions of likelihood with probability ranges and thereby develop a more intuitive feeling for probabilistic forecasts. This bridge could ease adoption and increase understanding of forecasting aggregation methods in the government.

Additionally, analysts could still make expert judgments that disagree with the forecasts. The engagement with the forecast itself, even if only to explain why the analyst thinks the crowdsourced forecast is incorrect, would increase the transparency of intelligence analysis for end users, which could grow policymaker confidence in the products.

Finally, crowds make mistakes, and this option would allow experts to explain why they thought the crowd was over- or under-confident or otherwise misguided or biased in its predictions. This will help analysts sharpen their arguments, while also building data, over time, on what kinds of methods are more likely to be accurate on what kinds of questions.

**Recommendation: Require use of explicit probabilities from a forecasting platform in relevant products, including analyst engagement when they disagree with the forecast.**

---

<sup>75</sup> One could also create other tools also with this approach. For example, better search capacity to find relevant forecasting questions.



## PREDICTION MARKETS VS. FORECASTING AGGREGATION PLATFORMS

One question for any new, crowdsourced USG geopolitical forecasting effort is whether the platform should be set up as a prediction market, like the ICPM, or as a prediction poll, like COSMIC BAZAAR in the United Kingdom. To help visualize the difference between these types of platforms, it is useful to contrast Good Judgment Open's<sup>76</sup> forecasting aggregation platform with PredictIt's prediction market interface.<sup>77</sup> Conventional wisdom had long been that prediction markets may be more accurate because they force people to back their forecasts with something they feel strongly about—real or play money. However, recent research demonstrates that prediction polls, especially when combined with algorithmic aggregation methods, and well-calibrated forecasting teams (i.e., superforecasters),<sup>78</sup> can outperform prediction markets.<sup>79</sup>

Prediction markets are especially challenging to implement in classified environments because classified markets will necessarily have large limitations on participation, requiring the use of algorithmic correctives to solve liquidity problems. Good liquidity, like that of a well-functioning stock market, is difficult to achieve in prediction markets like the ICPM, requiring prediction markets to have corrective tools like setting liquidity parameters and using automated market makers, which attempt to simulate efficient market behavior in electronic prediction markets.<sup>80</sup>

The theorized upsides of prediction markets, moreover, come with trade-offs. In theory, individual bettors on any forecasting platform with sufficiently large incentives could engineer situations in which they directly profit, by attempting to influence outcomes of events. In prediction markets, payoffs are potentially higher for events that the market deems less likely, so individual players have increased incentives to manipulate real outcomes away from the forecasting crowd's consensus. In the case of the Policy Analysis Market in 2003, the perception of government markets incentivizing betting on surprising events, such as acts

of terrorism or assassinations, is what ultimately led to its cancellation.

Additionally, time and effort spent thinking about these strategic decisions vis-à-vis other players in the market amounts to time and effort that busy government employees could fruitfully spend elsewhere—researching their forecasts or going about their day-to-day jobs—making prediction market participation relatively more burdensome. Finally, there is no incentive for teaming or sharing information in a prediction market, because it is a first-mover market where “profiting” requires someone else to make a bad bet.

Prediction markets also complicate scoring for individual accuracy. Scores of players in prediction markets reflect trading skill as much as forecasting skill because success depends not just on (early) accuracy but on tactical trading. Prediction markets therefore also make it difficult to identify top performers or “superforecasters.” Because forecasting aggregation platforms do not involve this kind of market behavior, they are better suited for an organization that cares about understanding and tracking individual or group performance.

Finally, since prediction markets use money (real or “play”), varying time preferences on the value of money can introduce distortions, especially on longer-term questions. Research on prediction markets illustrates that “prediction markets are reasonably well calibrated when time to expiration is relatively short” but not on longer timescales—after all, most people value money now more than money later.<sup>81</sup> Such time discounting issues can introduce biases and further elements of strategic playing in prediction markets.<sup>82</sup>

Since in methods of aggregation, crowdsourced forecasts are just as accurate as prediction markets, future USG efforts should use prediction polls to address many of the revealed difficulties of using prediction markets.

**Recommendation: Use prediction polls, aggregating them with algorithms and other proven methods, rather than prediction markets.**

76 Good Judgment, “Welcome to Good Judgment Open,” <https://www.gjopen.com/>.

77 PredictIt, “Which Party Will Win the Electoral College?” <https://www.predictit.org/>.

78 Mellers et al., “Identifying and Cultivating Superforecasters as a Method of Improving Probabilistic Predictions.”

79 Pavel Atanasov, Phillip Rescober, Eric Stone, et al., “Distilling the Wisdom of Crowds: Prediction Markets vs. Prediction Polls,” *Management Science*, 63(3), April 22, 2016, <https://pubsonline.informs.org/doi/pdf/10.1287/mnsc.2015.2374>; Jason Dana, Pavel Atanasov, Philip Tetlock, and Barbara Mellers, “Are Markets More Accurate Than Polls? The Surprising Informational Value of ‘Just Asking,’” *Judgment and Decision Making*, 14(2), March 2019, <http://journal.sjdm.org/18/18919/jdm18919.html>.

