Subject: Supplemental discussion: 1.) Content analysis and the NYTimes on Google's translation achievements; 2.) Faster national learning rates in world politics.

Dear Dr. Fischhoff & Colleagues:

      Two notes about topics that I raised earlier:

1.) <u>Content analysis</u>
      Concerning the potential for rapid learning systems about a wide range of issues, based on a strategic plan to develop new content analysis capabilities: I am forwarding (below) a copy of an article in this morning's <u>New York Times</u> that gives more detail about the new statistical processing strategies/capabilities for translation that I discussed in memorandum # 15. NSA's capabilities probably equal or exceed Google's. It would be an inspired idea to place these large reference/ R&D databases and analysis tools in the public domain to accelerate further innovations via universities and others. The translation capabilities are only one step in the agenda to develop the full power of content analysis, but they are an important early step.

2.) <u>National Learning Rates</u>
      I raised the concern, in several earlier messages, about continuing misperceptions and the opportunity to establish databases and analysis tools that can improve national learning rates in world politics in the 21st century.

      An interesting project to recommend might be a propositional inventory/variables list based on the Breslauer and Tetlock edited volume from the National Academy, <u>Learning in US and Soviet Foreign Policy</u> (1991) and other discussions of political behavior (e.g., Jervis). A propositional inventory and visual display map (along the lines that Brewster Smith created, and that I developed to establish baselines in domestic policy analysis (<u>The Case of the Unreturned Cafeteria Trays</u>), could be useful to mine this volume, organize its contributions syste-

matically, and develop the field further.

I never had the time to do this. But it could be a useful step to recommend, or for your staff to undertake.

[Re learning: You also might want to push - hard - on the question of whether the Reagan [Shultz/Gorbachev] variables involved in ending the Cold War are included in Breslauer and Tetlock's tour de horizon of earlier work. Mann's new <u>The Rebellion of Ronald Reagan</u> contains further detail from these years, including the revelation that the CIA's lead analyst of Soviet behavior (Robert Gates, now Secretary of Defense) was allied with almost everyone else in the intelligence community and Republican Party in opposing Reagan's project to end the Cold War and the arms race. (Gates, now, says that he was wrong - but I am not aware of any deep, self-reflective analysis about why he was wrong.) The Cold War and nuclear arms race might not have ended with most other Realpolitik analysts and political leaders on the US side: Reagan, it now turns out, invited Nixon to a secret meeting at the White House to elicit his support (but did not get it). I don't think that we have yet mined this case or understood these variables: Successful politicians can think about the world very differently than professional career analysts with huge data systems and applying statistical analysis.]

Yours truly,
Lloyd Etheredge

The New York Times
March 8, 2010
Google's Computing Power Refines Translation Tool
By MIGUEL HELFT

MOUNTAIN VIEW, Calif. — In a meeting at Google in 2004, the discussion turned to an e-mail message the company had received from a fan in South Korea. Sergey Brin, a Google founder, ran the message through an automatic translation service that the company had licensed.

The message said Google was a favorite search engine, but the result read: "The sliced raw fish shoes it wishes. Google green onion thing!"

Mr. Brin said Google ought to be able to do better. Six years later, its free Google

Translate service handles 52 languages, more than any similar system, and people use it hundreds of millions of times a week to translate Web pages and other text.

"What you see on Google Translate is state of the art" in computer translations that are not limited to a particular subject area, said Alon Lavie, an associate research professor in the Language Technologies Institute at Carnegie Mellon University.

Google's efforts to expand beyond searching the Web have met with mixed success. Its digital books project has been hung up in court, and the introduction of its social network, Buzz, raised privacy fears. The pattern suggests that it can sometimes misstep when it tries to challenge business traditions and cultural conventions.

But Google's quick rise to the top echelons of the translation business is a reminder of what can happen when Google unleashes its brute-force computing power on complex problems.

The network of data centers that it built for Web searches may now be, when lashed together, the world's largest computer. Google is using that machine to push the limits on translation technology. Last month, for example, it said it was working to combine its translation tool with image analysis, allowing a person to, say, take a cellphone photo of a menu in German and get an instant English translation.

