THE SENTINEL PROJECT
FOR GENOCIDE PREVENTION

Tapping into the "Smart Crowd" to Predict Genocide

Leveraging Group Intelligence for Risk Assessment

Christopher Tuckwood
The Sentinel Project for Genocide Prevention
Toronto, Canada
chris@thesentinelproject.org

Collin Sullivan
The Sentinel Project for Genocide Prevention
Berkeley, California
collin@thesentinelproject.org

Abstract

Genocide is generally accepted as being a predictable phenomenon, which also makes it preventable. However, one of the challenges in establishing an effective early warning system is bridging the remaining tension seen between qualitative and quantitative approaches to risk assessment. These two approaches are actually complementary and one possible compromise is to develop a system of "crowdsourcing" risk assessments by aggregating the collective intelligence of groups of genocide experts. This is a potentially powerful idea built on the theory that large numbers of people essentially guessing can sometimes answer complex questions or forecast the future more accurately than individual specialists and subject matter experts. Various intelligence agencies and academic projects are already studying the power of crowdsourcing for a variety of purposes. With regard to predicting genocide, the proposed solution is to assess the risk of genocide in a given country with the help of a "smart crowd" of pre-selected genocide experts whose anonymized responses would be based on a standardized set of background information and weighted according to their areas of expertise before being aggregated into an overall risk score. There are many aspects of such a system which require significant further research and testing in order to refine its accuracy once it has been established.

Introduction

The purpose of this paper is to highlight an existing problem and propose a solution while also providing some direction for future work and inviting readers to join as participants. Constructive criticism is welcomed since this research project is at an early and formative stage. The problem in question is one of the challenges inherent to the early warning of genocide - bridging the significant gap between qualitative and quantitative analysis in order to produce more effective forecasts. As several researchers have shown, genocide is the result of a common set of preconditions, or risk factors, out of which the genocidal process grows. The main point of contention between experts is what these specific factors are and then what methodology to use in assessing and measuring them.

Many in the genocide studies and prevention field advocate statistical modeling as the most effective and reliable method of determining the likelihood of genocide taking place in a given country within a specified timeframe. Others are strong proponents of taking a more

qualitative approach which is rooted in consulting subject matter experts and their knowledge of either the phenomenon of genocide itself or else the situation within the country in question. Both approaches have their own strengths and weaknesses, such as the tendency to view qualitative assessments as highly susceptible to bias and the argument that while statistical modeling significantly reduces bias and provides a structured process, its attempt to reduce complex human dynamics down to mathematical calculations contributes little to prevention efforts which require in-depth understanding of a situation.

In reality, qualitative and quantitative methods of assessments are complementary and, when combined, produce a stronger and more actionable forecast since qualitative analysis can provide context to statistical modeling while the results of those calculations can lend a degree of "scientific" rigour to qualitative assessments. One way of actually fusing these two approaches is to harness the potential of group intelligence - commonly referred to as the "wisdom of crowds" - for quantifying subjective qualitative assessments of risk with a system for aggregating a wide range of expert opinions. There are a variety of approaches to doing this, as well as multiple ways to evaluate the resulting data, which will be explored below in a relatively simple proposed initial study.

Group Intelligence for Predictions and Decisions

Simply put, the challenge here is to find a way of quantifying the unquantifiable, as Philip Tetlock has phrased it. Since individual expert opinions cannot be measured in the same way that demographic or economic information can be measured when conducting a risk assessment, combining the opinions of multiple experts can provide opportunities for measurement while also helping to eliminate the influences of individual bias. This idea is at the core of the concept of group intelligence, the proponents of which, such as James Surowiecki,[1] argue can more accurately answer questions than individual experts on a variety of subjects. The argument that large numbers of non-experts essentially guessing can answer a question more accurately than any one member of that group or even more so than individual subject matter experts has naturally raised some eyebrows but it does have merit under specific circumstances.

"The wisdom of the crowd," as it is sometimes called, can in fact be more effective when answering questions for which there are precise and known answers, such as those relating to geography or the characteristics of objects (e.g. guessing the circumference of the earth or the number of jelly beans in a jar). One commonly cited example of this is the performance of investment fund managers, the vast majority of whom have been found in several studies to be less effective than the market itself at producing financial returns when considered on an individual basis. However, averaging together the performance of large numbers of managers provides results which more closely follow the movements of the markets - essentially, the group as a whole is more effective than almost all (there are a few exceptions) of the supposedly expert individuals who comprise it.[2]

Even more relevant to the problem of predicting genocide is Tetlock's research on the accuracy of political experts when forecasting events in the international arena. Over the course of 18 years, Tetlock collected over 28,000 forecasts from 284 different experts and found that

[1] Surowiecki, James. *The Wisdom of Crowd.* London: Abacus, 2004.
[2] Ibid.

not only were many wrong in their predictions but also that the main discernible trend from the study was a negative correlation between confidence and accuracy. The more confident that individuals were in the accuracy of their predictions the less likely it was that those predictions would actually be accurate.[3] Theoretically, the right type of aggregation system would help to mitigate these flawed individual judgements, which will be one of the questions addressed by the study proposed below.

