

WHO SHOULD WE BELIEVE ABOUT IRAN?

We scored 113 predictions by 15 foreign policy experts—hawks and doves alike—to build a data-driven forecast of the US-Iran conflict.

March 1, 2026

I. The Fog of Prediction

In the twenty-three years since the invasion of Iraq, the American foreign policy establishment has generated an enormous volume of predictions about the Middle East. Some of those predictions were spectacularly right. Many more were spectacularly wrong. And yet, when a new crisis erupts—as it has now, with Washington and Jerusalem weighing military options against Iran—the same rotating cast of analysts, think-tankers, and former officials is summoned to offer guidance. The question almost nobody asks is: *which of these people have actually been right before?*

This report attempts to answer that question with data rather than credentials. We identified fifteen prominent analysts who have written extensively on Iran, the broader Middle East, and American military interventions. We then went back through the public record—op-eds, congressional testimonies, books, television appearances, and policy papers—and scored 113 specific, falsifiable predictions they made over the past two decades. We graded each prediction on a simple five-point scale from FALSE to TRUE, based on what actually happened.

The fifteen experts were chosen to represent the full ideological spectrum. Seven lean dovish or anti-interventionist. Six are hawkish or pro-interventionist. Two defy easy categorization. Critically, we did not weight anyone’s opinion by their ideology. We weighted them by their track record. If the hawks had been more accurate than the doves, the model would reflect that. As it turns out, the data tells a more nuanced story.

* * *

II. The Realists Who Called Iraq

John Mearsheimer is a professor of political science at the University of Chicago and one of the most cited scholars in international relations. In September 2002, he co-authored an advertisement in the *New York Times* under the headline “WAR WITH IRAQ IS NOT IN AMERICA’S NATIONAL INTEREST.” Signed by 33 university-based analysts, it warned that an invasion would be a strategic disaster. In a paper co-authored with Stephen Walt in *Foreign Policy* (“An Unnecessary War,” January 2003), they argued that Iraq could be contained and that invasion would destabilize the region.

Source: Mearsheimer, J.J. and Walt, S.M., “An Unnecessary War,” *Foreign Policy*, Jan/Feb 2003, pp. 50–60. Also: “War with Iraq Is Not in America’s National Interest,” *New York Times* advertisement, September 26, 2002.

Twenty-three years later, that prediction looks almost painfully prescient. Iraq became a quagmire that cost trillions of dollars and hundreds of thousands of lives, destabilized the entire region, and empowered Iran—the very outcome Mearsheimer warned about. He went on to predict that NATO expansion would provoke Russian aggression against Ukraine (2014), that Libya intervention would create a failed state (2011), and that the Ukraine conflict would become a prolonged war of attrition (2022). All of these came true.

Of eight major predictions we scored, Mearsheimer was right on seven and mostly right on the eighth. His overall accuracy: **86.1%—the highest in our panel.**

Ali Vaez, the Iran director at the International Crisis Group, is less famous but nearly as accurate. Vaez predicted in 2018 that US withdrawal from the JCPOA nuclear deal would accelerate Iran’s uranium enrichment—which it did, dramatically. He predicted that maximum pressure sanctions would fail to change regime behavior, that Iranian protests would be met with crackdowns but not regime collapse, and that Iran would accelerate its nuclear program after the JCPOA fell apart. He was right on all counts. His accuracy: **82.1%.**

Sources: Vaez, A., *International Crisis Group reports on Iran, 2018–2022*. See especially “How Europe Can Save the Iran Nuclear Deal,” *ICG*, 2018; and “The Iran Nuclear Deal at Four: A Requiem,” *ICG Middle East Report No. 210*.

Stephen Kinzer, a former *New York Times* correspondent and author of *All the Shah’s Men* and *Overthrow*, built his career arguing that American regime change operations consistently produce outcomes worse than the status quo. From the 1953 Iranian coup to Libya in 2011, Kinzer’s thesis has held up remarkably well. His accuracy: **78.1%.**

* * *

III. When Being Right Destroys Your Credibility

Not all doves turned out to be reliable. **Scott Ritter**, the former UN weapons inspector, made one of the most consequential correct calls of the 21st century. In

2002, while the Bush administration was insisting that Iraq possessed vast WMD stockpiles, Ritter went on CNN, testified at forums, and wrote a book arguing that Iraq had been effectively disarmed. He told a Harvard audience in September 2002 that inspectors had “fundamentally disarmed Iraq” by December 1998.

Source: “*Former weapons inspector: Iraqi arms ‘gone’ as of 1998,*” *Harvard Gazette*, September 2002. **Also:** Ritter, S. and Pitt, W.R., *War on Iraq: What Team Bush Doesn’t Want You to Know*, 2002.

He was right. The Iraq Survey Group, led by Charles Duelfer, ultimately confirmed that Iraq had shuttered its WMD programs after the first Gulf War. As the journalism watchdog FAIR documented, Ritter was one of the very few prominent voices who got the biggest question of the decade correct.

