Post-Mortem: Lessons Learned in Superforecasting
the Russian Invasion of Ukraine

Superforecasters at Good Judgment Inc are an international team of skilled generalists with a decade-long track record[1] of making highly accurate and well-calibrated forecasts.

Between January 21 and February 10, 2022[2], our Superforecasters thought that the probability of a Russian invasion of Ukraine was under 50%. Starting on February 11, however, the consensus forecast began pointing toward invasion as the more likely outcome. *Should we have come to that conclusion sooner? Did our forecasts appropriately reflect the publicly available information at each point in time? Were we over- or underconfident about our forecasts, individually and collectively? Could we have been more accurate, earlier?*

A *post-mortem* exercise is a key step in the Superforecasting process. For every forecast we make, we keep score so we can learn and improve. Here, we examine our forecasting on the Russia-Ukraine question and present lessons learned. Superforecaster comments appear in italics.[3]

For this question, while Superforecasters have differing perspectives about their individual and collective forecasts, many see the following elements as contributing to a consensus that was "on the wrong side of maybe" for more than two weeks:

- **too much attention to certain base rates;**
- **underestimation of Putin's willingness to take major risks;**
- **and too much discounting of US intelligence releases as efforts to *influence* events rather than to *inform*.**

---

[1] Good Judgment's methodology during the four-year IARPA forecasting tournament generated up to a 75% increase in accuracy over baseline (see Mellers B., et al. 2015. "Identifying and Cultivating Superforecasters as a Method of Improving Probabilistic Predictions." *Perspectives on Psychological Science* 10(3): 267-281. doi: 10.1177/1745691615577794). In 2013-14, many of these Superforecasters were compared directly to 4,300 members of the US Intelligence Community on the same set of 139 geopolitical forecasting questions. Their aggregated forecasts were about 35% more accurate (see Goldstein S., et al. 2015. "Assessing the Accuracy of Geopolitical Forecasts from the US Intelligence Community's Prediction Market." Unpublished manuscript approved for public release, available at https://goodjudgment.com/wp-content/uploads/2020/11/Goldstein-et-al-GJP-vs-ICPM.pdf).

[2] In the period of 13-15 February, Superforecasters assigned a higher probability to an invasion (51% to 59%) but on 16-17 February reverted for 48 hours to the view that an outright invasion was less likely than de-escalation.

[3] Clients who would like to see more detailed commentary than is included in our standard reports are encouraged to contact their account representative to discuss options.

We discuss each of these points below. We take these internal critiques seriously, but we are also aware of the importance of avoiding "certainty-of-hindsight" bias. Just because an invasion did occur does not mean that it was inevitable.

Drivers of the Consensus Forecast

The forecasting question "Between 25 January 2022 and 31 May 2022, will NATO and/or a NATO member state accuse Russian national military forces of invading Ukraine?" opened on 21 January 2022 with:

- 60% probability for "No" and
- 40% probability for a combination of the "Yes" answer options
    - "Yes, invading the Donetsk and/or Luhansk Oblasts as well as other Ukrainian territory" (28% probability),
    - "Yes, invading the Donetsk and/or Luhansk Oblasts only" (8% probability), and
    - "Yes, invading Ukrainian territory outside of Donetsk and/or Luhansk Oblasts only" (4% probability).