Broadly speaking, there are two common approaches to harnessing group intelligence - the first being to conduct a relatively straightforward survey and the second being the more complex option of establishing what is usually called a decision market or a prediction market. Creating a prediction market, which will be addressed first, involves the establishment of a system in which individuals trade futures contracts (such as is done for commodities in traditional financial markets) the value of which are linked to the outcomes of real-world events. For example, a particular contract may concern the likelihood of a terrorist attack in a given location within a certain period of time. If this event occurs, those who own the contracts would receive a certain payout, with the value of the contract prior to that time representing the collective market assessment of the attack's likelihood as the contract changes hands between traders based on information that they each have. One notable example of such a system is the Iowa Electronic Markets (IEM) operated by the Tippie College of Business at the University of Iowa. The IEM began as an attempt to predict the outcome of US presidential elections while offering an alternative to traditional opinion polls, but it has since expanded its mandate to answering a wide range of questions on other issues.[4]

Unfortunately, the IEM remains somewhat unique because proponents of creating other prediction markets have met with some challenges to their implementation due to regulatory restrictions under which they would be classified as internet gambling sites and therefore illegal in many places.[5] It is also possible to operate them without participants placing real money at stake but there are questions regarding how this removal of a real-world incentive might impact the accuracy of market predictions. An additional obstacle has been the negative connotations that some people associate with the concept of participants betting on - and therefore potentially profiting from - violent events such as terrorist attacks, assassinations, and atrocities. This was the source of the somewhat misguided outcry which met the establishment of the Policy Analysis Market (PAM), which was a US Government-funded prediction market set up in 2003. Although intended to support counterterrorism efforts, two senators successfully campaigned against what they viewed as a "grotesque" idea and the PAM was shut down.[6] Some similar projects have been able to operate in the private sector without significant constraints because they have been kept internal to the employees of the corporations which use them to make business decisions. However, due to these logistical and image-related difficulties, prediction markets have been deemed inappropriate for predicting genocide at this time.

---

[3] Tetlock, Philip. *Expert Political Judgment.* Princeton: Princeton University Press, 2012.
[4] Iowa Electronic Markets. Tippie College of Business, University of Iowa, 2012. Internet, accessed 20 June 2012 <http://tippie.uiowa.edu/iem>
[5] Arrow et al. "The Promise of Prediction Markets," *Science Magazine*, 16 May 2008.
[6] BBC News. "Pentagon axes online terrorist bets," 29 July 2003. Internet, accessed 20 June 2012 <http://news.bbc.co.uk/2/hi/americas/3106559.stm>

The other option mentioned above, conducting expert surveys, offers a simpler and therefore initially more attractive option for aggregating opinions into an overall group response. In its simplest form, such an approach simply requires the creation of a mechanism for creating properly structured questions, soliciting responses, and properly combining them. The method of aggregation used may range from the calculation of a simple mean to more complex weightings of individual responses based on participant expertise and past performance. One such system currently in place is known as Forecasting ACE, a system funded by the Intelligence Advanced Research Projects Activity of the US Government.[7] In order to qualify as a "smart crowd" according to Surowiecki, the system and the participants involved must meet four criteria: diversity of opinion, independence, decentralization, and aggregation.[8] These requirements will be addressed below.

Admittedly, there are some general weaknesses to this approach which will have to be addressed. For example, as mentioned above, groups are most suitable for answering questions which have known answers. Without the ability to test the accuracy of the group response by checking it against reality, the value of such a system is difficult to assess. This will naturally be a challenge for a system seeking to predict genocide since the only way to determine the accuracy of any prediction is to wait and see whether or not the predicted events occur. Naturally, predicting genocide does not present us with the benefit of knowing the "correct" answer. Furthermore, as Vul & Pashler observed when studying "the crowd within," individuals making multiple probabilistic judgements (though this also applies to other types of answers) about the likelihood of an event occurring tend to be biased towards the extremes.[9] It is possible that this same effect could carry over into groups, with respondents tending to indicate either no risk or else very high risk of genocide but with few responses in between.

Proposed System for Assessing Genocide Risk

Clearly, the concept of using the "smart crowd" concept to aggregate multiple opinions has applicability to assessing the risk of genocide – with some modification – since it enables subjective qualitative judgements to be quantified in a way that should mitigate bias and produce a more accurate response than individuals can. This should provide an ideal complement to more traditional methods of statistical modeling to forecast genocide. This system can be implemented in five phases:

1. *Assemble the "smart crowd"* – Although crowds are conventionally considered to be wise when they are drawn from a wide (and preferably random) range of backgrounds, with a specific and frequently misunderstood topic like genocide it may be detrimental to the accuracy of the system to merely use random members of the general public as participants. For this reason, and also because the goal is to aggregate and quantify expert opinions, individuals will be required to possess at least a basic understanding of what genocide is and the nature of the phenomenon. Participants will be recruited from academic institutions and civil society organizations focused on genocide.