Source: “*Wrong on Iraq? Not Everyone,*” *FAIR (Fairness & Accuracy in Reporting)*, 2006.

But Ritter’s subsequent track record has been dismal. He predicted an imminent US attack on Iran in June 2005—it never came. He claimed Syria’s al-Kibar facility was not nuclear; it was. He predicted a decisive Russian military victory in Ukraine in 2023; that didn’t happen either. He denied Russian responsibility for the Bucha massacre, a position unsupported by the evidence. His overall accuracy collapsed to **37.5%**. A cautionary tale: being right once, even on the most important question of your era, does not make you a reliable forecaster.

Stefan Molyneux, an independent commentator, fared even worse. He predicted Western civilization would collapse from immigration, that Brexit would trigger EU dissolution, and that Iran would capitulate after the JCPOA withdrawal. None of this happened. His accuracy: **12.5%—the lowest in our panel**. The model gives him essentially zero weight.

* * *

IV. The Hawks: Strong on Iran, Wrong on Iraq

To ensure our model wasn’t just an echo chamber of anti-interventionist scholars, we added six analysts from the hawkish side of the spectrum—mostly affiliated with the Foundation for Defense of Democracies (FDD) and the American Enterprise Institute (AEI). These are the people who designed the “maximum pressure” sanctions campaign, who argued most forcefully that the JCPOA was fatally flawed,

and who have been advocating for a harder line on Iran for decades. Their predictions tell a fascinating, complex story.

Start with the worst performer. **Reuel Marc Gerecht**, a former CIA case officer and senior fellow at FDD, wrote a *New York Times* op-ed in November 2002 with perhaps the most unfortunate headline in the paper's history: "*An Iraq War Won't Destabilize the Mideast.*"

Source: Gerecht, R.M., "An Iraq War Won't Destabilize the Mideast," *The New York Times*, November 26, 2002.

The Iraq War did, of course, destabilize the Mideast—catastrophically. Gerecht also predicted that a US invasion of Iraq would provoke a democratic revolution in Iran and that Iran was "in no shape for prolonged confrontation" with the United States. Both predictions were wrong. His accuracy: **33.3%**.

But here is where the data gets interesting. Not all hawks share Gerecht's track record. **Behnam Ben Taleblu**, the senior director of FDD's Iran Program, has been consistently right on the things he knows best. As early as 2020, he co-authored analysis in *Breaking Defense* warning that Iran would transfer drones and missiles to Russia once international arms embargoes were lifted. This prediction was validated dramatically when Iranian Shahed-136 drones began raining down on Ukrainian cities in late 2022.

Source: Hardie, J., Brobst, R. and Ben Taleblu, B., "Iranian drones could make Russia's military more lethal in Ukraine," *Breaking Defense*, July 27, 2022. Also: FDD analysis, "Iran, Russia Expedite Building Drone Factory in Russia," February 9, 2023.

Ben Taleblu also correctly identified Houthi missile capabilities as a genuine threat years before the Red Sea crisis, and warned that Iran's cruise missile capability was underestimated by Western intelligence. His accuracy: **71.9%—the highest among the hawks**, and higher than several of the doves.

Ray Takeyh, a senior fellow at the Council on Foreign Relations, occupies a unique position: a sophisticated hawk who has been right about many things that simpler hawks have gotten wrong. Takeyh correctly predicted that the JCPOA's sunset provisions would leave Iran well-positioned for a nuclear breakout, that Iran's defense budget would surge after the deal, and—crucially—that the Iranian regime is too multi-layered to be toppled by bombing alone. Unlike Gerecht and some of the more optimistic hawks, Takeyh explicitly argues that air power can destroy nuclear

and missile programs, but that the regime will survive and rebuild. His accuracy: **66.7%**.

Mark Dubowitz, the CEO of FDD and one of the principal architects of the maximum pressure sanctions campaign, has a mixed record. He was right that the JCPOA created a “patient pathway to the bomb” through sunset provisions, right that killing Soleimani wouldn’t trigger World War Three, and right that Biden-era lax sanctions enforcement would enable an Iranian oil revenue surge. But his core strategic prediction—that maximum pressure would bring Iran back to the negotiating table for a better deal—has not materialized. A 2020 FDD paper co-authored with Ben Taleblu advocated a combined strategy of “maximum pressure” and “maximum support” to empower the Iranian people to dismantle the regime. That, too, has not happened. His accuracy: **63.9%**.

Sources: Dubowitz, M. and Ben Taleblu, B., “Two Years On, the Trump Administration’s Iran Policy Continues to Make Sense,” FDD, May 7, 2020. Also: Dubowitz, M., “Iran’s Nuclear Disarmament,” FDD Monograph, March 2025.