80 Ben Golden, “What Is Liquidity and How Does it Affect Prediction Markets?,” *Cultivate Labs*, <https://www.cultivatelabs.com/posts/what-is-liquidity-and-how-does-it-affect-prediction-markets>; Abraham Othman and Tumoas Sandholm, “Automated Market Makers that Enable New Settings: Extending Constant-Utility Cost Functions,” In: P. Coles, S. Das, S. Lahaie, and B. Szymanski (eds.), *Auctions, Market Mechanisms, and Their Applications: Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering*, vol. 80 (Berlin, Heidelberg: Springer, AMMA 2011), <https://www.cs.cmu.edu/~sandholm/automatedMarketMakersThatEnableNewSettings.AMMA-11.pdf>.

81 Lionel Page and Robert T. Clemen, “Do Prediction Markets Produce Well-Calibrated Probability Forecasts?” *The Economic Journal*, 123(568): 491-513, May 2013, [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=2257139](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2257139).

82 As referenced above, COSMIC BAZAAR in the United Kingdom uses a prediction poll structure, rather than a prediction market.



## FORECASTING PLATFORM CLASSIFICATION LEVEL

Both classified and unclassified platforms have advantages and disadvantages, and a new forecasting effort could launch both types, potentially as an integrated system. Existing USG crowdsourced forecasting efforts generally operate on classified markets, outside of unclassified S&T initiatives. In contrast, recent UK efforts provide a template for a more accessible government forecasting effort. There are benefits to classified forecasting platforms. Classified platforms allow for the asking of questions where even asking them might reveal important information about U.S. interests. Classified platforms also mean participants can use classified information and knowledge to forecast, which would be hard to compartmentalize for use on an unclassified platform. Moreover, by adopting the rest of the recommendations in this white paper, it will be more plausible to generate a significant pool of classified forecasters, rather than just a small group of experts.

Yet, an unclassified platform also has significant advantages. An unclassified forecasting platform can help optimize wisdom of crowd effects by increasing participation and diversifying the pool of forecasters. A classified system like the ICPM restricts the number of people who may participate in the forecasting challenges to those government employees or contractors who hold certain clearances and have regular access to top secret computer networks. An unclassified system could expand the pool of participants throughout the government. Even if the goal is simply the identification of superforecasters, the wider the net, the larger the number of superforecasters that a forecasting initiative will reveal.

Enlarging the pool of forecasters through an unclassified network will also likely diversify that pool. A more diverse forecasting pool will improve accuracy because the wisdom of crowds works best when people with different information in aggregate are forecasting.<sup>83</sup> Moreover, there may be systemic biases within the intelligence community that an unclassified forecasting platform could help even out. In particular, an unclassified system could generate substantial participation from State Department foreign service officers and others deployed around the world, increasing the scope of participation in ways that can enhance accuracy.

Running an unclassified prediction market may also help alleviate some of the problems of the secrecy heuristic—the tendency to give more credence to classified information, even if there are not principled reasons to privilege that information. As described above, there is already evidence from the Good Judgment Project in the ACE program that forecasters with access to open-source unclassified information can outperform intelligence analysts on the same questions, in some cases.

This secrecy bias may be pervasive in the intelligence community and may have contributed to the relative success of open-source superforecasters in the ACE program over ICPM forecasters with access to classified information. Experimental research on using secrecy as a heuristic for informational quality demonstrates that people tend to weigh secret information more heavily than publicly available information, viewing secret information as higher quality than public information.<sup>84</sup> Secrecy does matter, especially in situations where information asymmetry exists, but a pervasive secrecy bias may negatively affect the accuracy of a classified crowd in some cases.

An unclassified geopolitical forecasting platform, by creating a more accessible system, could also host questions relevant for many different agencies, increasing overall bureaucratic support for geopolitical forecasting initiatives. The easier ability to use the system would make it more useful for a variety of agencies. Additionally, the forecasts themselves could be shared across government, informing and improving everyone's work without fear of accidentally disclosing secret information (e.g., in forecast justifications written by analysts). Unclassified forecasts could also be uploaded onto more secure or classified networks of the government.

**Recommendation: Create both classified and unclassified forecasting platforms. Launch an unclassified platform first if necessary.**

83 James Surowiecki, *The Wisdom of Crowds: Why the Many Are Smarter Than the Few and How Collective Wisdom Shapes Business, Economies, Societies and Nations* (New York: Doubleday, 2004).

84 Mark Travers, Leaf Van Boven, and Charles Judd, "The Secrecy Heuristic: Inferring Quality from Secrecy in Foreign Policy Contexts," *Political Psychology*, 35(1), 2014, <https://online.library.wiley.com.proxy.library.upenn.edu/doi/pdfdirect/10.1111/pops.12042>.