[7] Forecasting ACE. Applied Research Associates, Inc. Internet, accessed 20 June 2012 <http://www.forecastingace.com/index.php>
[8] Surowiecki, *The Wisdom of Crowds.*
[9] Vul, E & Pashler, H (2008) "Measuring the Crowd Within: Probabilistic Representations within Individuals," *Psychological Science*. 19(7) 645-647.

2. *Profile participants* – Although participants will remain anonymous throughout the study and not communicate with each other, gathering some basic information on each of them will be useful for later analysis. In particular, identifying any relevant regional or thematic expertise which they may have will be useful for later analysis since it can be used to experiment with weighting participant responses, a concept which is explained further below.

3. *Distribute "seed" knowledge* – Traditionally, wise crowds are intended to make use of knowledge locally available to each participant. However, due to the significant complexity of questions relating to genocide risk, establishing a base level of understanding throughout the group will be necessary in order to eliminate pure guessing. This will be accomplished by preparing and distributing to participants small briefing packages which profile the country in question. A possible consideration in order to eliminate bias will be to remove country names from these packages. It is also recognized that some participants may have prior knowledge of the country, which will hopefully be identified in phase two above and then considered during the aggregation phase below.

4. *Pose questions* – Participants will be provided with carefully-worded questions in order to elicit their responses. Since the goal here is quantification, the answers requested will likely be in the format of ratings on a simple numerical scale, with a free text option for entering further information.

5. *Aggregate responses* – Once individual participants have submitted their responses through a private, secure means, these will be combined to arrive at a group answer. Initially, a simple mean will be calculated for the numerical ratings, though further analysis can be conducted based on the methods proposed in the section below.

Variations, Comparisons, and Testing

Although the initial trials using this system to assess the risk of genocide will simply calculate the simple mean of multiple individual risk scores in order to aggregate participant opinions, more advanced work can be done to build on this basis. This includes variations on the types of questions asked, comparing the results of adjustments to the respondent group characteristics, and tests for accuracy. Specifically, future enhancements will include the following:

- *Experts vs. general population (comparison)* - One of the key assumptions of the proposed system is that genocide experts will produce more effective predictions of where genocide is expected to take place. In order to validate this, it will be necessary to compare results to a non-expert sample recruited from the general population.

- *Weighted assessments vs. simple mean of risk rankings (comparison)* - Initial testing will be conducted with an aggregation mechanism based on calculating the simple mean of all participant risk scores. However, more advanced versions will need to incorporate weighting which considers the degree to which the area of expertise of each participant

influences the accuracy of his or her assessments. The underlying hypothesis is that, for example, experts who specialize in studying the preconditions which lead to genocide would provide more accurate predictions than those who study post-extermination aspects of genocide such as commemoration.

- *Confidence vs. accuracy* - As pointed out above, Tetlock has identified high expert confidence as being correlated to low degrees of accuracy. If this is true, it will be necessary to assess and track the degree of participant confidence over time and then compare this with their individual levels of accuracy. Aggregating group confidence and comparing with group accuracy will add a further level of complexity which is worth examining.

- *Vary group sizes* - Conventional wisdom dictates that larger sample sizes tend to lead to more accurate surveys, which would logically extend to group intelligence situations as well. This possibility will need to be tested in order to identify the point (if at all) at which accuracy returns begin to diminish significantly, indicating an optimal group size.

- *Back-testing* - Assessing the accuracy of this system requires the ability to check survey results against real-world outcomes, which may be very difficult, impossible, or simply too slow of a process when it comes to predicting future events such as potential genocide. Therefore, it will be necessary to test the concept of the system by providing information packages based around anonymized profiles of countries which have experienced genocide in the past.

- *Add more dimensions to the answers requested* - In addition to simply assessing the likelihood of genocide taking place in a given country, other parameters must be specified in order to produce useful forecasts. Eventually, participants can also be asked to answer questions relating to time horizon (i.e. when genocide will take place if it is likely), likely triggers, and the projected magnitude of its impact.

- *Tracking participant accuracy* - Building on the idea mentioned above which relates to weighting individual participant responses based on their areas-of-expertise, a more advanced option will be dynamic weighting which changes over time in relation to the accuracy of individual participant. For example, consistently accurate participants will be given increasing weight while inaccurate participants will have their weightings reduced, thus increasing the collective accuracy of the overall group.

Conclusion

In closing, this remains an untested concept with regard to assessing the risk of genocide but holds exciting potential for quantifying large numbers of expert opinions and helping to bridge the continuing gap between quantitative and qualitative approaches to this issue. The project will now enter a testing phase for which individuals with expertise in the genocide studies and prevention field are encouraged to volunteer.