Figure 1. Consensus forecast
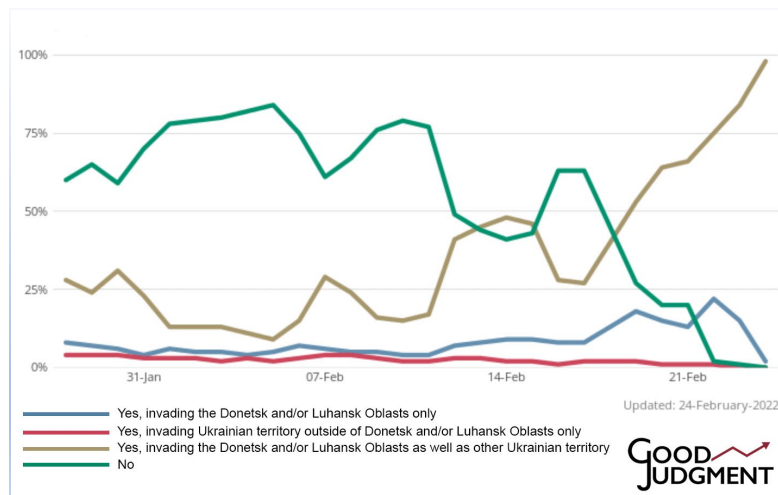
The consensus forecast shifted to the answer option "Yes, invading the Donetsk and/or Luhansk Oblasts as well as other Ukrainian territory"—the eventual resolution—between 13 and 15 February and then again, sharply, from 18 February onward. But for 19 days out of the 32 that this forecasting question was open, we thought "no invasion" was more likely than invasion.

Base rate. Superforecasters take seriously Nobel laureate Daniel Kahneman's injunction to be "base rate-sensitive" and to start work on any new forecasting question by considering the "outside view," asking, "How often does this sort of thing happen?" The base rate of a war breaking out in Europe in any given year is low. Therefore, forecasting a high likelihood of invasion early on would normally have suggested a forecaster was being base rate-*in*sensitive. However, we likely overvalued the "outside view" of "no war in Europe" in this instance, whereas the "inside view"—e.g., Putin's consistent denial of Ukraine's right to sovereignty, signals from US intelligence—should have received more weight in our forecast earlier.

> *"It was a mistake for the broader team not to have taken Putin's own words, delivered in his summer letter, more seriously. While an invasion might not have been a foregone*

conclusion (I didn't think it was, as my forecasts reflect), when a man who has launched military operations in Georgia, Syria, and Ukraine makes accusations of 'genocide' and 'weapons of mass destruction' being used against Russia, and then assembles an invasion army—for the second time, after his first attempt at coercive diplomacy failed— it's time to worry. A lot."

Superforecasters know that it is critical to identify the *right* base rate(s) for comparison. The pattern of Putin's behavior that this forecaster identified might be considered a base rate for how Vladimir Putin's Russia will behave, which arguably should have received as much or more weight than the more generalized base rate of whether a war would break out in Europe.

"Changes in Russian activity over the last two years justify a substantial increase in the forecasted risk of invasion over the base rate. That base rate of large-scale, overt Russian military campaigns has been low, but the following observations point towards a transition from a period of Moscow's tolerance of the status quo *to a period characterized by 'military and technical measures' directed towards confronting it: 1. Increased militarization of Russian civil society normalizes a long-term wartime posture; 2. Divergence of Russian leaders' interests from Western understanding of Russian interests sets the stage for a 'failure of imagination' in Western risk assessment. The retention of regime members' personal status and power, and the reversal of the post-Soviet status quo, are far more salient to Moscow than avoiding Western sanctions; 3. Rejection of the legitimacy of the status quo in Ukraine. Moscow has consistently stated that it considers the leadership and security concerns of Ukraine, and its very foundation as a country, illegitimate. If anything, Russia has been employing increasingly crude and contemptuous language towards the Ukrainian state."*

Rationality arguments. Many of us mistakenly imposed our schema of rationality and values on Russian actors. In fact, it seems Putin was willing to put more at risk than several Superforecasters had assumed.

"People talked about the trade-offs of the invasion and concluded that Putin could get what he wanted without an invasion. So, the rational thing would be to not invade. Even now, the media is having a field day pontificating that Putin might win in the short term but in the long term will pay a heavy price. It just goes to reinforce the idea that we can't figure out why an invasion was the smart thing to do."

"While telling myself I could not get inside Putin's head, when forecasting I made assumptions based on what I would do if I were sitting in the Kremlin. Would have been better to accept that Putin's thinking was beyond the frontier of irreducible uncertainty, at least for me."