## TRACKING INDIVIDUAL PERFORMANCE AND LEADERBOARDS

Another question of platform design centers on performance tracking, gamified rankings, and the use of leaderboards. An advantage of any falsifiable forecasting system is the ability to clearly track the accuracy of the system in a transparent fashion. Crowdsourced forecasting systems, therefore, also provide opportunities for accountability. Officials could track, for example, the accuracy of analysts who focus on China when it comes to forecasting questions involving China. Accuracy on those forecasts can also provide feedback in a way that generates opportunities for continual job training and improvement by adopting de-biasing strategies.

However, a key element of most crowdsourced forecasting systems is a public (to participants) leaderboard that transparently displays all participants ranked according to whatever scoring system the platform uses (e.g., a Brier score). Although leaderboards theoretically help encourage participation, especially among top performers, they also raise questions that can create a disincentive to broad-based participation, including the large number of “losers” created by a ranking system. How can a revised approach to crowdsourced forecasting in the USG balance the advantage of accountability and job training with the risks of participation disincentivizes?

### *Use of Leaderboards and Performance Tracking in Forecasting*

Crowdsourcing initiatives like prediction markets and prediction polls often encourage intrinsic motivation to participate through methods known as “gamification,” including the use of leaderboards.<sup>85</sup> Leaderboards, for example, by providing positive feedback to successful forecasters, increase their incentives for continuing participation.<sup>86</sup> Such leaderboards also serve a simple purpose of identifying top performers. Leaderboards do

not, however, increase participation among the majority of forecasters.<sup>87</sup> For most forecasters, evidence that they are not among the leaders may actually serve as a disincentive for broader participation, as most people feel like they are “losing.”

Leaderboards within a government bureaucracy may also stimulate further resistance to institutionalizing forecasting platforms,<sup>88</sup> due to concerns that less experienced employees or “amateurs” may outperform their superiors—the status quo incentives mentioned above.<sup>89</sup> Leaderboards are therefore useful but present a series of challenges themselves for USG forecasting efforts.

Finally, the average difficulty of questions may vary between departments, so comparing Brier scores across teams might be legitimately “unfair.” This could be partly mitigated by focusing on measuring performance over a cheap baseline, meaning a macro average of overall forecasting accuracy (and comparing how much different teams outperform that cheap baseline, with the cheap baseline varying between teams in response to differing average question difficulty), but recognizing these issues will be important.

### *Improvements for Future Leaderboard Use*

Several adjustments could preserve the useful incentives of gamified leaderboard interfaces while mitigating the disincentives mentioned above. These include:

- ranking performance only by “teams” or within “challenges;”
- resetting periodically;
- rewarding participation as well as performance; and
- introducing individual performance sheets.

First, there need not be individuals ranked on a single leaderboard for the entire platform. Leaderboards could also rank teams, rather than individuals, that group departments as a whole.<sup>90</sup> A team mentality, based on

85 B. Morschheuser, J. Hamari, J. Koivisto, and A. Maedche, “Gamified Crowdsourcing: Conceptualization, Literature Review, and Future Agenda,” *International Journal of Human-Computer Studies*, 106: 26-43, 2017, [https://trepo.tuni.fi/bitstream/handle/10024/118201/gamified\\_crowdsourcing\\_conceptualization\\_2017.pdf?sequence=2&isAllowed=y](https://trepo.tuni.fi/bitstream/handle/10024/118201/gamified_crowdsourcing_conceptualization_2017.pdf?sequence=2&isAllowed=y).

86 Thomas A. Montgomery, Paul M. Stieg, Michael J. Cavaretta, and Paul E. Moraal, “Experience from Hosting a Corporate Prediction Market: Benefits beyond the Forecasts,” *KDD '13: Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, August 2013, [https://dl.acm.org/doi/pdf/10.1145/2487575.2488212?casa\\_token=ydFSI8pAQcQAAAAA%3Av-gR2gyQ87Xm31eMz1-xtCDd980OBJwafOEyC\\_mrdqiU0q-Ji0krCccF8G2lHh6E7B6S4Lgbapkuvq](https://dl.acm.org/doi/pdf/10.1145/2487575.2488212?casa_token=ydFSI8pAQcQAAAAA%3Av-gR2gyQ87Xm31eMz1-xtCDd980OBJwafOEyC_mrdqiU0q-Ji0krCccF8G2lHh6E7B6S4Lgbapkuvq).

87 Panagiotis G. Ipeirotis and Evgeniy Gabrilovich, “Quiz: Targeted Crowdsourcing with a Billion (Potential) Users,” *WWW '14: Proceedings of the 23rd International Conference on World Wide Web*, April 2014, [https://dl.acm.org/doi/pdf/10.1145/2566486.2567988?casa\\_token=RzbV-KrWloAAAAARpb9V4A2alawCv3s0S4nXZWRZ53rYuCW4BbAOwHelAioupM5eM0gsipgKWdN0qWNi7NxYkiWMSRADQ](https://dl.acm.org/doi/pdf/10.1145/2566486.2567988?casa_token=RzbV-KrWloAAAAARpb9V4A2alawCv3s0S4nXZWRZ53rYuCW4BbAOwHelAioupM5eM0gsipgKWdN0qWNi7NxYkiWMSRADQ).