"Machine translation is one of the best examples that shows Google's strategic vision," said Tim O'Reilly, founder and chief executive of the technology publisher O'Reilly Media. "It is not something that anyone else is taking very seriously. But Google understands something about data that nobody else understands, and it is willing to make the investments necessary to tackle these kinds of complex problems ahead of the market."

Creating a translation machine has long been seen as one of the toughest challenges in artificial intelligence. For decades, computer scientists tried using a rules-based approach — teaching the computer the linguistic rules of two languages and giving it the necessary dictionaries.

But in the mid-1990s, researchers began favoring a so-called statistical approach.

They found that if they fed the computer thousands or millions of passages and their human-generated translations, it could learn to make accurate guesses about how to translate new texts.

It turns out that this technique, which requires huge amounts of data and lots of computing horsepower, is right up Google's alley.

"Our infrastructure is very well-suited to this," Vic Gundotra, a vice president for engineering at Google, said. "We can take approaches that others can't even dream of."

Automated translation systems are far from perfect, and even Google's will not put human translators out of a job anytime soon. Experts say it is exceedingly difficult for a computer to break a sentence into parts, then translate and reassemble them.

But Google's service is good enough to convey the essence of a news article, and it has become a quick source for translations for millions of people. "If you need a rough-and-ready translation, it's the place to go," said Philip Resnik, a machine translation expert and associate professor of linguistics at the University of Maryland, College Park.

Like its rivals in the field, most notably Microsoft and I.B.M., Google has fed its translation engine with transcripts of United Nations proceedings, which are translated by humans into six languages, and those of the European Parliament, which are translated into 23. This raw material is used to train systems for the most common languages.

But Google has scoured the text of the Web, as well as data from its book scanning project and other sources, to move beyond those languages. For more obscure languages, it has released a "tool kit" that helps users with translations and then adds those texts to its database.

Google's offering could put a dent in sales of corporate translation software from companies like I.B.M. But automated translation is never likely to be a big moneymaker, at least not by the standards of Google's advertising business. Still, Google's efforts could pay off in several ways.

Because Google's ads are ubiquitous online, anything that makes it easier for

people to use the Web benefits the company. And the system could lead to interesting new applications. Last week, the company said it would use speech recognition to generate captions for English-language YouTube videos, which could then be translated into 50 other languages.

"This technology can make the language barrier go away," said Franz Och, a principal scientist at Google who leads the company's machine translation team. "It would allow anyone to communicate with anyone else."

Mr. Och, a German researcher who previously worked at the University of Southern California, said he was initially reluctant to join Google, fearing it would treat translation as a side project. Larry Page, Google's other founder, called to reassure him.

"He basically said that this is something that is very important for Google," Mr. Och recalled recently. Mr. Och signed on in 2004 and was soon able to put Mr. Page's promise to the test.

While many translation systems like Google's use up to a billion words of text to create a model of a language, Google went much bigger: a few hundred billion English words. "The models become better and better the more text you process," Mr. Och said.

The effort paid off. A year later, Google won a government-run competition that tests sophisticated translation systems.

Google has used a similar approach — immense computing power, heaps of data and statistics — to tackle other complex problems. In 2007, for example, it began offering 800-GOOG-411, a free directory assistance service that interprets spoken requests. It allowed Google to collect the voices of millions of people so it could get better at recognizing spoken English.

A year later, Google released a search-by-voice system that was as good as those that took other companies years to build.

And late last year, Google introduced a service called Goggles that analyzes cellphone photos, matching them to a database of more than a billion online images, including photos of streets taken for its Street View service.

Mr. Och acknowledged that Google's translation system still needed improvement, but he said it was getting better fast. "The current quality improvement curve is still pretty steep," he said.

Dr. Lloyd S. Etheredge - Director, Government Learning Project
Fellow, World Academy of Art & Science
Policy Sciences Center Inc.
127 Wall St., Room 322 - Box 208215
New Haven, CT 06520-8215
URL: www.policyscience.net
301-365-5241 (v); lloyd.etheredge@policyscience.net (email)