"I was very wrong, and from reading other comments, I guess the main reason was that I was assuming rational actions too. As I highlighted in my forecasts, the best time to

invade would have been much earlier, taking advantage of surprise, before Putin's intentions were telegraphed all over and giving Ukraine, NATO, and the US time to prepare their responses. However, I should have certainly paid more attention to the continued build-up that went beyond mere exercises."

Pragmatism arguments. Related to the rationality arguments were the views of some Superforecasters that they failed to see sufficient pragmatic reasons for Russia to invade Ukraine, as the invasion was guaranteed to bring on Western sanctions and a pledge from Ukraine to fight back—in many cases using modern weapons supplied by Western allies.

"When the Soviet regime invaded Czechoslovakia in August 1968, it did so with 250k troops. They were expecting to just need to control the civilian population, not expecting any resistance from the Czechoslovak army, which was communist-controlled. They then needed an additional 250k troops for the occupation. Thus, somewhere approaching 500k troops just to subdue the civilian population of a country of about 15 million people. Ukraine has 44 million people, with a military of over 255k, and in this case the military will fight hard and oppose an invasion."

"I did not foresee the unprovoked invasion at all, thinking that a forever 'imminent threat' would be enough."

"The argument for 'No' hinged heavily on all the Russian military movement being an elaborate bluff. One possible lesson is that when bluffs get too elaborate, they aren't bluffs. But elaborate bluffs occur and turn out to be either bluffs or the diplomats manage to avoid war (e.g., Cuban missile crisis)."

"At the point we started considering this question, it was potentially an ornate bluff meant to convince not only foreigners but everyone in Russia. Second, especially near the end a few weeks back, it was a game of chicken with the Ukrainian people and government. Who would play chicken with Putin?"

"Some people focused on the logic of it. Some focused on the facts of it. There was a huge gap between those two groups. Those focused on 'it doesn't make sense to me' could not be persuaded."

Misreading "inside view" information: US intelligence signals. Superforecasters discussed early on the fact that people with access to classified information would be able to spot the signal of an impending invasion before it appeared in public channels. They took note of Washington's unusual tactic of sharing intelligence assessments on Russia with the media, but many focused on past intelligence failures (e.g., Iraq 2003) and were perhaps more skeptical of those signals than they should have been.

"I was unsure how much of the release of US intelligence was intended to influence vs inform. I erred in upweighting the former. Turns out the latter was important too. US intel did a very fine job of calling what was going to happen, how, when, where. I should have increased my forecast for invasion sooner."

"Did we ignore US intelligence assessments more than we should have? That's a hard question given the spotty track record of US intelligence estimates and the fact that the release of those estimates was less to inform than it was an attempt to ward off, and/or define, the conflict."

"The accuracy of US intelligence was surprising. While they get lots of calls right, people may have the perception that they're less accurate when it comes to highly politicized issues. The US and UK were trying to influence the narrative, which made some people skeptical, but it doesn't seem like they were selectively releasing intelligence. Rather, they were making publicly available what they would have relied upon anyway."

Lack of a Red Team/*Pre-Mortem* Process: In forecasting, getting it "right" is only half of the picture. Good Judgment emphasizes the importance of "being right for the right reasons." To check their thinking, Superforecasters always strive to consider alternatives, actively looking for reasons why their forecast may be wrong.

"I think my forecast was right for the right reasons, in that I had a view about the most likely sequence of events, and that's basically what happened. However, I think my forecast was 'brittle' in the sense that a slightly different set of inputs could have produced a dramatically different forecast. Other sequences of events were possible and if they had occurred, I would have been way off."

Good forecasters engage in *pre-mortems*—assuming that the future has unfolded differently from the way they forecast and thinking about how that unexpected future might have come about. We often assign a "Red Team" of Superforecasters to lead such *pre-mortem* thinking but failed to do so this time. That was a mistake.

Divergent Views of Superforecasters Who Saw Invasion As Highly Likely

Our consensus forecast masked the fact that there was a wider variety of divergent views among the Superforecasters. A significant minority thought invasion was highly likely throughout their forecasting on this question.