88 Bo Cowgill and Eric Zitzewitz, “Corporate Prediction Markets: Evidence from Google, Ford, and Firm X,” *The Review of Economic Studies*, 82(4), 2015, <http://www.restud.com/wp-content/uploads/2015/03/MS14671manuscript.pdf>.

89 Daren C. Brabham, “The Myth of Amateur Crowds: A Critical Discourse Analysis of Crowdsourcing Coverage,” *Information, Communication & Society*, 15(3), 2012, [https://www.tandfonline.com/doi/pdf/10.1080/1369118X.2011.641991?casa\\_token=og9R1yx8yMoAAAAA%3AR55-sG\\_QZAzk2UMhSM3A52Jtn8wGrGF\\_Xr6RLgtTNz3M0tSGwP-nKrY-IOvrt135-qzWY3lwWwz1L1g&](https://www.tandfonline.com/doi/pdf/10.1080/1369118X.2011.641991?casa_token=og9R1yx8yMoAAAAA%3AR55-sG_QZAzk2UMhSM3A52Jtn8wGrGF_Xr6RLgtTNz3M0tSGwP-nKrY-IOvrt135-qzWY3lwWwz1L1g&).

90 Markus Rokicki, Sergej Zerr, and Stefan Siersdorfer, “Groupsourcing: Team Competition Designs for Crowdsourcing,” *WWW '15: Proceedings of the 24th International Conference on World Wide Web*, May 2015, [https://dl.acm.org/doi/pdf/10.1145/2736277.2741097?casa\\_token=1SXHkpPzAAQwAAAAA%3AAtHOYORsm8mNvYmgzMGU.S.f7slecWAQRAFUNo5FhdCV9uFz9LLtmQFZUP2gx:3x2K6S1H6\\_nudpQUVg](https://dl.acm.org/doi/pdf/10.1145/2736277.2741097?casa_token=1SXHkpPzAAQwAAAAA%3AAtHOYORsm8mNvYmgzMGU.S.f7slecWAQRAFUNo5FhdCV9uFz9LLtmQFZUP2gx:3x2K6S1H6_nudpQUVg).

offices or departments, so team size would vary, could lead to a broader level of participation as it would not directly distinguish between individuals, allowing for collective success. Additionally, team leaderboards would not distinguish between supervisors and their departments, potentially reducing some of the bureaucratic resistance to implementing crowdsourcing. Relatedly, organizations could structure leaderboards around specific challenges (e.g., “The Future of U.S.-China Competition,” “The Trajectory of Globalization,” “New Developments in African Politics”). Each new challenge could have its own leaderboard, refreshing the organization’s competitive spirit frequently, and reducing the feeling of being “stuck” at the bottom of a long-lived leaderboard. Good Judgment Open uses the challenge leaderboard model to address these issues.<sup>91</sup>

Rather than focusing on the accuracy of participants’ predictions, the boards could also rank participation levels, meaning that those who participated more frequently would be ranked higher, regardless of the accuracy of their predictions.

Finally, in lieu of or in addition to leaderboards, participants could also receive private profiles that track the accuracy of their predictions, their participation levels, and their improvement in accuracy over time. This could reduce the impact of leaderboards on bureaucratic tensions. They could also reduce concerns about disincentives, as the private dashboards would remove the social pressures that could prevent participation by low-scorers. The emphasis on measuring participation and improvement would attach an intrinsic value to forecasting independent of accuracy scores.

In short, a menu of options to modify the use of leaderboards in future USG forecasting platforms could preserve the incentives gamification creates while mitigating participation disincentives.

**Recommendation: Implement a flexible approach to leaderboards, tracking data on success over time for every forecaster, but creating options to frame it based on self-improvement, in-office “pods,” and other ways that create continuing incentives to participate, rather than undermine motivation.**

## INTEGRATION WITH OTHER FORECASTING APPROACHES

The USG currently uses several different methods for geopolitical forecasting, even if those methods are often used implicitly for forecasting purposes.

Crowd forecasts are not only complementary with other methods, but can also amplify the successful elements of other analytical tools. For example, a crowd forecasting platform could include other intelligence analysis methods in inputs available to forecasters, meaning its aggregative effects could help identify under-used methods and show which of these methods are best suited to certain problems. Just as these mechanisms can help to compare sources (see above), therefore, they can also help to compare and assess methods and improve the overall quality of analysis.

While crowdsourcing offers many advantages for forecasting geopolitical events, these methods have clear limits. Some questions about the world are more easily converted into quantifiable propositions. Whether North Korea will test a nuclear weapon by a given date, for example, is easier to convert into a quantifiable proposition than understanding the conditions in which Kim Jong Un will open his economy more to the West.<sup>92</sup>

Some examples of other forecasting methods include:

- scenario-planning exercises;
- war games;
- unstructured qualitative analysis; and
- decision-science approaches based on machine-learning models or out-of-sample forecasts from time-series models.

Each of these approaches has its own advantages and disadvantages. Scenario-planning exercises and war games have utility for revealing unanticipated interactions between actors in a crisis, the potential behavioral reactions to different approaches, and the causal logic that could govern how decision makers will approach a situation. These methods also have value beyond pure forecasting in encouraging broader thinking and alternative analyses. Such complementary methods may even include “futurethink” techniques, such as science-fiction narratives, whose “possibilistic” frames can help round out the shortcomings of probabilistic approaches if deployed appropriately.<sup>93</sup>

91 Good Judgment, Inc., “Challenges,” Leaderboards, Good Judgment, Inc., <https://www.gjopen.com/challenges>.

92 As explained above, it is possible, of course, to convert more complicated issues into discrete, falsifiable questions through the generation of clusters of questions, each of which sheds light on a broader puzzle. However, illustrating the relative value of such an approach for policymakers may be more complicated, especially in the short to medium term.

93 Alexander H. Montgomery and Amy J. Nelson, *The Rise of the Futurists: The Perils of Predicting with Futurethink* (Washington, DC: Brookings Institution, 2020), <https://www.brookings.edu/research/the-rise-of-the-futurists-the-perils-of-predicting-with-futurethink/>.

Less structured qualitative analysis can draw on years or sometimes decades of experience from analysts who use that knowledge to assess available data and generate a professional judgment on what is most likely to occur. Forecasting from decision science, whether using machine-learning models or simply out-of-sample forecasts from time-series models, can aggregate huge quantities of data and find patterns that are hard for humans to identify.

Each of these methods complements prediction polls in different ways. Prediction polls could feed topics for inquiry into scenario-planning exercises and war games, for example. As referenced above, prediction polls could also serve as triggers for qualitative analysis, whereby, as the probability of a particular event reaches a given threshold, it would create a requirement for the generation of a report based on existing evidence—a deeper dive. Prediction polls are more flexible than decision-science approaches, because they do not require large quantities of data to train models, but for areas where those models have enough data to make forecasts, they are powerful.

Given that the USG could certainly serve to improve at geopolitical forecasting, adding prediction polls to the forecasting tool kit could help bring a greater diversity of methods to the table. Focusing on the complementarities with existing approaches, rather than competition, may not only create more effective overall forecasting, but it could also improve the probability of implementation success within complex bureaucracies.

**Recommendation: Promote crowdsourced geopolitical forecasting as complementary to existing forecasting methods.**

## RESEARCH AND DEVELOPMENT VS. APPLICATION

A key challenge for any new forecasting initiative is becoming a program of record. However, there are continuing areas for research on quantitative geopolitical forecasting that should continue in tandem with implementation of a USG crowdsourced forecasting platform. Ongoing research and development could ensure that scientific advances are continually integrated in ways that increase the utility of the approach.

There are many possibilities for research and development efforts that could complement the relaunch of a geopolitical forecasting platform using one of the organizational models suggested below. These efforts

could focus on areas that build on the successes of past projects, like the ACE program, in narrowly tailored ways. Opportunities for further research could include:

- A:B testing on how different approaches to leaderboards influence accuracy and participation over time;
- A:B testing on methods of communicating quantitative forecasts to policymakers, especially attempts to integrate crowdsourced forecasts with explanations of the causal logic of those forecasts;
- A:B testing on different user interfaces for eliciting forecasts, especially continuous-variable forecast interfaces;
- designing “noise-reduction” strategies that help analysts filter the huge mass of complicated data they have to sort through when making judgments;<sup>94</sup>
- leveraging advances in decision science using artificial intelligence to revive attempts to combine human and machine forecasting models into a more synthetic approach;
- launching experiments focused on different types of forecasts on which there is currently little research, including conditional forecasting and longer-term forecasts;
- A:B testing on clusters of questions designed to provide insights on broader national security challenges; and
- research on the development of methods to accurately explain and communicate why the future events have the probabilities that the forecasting platform assigns them.

A permanent open forecasting tournament could enhance research efforts. Such a tournament could be structured like the Geopolitical Forecasting Challenge or Mercury Challenge, but without a fixed end date. The open tournament could serve both as an incubator of innovative forecasting methods and as a proving ground for any team or contractor to demonstrate the effectiveness of their forecasting approach.

**Recommendation: Fund research and development designed to improve the accuracy of crowdsourced geopolitical forecasting methods, and efforts to effectively communicate quantitative insights to policymakers. Run an open forecasting tournament as a research incubator and proving ground.**

<sup>94</sup> Vile Satopaa, Marat Salikhov, Philip Tetlock, and Barb Mellers, “Bias, Information, Noise: The BIN Model of Forecasting,” *Social Science Research Network*, April 7, 2020, [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=3540864](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3540864).

## SUPERFORECASTERS

A key output from Good Judgment's success in the ACE program was the ability of the Good Judgment Project to take top forecasters from the early years of the program, give them additional training, and place them in teams designed to leverage their combined expertise and ability. The striking results, which outperformed other methods, illustrate the power of identifying and cultivating the absolute best forecasters in an organization.<sup>95</sup>

One question for any forecasting platform is whether the goal is to encourage substantial participation over time from a broad pool or to identify and cultivate superforecasters and then focus on them as the “crowd” for future forecasting. While having superforecasters is not inconsistent with broad participation, the prioritization of approaches could influence a number of design choices. For example, concerns about leaderboards and their impact on motivation would be less relevant if the point of a forecasting platform is the identification of superforecasters, rather than broad participation over time.