Losing face. Once a significant number of Russian troops were placed in strategic positions, some Superforecasters argued there was little realistic possibility for Putin to de-escalate without either getting major concessions or losing face.

> *"Russia has doubled down on unrealistic demands both internally and internationally. Backing down now would mean that they are unlikely to be taken seriously again. Not much more room to escalate for next time…. Therefore, Russia knows that there will be consequences. What would make the (risk of) consequences worth it? Merely expanding the costly frozen conflict in the two* oblasts *seems like a waste of time. Creating a land bridge to Crimea may be goal worth the pain."*

Putin's views. Putin's rejection of Ukraine's right to sovereignty was mentioned by some Superforecasters as a crucial signal. It subsequently played out prominently in Putin's aggressive speeches of February 21 and February 24. Another factor some Superforecasters underlined was Putin's focus on his "legacy" and his increasingly small inner circle, raising the possibility of a skewed cost analysis within the Kremlin.

> *"I still expect some form of military action (standoff strikes against Ukrainian military targets, etc.). I cannot see how Putin will climb down the escalation ladder without some significant concessions based on his demands, and those key ones are not forthcoming. It is also possible that Putin might have a different cost calculus and may be more optimistic about his success than Ukrainians and/or others. A lot of wars did start after all because one side was more optimistic about its chances than they actually were."*

Lessons Learned

Many of us paid too much attention to the low base rate of war in Europe. That base rate may be undergoing significant change in the 21st century, and going forward, we will keep in mind the possibility that thresholds for state-on-state violence are now lower in the post-Soviet space. We will also be more careful to consider multiple base rates.

Many of us underestimated how much Russian President Vladimir Putin was willing to put at risk. It is always harder to forecast on questions where the outcome is largely under the control of one person. Predicting the thinking, values, and trade-offs of a group is easier than predicting the decisions of a single person. Going forward, we will work to remain epistemically humble about what one can possibly know about a single actor's thinking.

Epistemic humility also goes to the issue of possible overconfidence in either *direction.* We cannot know what the correct probabilities would have been at each point in the forecasting window, but find this Superforecaster comment insightful:

> *"I find it difficult to believe that anything more than a 50% chance of invasion was realistic [early on], with a high degree of irreducible certainty, with so many parties involved, and so many bilateral and multilateral confabulations happening."*

We questioned whether releases of US intelligence were meant to influence more than to inform, and too many of us recalled recent intelligence failures with more acuity than the much

larger set of intelligence successes. We discussed early on the fact that people with access to classified information would be able to spot the signal of an impending invasion before it appeared publicly. In hindsight, our forecast of a full-scale invasion should have shot upward as the White House continued to release intelligence about the impending full-scale invasion.

Methodological Notes and the Importance of Keeping Score

Keeping score is what makes a *post-mortem* a key step in the Superforecasting process. By keeping score, we can learn and get better. This, in turn, leads to better forecasting and is the reason why we have had an impressive track record in forecasting accuracy as the Good Judgment Project in 2011-2015 and Good Judgment Inc since 2015.

But keeping score is complicated; unless a forecast was either 0% or 100%, it is impossible to say with certainty whether it was "right" or "wrong." Rather, determining the accuracy of any prediction that is not 0% or 100% requires examination of a set of forecasts and the determination of how *calibrated* they are. For example, the set of forecasts of 75% are "right" if those events

Figure 2. Several Superforecasters cautioned that because the base rate for invasions in Europe is very low, a 1-in-3 chance of a Russian invasion was an alarming signal.

occur 75% of the time. So consider our 17 February consensus forecast of 63% for de-escalation. Was that "wrong"? Or was it "right" and we are now in the universe that was seen as 37% likely at that time?

As part of our ongoing *post-mortem*, we will examine the calibration of all our forecasts related to military action and to events involving Russia. We also routinely re-examine all our forecasting methodologies, including aggregation algorithms and the ways we communicate forecasts to clients.