The most successful implementation of crowdsourced forecasting may involve attempting to maximize both of these factors. After all, research results show that while superforecasters are superior to aggregated forecasts, combined with algorithms, they are not that much better.<sup>96</sup> And it is not clear if they are superior enough to influence policy. Moreover, the job training and accountability aspects of participating in falsifiable forecasting may be reasons to prefer a broad pool over time. However, it would be very attractive for the USG to identify a group of superforecasters who are already within the government community. They could then potentially have some of their responsibilities reassigned to focus more on forecasting. An alternative might be to contract with external superforecasters, either building a large enough team of contractors with relevant security clearances or having them forecast on unclassified questions.

**Recommendation: Prior to relaunching a crowdsourced forecasting platform, ensure that the identification of top forecasters as a programmatic goal does not detract from broad and consistent participation.**

---

<sup>95</sup> Good Judgment, Inc. forecasters also proved very successful in other IARPA programs.

<sup>96</sup> Mellers et al., “Identifying and Cultivating Superforecasters as a Method of Improving Probabilistic Predictions.”



# ➤ PATHS FORWARD FOR USG GEOPOLITICAL FORECASTING

**A renewed attempt at encouraging the adoption of crowdsourced geopolitical forecasting methods within the United States could occur in several different ways, each of which has its own upsides and downsides. This section outlines several different possible locations for such an endeavor, describing both the reasons to consider that location and potential downsides.**

## **THE INTELLIGENCE COMMUNITY**

The most natural home for a future geopolitical forecasting effort in the USG lies with the agency responsible for geopolitical forecasting in the first place—the intelligence community. The familiarity of intelligence analysts with forecasting and the initial founding of the ICPM could ease the process of adoption of a new platform.

A natural home for a geopolitical forecasting initiative would be a National Forecasting Office placed in the Office of the Director of National Intelligence. It could incorporate crowdsourced forecasting efforts, quantitative forecasting methods, including decision-science approaches that incorporate artificial intelligence and more qualitative approaches.

Moreover, if the next chair of the National Intelligence Council is favorably inclined toward geopolitical forecasting, it could be a potential location for a forecasting platform. While the challenges involved in the National Intelligence Council's administration of the ICPM would raise questions about the success of a future platform run by the council, as a futures-focused area of the intelligence community, it is a logical location.

Additionally, given that effective adoption will likely include intelligence analysts incorporating crowdsourced forecasts into their work products, basing a platform in the intelligence community itself may make adoption by analysts more likely. As a tool within the intelligence community, analysts may have more trust than if the platform is deployed elsewhere.



Basing a new forecasting platform within the intelligence community would also have downsides, however. Issues surrounding the use of the ICPM described above suggests at least some resistance to a method of intelligence analysis that is unfamiliar to most intelligence analysts.<sup>97</sup>

## DEFENSE DEPARTMENT

A second option would be to place a geopolitical forecasting platform in the DoD. Given the resources of the Pentagon and the inherent interest of the defense community in forecasting, the DoD is a plausible political location for a crowdsourced geopolitical forecasting platform. Moreover, the pursuit and adoption of emerging technologies will be critical to the DoD in the decades ahead, especially artificial intelligence. Decision-science approaches to leverage knowledge from algorithms as decision aids for commanders, to ease the cognitive load and help them make better judgments, are already growing in prominence in the research and engineering communities. Forecasting aggregation methods that use algorithms to aggregate prediction polls could easily tie into this growing interest in decision science. Connecting the broader DoD imperative to invest more in artificial intelligence to the question of geopolitical forecasting could provide an opportunity for growth.

There are also preexisting directorates of the military services focused on early warning—the “2’s.” Their responsibilities include forecasting. This could make the military services a potential locus for support, or even a home for a forecasting platform. Additionally, the Defense Intelligence Agency could be a potential bridge between DoD forecasting efforts and the intelligence

community. Forecasting efforts could fit within the Defense Intelligence Agency’s work on developing new analytic tools to help analysts better grasp probabilities in their subject matter areas.

The downside of placing a geopolitical forecasting platform in the DoD is the vastness of the defense bureaucracy. Though DoD’s size and scope could make it easier to establish a forecasting platform in the defense space, it could also make it challenging for the results of a forecasting platform to get visibility with decision makers. Moreover, if a key consumer of these forecasts will be intelligence analysts using them for reports, they may resent DoD being a key source of data, as opposed to the data coming from within the intelligence community. That could make adoption more challenging and lead to bureaucratic infighting. Overcoming that obstacle would be important.

## THE WHITE HOUSE

A third option for promoting geopolitical forecasting in the USG would involve putting a future platform even closer to the president, in the National Security Council (NSC), by standing up an NSC staff office focused on forecasting, early warning, and/or crisis prevention.<sup>98</sup> The advantage of placing a forecasting office within the broader White House architecture is that proximity to the president may make it more likely that forecasts have impact and influence U.S. national security decision-making. Placing a forecasting office within the NSC staff would also prevent it from being buried at an agency.

Of course, a downside to placement in the NSC is that the same thing that could make such a move impactful—proximity to the president—could also create challenges.

---

<sup>97</sup> Note that this also illustrates that education and training may be necessary to ensure successful adoption.

<sup>98</sup> Dafna H. Rand and Michael C. Horowitz, “The Crisis-Prevention Directorate,” *The National Interest*, May 21, 2014, <https://nationalinterest.org/feature/the-crisis-prevention-directorate-10500?page=0%2C1>.



Perceptions that a forecasting tool had led policymakers astray might lead to efforts to close such an office or to remove the national security advisor or even the president. A change of political parties in particular could place the effort at risk. The driver for change might not be a substantive disagreement over the utility of crowdsourced forecasting methods but instead the way that each new president, and sometimes even each new national security advisor, puts their stamp on the office and on U.S. national security decision-making, including by distancing themselves from the policies and procedures of their predecessor.

A related placement for a forecasting platform might be the Office of Science and Technology Policy (OSTP). The OSTP leads S&T policy coordination across agencies. To the extent that a new geopolitical forecasting platform might aggregate forecasts from people across the USG, a platform based in OSTP, like a platform based in the NSC, might have an easier time at encouraging participation across a multitude of executive agencies and offices. While important for U.S. S&T leadership, OSTP has a somewhat lower public profile than the NSC staff. This could help a geopolitical forecasting initiative survive, politically, over time. Basing a forecasting platform at OSTP might also encourage forecasters to ask new types of questions, beyond geopolitics, to track the likely success of government initiatives, perceptions of internal programs, and other efforts to improve the efficiency and efficacy of government. Going beyond geopolitical questions could also help illustrate the utility of forecasting aggregation methods, which could help generate support for geopolitical forecasting as well.

However, though OSTP's lower profile could increase the chances of a forecasting platform's survival, it might also make it less likely that analysts and decision makers use forecasts effectively. If a geopolitical forecasting platform were based at OSTP, it would be critical to have a clear mechanism, identified from the start, for the communication of those forecasts throughout the executive branch.

An additional option close to the president would be to open a "National Forecasting Office" in the White House. Such an office could help to launch a whole-of-government forecasting platform and serve as a one stop shop for coordinating forecasting efforts across different agencies and departments. Such an office would also help to address the core competency issues raised earlier, as its employees would view forecasting as their core mission. Successful models for a National Forecasting Office may be found in the Office of Information and Regulatory Affairs' core competency in cost-benefit analysis or in the United Kingdom's Behavioural Insights Team, housed close to the president and prime minister, respectively.

## CONGRESS

Another organizational option for a geopolitical forecasting platform would be in Congress. If the 117th Congress, which convened in January 2021, restarts a version of the Office of Technology Assessment, which was shuttered in the 1990s, it could be an opportunity to create a forecasting platform as part of that endeavor. Another option would be to add running a geopolitical forecasting platform to the responsibilities of the Congressional Research Service—known for its detail-oriented, fact-based approach to research on many issues, including national security and foreign policy issues—or to those of the GAO's new Science, Technology Assessment, and Analytics office, a potential successor to the Office of Technology Assessment.

Placing a geopolitical forecasting platform under congressional supervision could help ensure it has bipartisan support, which could make it more likely to survive the inevitable political ups and downs of any forecasting effort. Moreover, the nonpartisan mandate of congressional agencies could, in theory, help insulate a forecasting platform from accusations of political bias.

However, the level of political polarization in the United States could make it difficult to authorize a congressionally operated platform in the first place. The platform could also become a political football in itself, with members attempting to push questions with biased wording designed to elicit particular answers that would support their world view. In theory, that forecasting questions eventually resolve, providing an empirical judgment on the world, would help guard against accusations of bias. Even if a member launched a biased question, eventually it would resolve in a way that would likely reveal their bias and undermine their position. In that case, though, the likelihood of this might make members potentially less likely to authorize the creation of a forecasting platform in the first place.

## OTHER OPTIONS

Other options for placing a crowdsourced geopolitical forecasting platform within the USG would be in the Office of International Affairs in the Treasury Department, in the U.S. Digital Service, or in the State Department. Given the way that numbers and data analysis play a crucial role in economic assessments already, the bureaucracy within the Treasury Department already designs and promotes methods for forecasting economic futures. Moreover, the Office of International Affairs is responsible, specifically, for issues surrounding geopolitical risk. While that is generally defined mostly in the context of economics, given the intersection of economic and security issues, it would be

plausible for the Treasury Department to expand regular forecasting efforts to include geopolitical risk.

The potential downside to placing a program in the Treasury Department would be the novelty of the content—geopolitical issues—relative to most of what the Treasury Department does. The majority of forecasters on a platform would likely be coming from outside of the Treasury Department, and most of the end users would also be outside of the Treasury Department. This could create a bureaucratic incentives challenge. One solution to the challenge might be to ensure that there are sufficient questions about economic issues to give the Treasury Department a clear stake in the platform.

The U.S. Digital Service's competency with government websites and networks, and its location in the Executive Office of the President, make it another promising home for a forecasting platform. Its focus on optimizing performance through design and technology could be useful in creating a user-friendly experience on the forecasting platform, a key element for ensuring adoption across government. The Digital Service's focus on public-facing platforms, however, may not make it the best option for an internal forecasting platform.

A key element of any successful implementation of a USG forecasting platform will be garnering users in the State Department, given the deployment of State Department personnel with on-the-ground insights around the world. The State Department is a plausible location for a forecasting platform given that the substantive content of geopolitical forecasting questions has clear relevance for the State Department, specifically in Policy Planning or the Bureau of Intelligence and Research. Placing a forecasting platform in the State Department might also make adoption by State Department employees more likely, which could improve forecasting accuracy. Foreign services officers around the world develop expertise on the micro-details of specific countries. This could give them unique insight to complement forecasts from other parts of the government. With limited budgets and a more qualitatively oriented organizational culture, implementation in the State Department could prove tricky, however. It might be especially vulnerable to turnover in political appointees, unless a platform could be quickly integrated in a way that ensures the buy-in of career civil servants.

## > CONCLUSION

**Forty-seven years after Henry Kissinger asked for the equivalent of gambling odds on important geopolitical events, moving forward with crowdsourced geopolitical forecasting efforts could provide decision makers with new and helpful information. In combination with quantitative forecasts derived from statistical models, and traditional sources of information and forecasts, crowdsourced forecasting could help busy decision makers more quickly make sense of a complicated world.**

Despite backsliding in the funding of crowdsourced forecasting methods over the last several years, there is an opportunity, as the Biden administration enters office, to move forward. Adoption of crowdsourced geopolitical forecasting by the USG will require addressing three sets of questions. The first set of questions involves the organizational location of any platform. Basing a platform in the intelligence community, Defense Department, Congress, or elsewhere could shape the available resources, the types of questions asked, and the availability of the forecasts. The second set of questions involves the details of the platform, wherever it is organizationally located, like whether to use a prediction poll or prediction market, the design of leaderboards, strategies to encourage participation, further research and development efforts, and more. The third set of questions involves how to design an effective system to communicate the results from any forecasting platform to end users. Without an effective communications strategy, adoption will not succeed.

All of these decisions, like forecasting itself, will need to be made under conditions of uncertainty. But the events of the last year have made clear the vital necessity of reinvigorating crowdsourced geopolitical forecasting by the USG—not as a replacement for traditional forecasting methods, but as a complement that can provide additional useful information to busy decision makers.

Effective implementation of the general ideas in this white paper will require reaching out to and influencing several groups of stakeholders whose buy-in will be critical to the adoption of crowdsourced or quantitative geopolitical forecasting by the USG as a complement to more traditional forecasting methods.

- **Current intelligence officials:** These officials have the most experience in geopolitical forecasting and helmed the last efforts to create and implement crowdsourced geopolitical forecasting methods in the USG. Their support will be important for the development and promotion of any forecasting platform, though its relative importance would be lower if a platform is developed and implemented in another agency.
- **Political appointees:** Launching a forecasting platform with the scale and scope to succeed over time, and the bureaucratic support for implementation, will also require high-level support from political appointees in the relevant agency chosen for the creation of the platform.
- **National security community:** Over the next year, a number of officials will enter the incoming Biden administration. Shaping the perceptions of the Washington, DC, foreign policy community during 2021 will therefore be important to increasing the probability of success over the next four years.
- **Congressional staff:** A key constituency for the success of any forecasting platform will be Congress. There are several members of Congress who could have interest in a crowdsourced geopolitical forecasting platform, from those focused on national security issues to those more supportive of evidence-based approaches to policymaking. This kind of initiative has the potential to gain bipartisan support, but that will require effective outreach to the Hill.

## APPENDIX: INTERVIEWS

This white paper benefited from interviews with a number of people who provided important information and helpful context. These include:

**Jeff Alstott, U.S. Government**

**Christopher Bird, Government of the United Kingdom**

**Robert Cardillo, Former Director of the National Geospatial Intelligence Agency**

**Welton Chang, Human Rights First**

**Catherine Day, Government of the United Kingdom**

**Charles Edwards, Government of the United Kingdom**

**Jeffrey Friedman, Dartmouth College**

**Marc Koehler, Good Judgment, Inc.**

**Jason Matheny, Center for Security and Emerging Technology, Georgetown University**

**Michael Page, Center for Security and Emerging Technology, Georgetown University**

**Adam Russell, University of Maryland**

**Adam Siegel, Cultivate Labs**

**Jana Schwartz, University of Maryland**

**Brad Stastny, Microsoft Corporation**

**Philip Tetlock, University of Pennsylvania**



**UNIVERSITY OF PENNSYLVANIA | PERRY WORLD HOUSE**  
3803 LOCUST WALK, PHILADELPHIA, PA 19104

215.573.5730

@PERRYWORLDHOUSE  
FACEBOOK.COM/PERRYWORLDHOUSE

**GLOBAL.UPENN.EDU/PERRYWORLDHOUSE**