* * *

V. The Scorecard

When you lay all fifteen experts side by side, a clear pattern emerges. The table below ranks every analyst by their empirical track record—the percentage of their predictions that turned out to be correct, scored on a 0–1 scale from FALSE to TRUE.

Expert	Affiliation	Camp	N	Accuracy	Model Weight
John Mearsheimer	U. Chicago (Realist)	Dovish	8	86.1%	Very High
Ali Vaez	Intl Crisis Group	Dovish	6	82.1%	Very High
Stephen Kinzer	Boston University	Dovish	7	78.1%	High
Narges Bajoghli	Johns Hopkins	Dovish	5	75.0%	High
Behnam Ben Taleblu	FDD Iran Program	Hawkish	7	71.9%	High
Jeffrey Sachs	Columbia / UN	Dovish	8	66.7%	Medium-High
Ray Takeyh	CFR	Hawkish	8	66.7%	Medium-High
Jonathan Schanzer	FDD	Hawkish	7	65.6%	Medium-High

Mark Dubowitz	FDD (CEO)	Hawkish	8	63.9%	Medium
Alastair Crooke	Conflicts Forum	Dovish	8	61.1%	Medium
George Friedman	Geopolitical Futures	Other	9	57.5%	Medium
Michael Rubin	AEI	Hawkish	8	50.0%	Low
Scott Ritter	Ex-UNSCOM	Dovish	9	37.5%	Very Low
Reuel Marc Gerecht	FDD / Ex-CIA	Hawkish	8	33.3%	Very Low
Stefan Molyneux	Independent	Other	7	12.5%	Near Zero

Several patterns jump out. First, the top three experts are all doves—but the fourth-ranked hawk (Ben Taleblu, 71.9%) outperforms several doves. Second, the Iraq War is the great dividing line: hawks who predicted Iraq would be a success (Gerecht, Rubin) carry that anchor forever in the data, while hawks who focused on Iran-specific technical analysis (Ben Taleblu, Takeyh) have strong records. Third, ideology alone doesn’t predict accuracy. The worst performer in the dove camp (Ritter, 37.5%) and the worst in the hawk camp (Gerecht, 33.3%) are nearly identical in their unreliability.

Averaging across camps, the doves score 69.5% and the hawks score 58.6%. That’s a meaningful gap, but not an enormous one, and it’s driven heavily by the Iraq-era predictions. On Iran-specific questions—nuclear deal sunset provisions, missile capabilities, proxy networks—the hawks often have sharp insights. The data suggests we should not dismiss either camp wholesale. We should listen most carefully to the people with the best records, regardless of which tribe they belong to.

* * *

VI. Building the Forecast

With fifteen scored experts in hand, we built a Bayesian forecasting model. The approach is grounded in a simple insight from the academic literature on prediction aggregation: people who have been right before are more likely to be right again, and their opinions should count for more.

For each expert, we constructed a statistical profile of their reliability using a Beta distribution—a standard tool in Bayesian analysis that captures both how accurate

they've been and how confident we should be in that estimate given the number of predictions we've scored. Mearsheimer, with eight predictions at 86% accuracy, gets a tight distribution centered high. Bajoghli, with only five predictions at 75%, gets a wider distribution—we're less certain about her true accuracy because we have less data.

We then asked: based on each expert's published analysis, what probability would they assign to four possible outcomes of a US-Israel military campaign against Iran?

Scenario 1: Quick Win / Regime Change. Air strikes destroy nuclear and military infrastructure; the regime collapses or capitulates within months. The optimistic hawk scenario.

Scenario 2: Prolonged Quagmire. Strikes trigger an extended campaign of Iranian retaliation, proxy warfare, Strait of Hormuz disruptions, and ongoing tit-for-tat escalation lasting years. The Iraq/Afghanistan pattern.

Scenario 3: Stalemate / De-escalation. Initial strikes are followed by back-channel diplomacy, a face-saving off-ramp, and a return to uneasy deterrence. Nobody wins decisively.

Scenario 4: Catastrophic Spread. The conflict spirals into a wider regional war involving Hezbollah, Gulf states, or direct great-power confrontation, potentially crossing the nuclear threshold.

Each expert's scenario probabilities were inferred from their writings, testimonies, and public statements—not surveyed directly. For example, Takeyh's published position that air power can damage Iranian programs but that the regime will survive and rebuild translates to moderate quagmire probability (42%) with low quick-win probability (12%). Dubowitz's more optimistic stance on regime vulnerability and sanctions effectiveness translates to the highest quick-win estimate in the panel (30%).

We then ran 100,000 Monte Carlo simulations. In each draw, we sampled each expert's reliability from their statistical profile, sampled their scenario predictions, and computed a weighted consensus. Experts with higher track records pulled the result toward their views; experts with lower track records had less influence. Crucially, we also accounted for correlations between experts: the four FDD analysts (Dubowitz, Gerech, Ben Taleblu, Schanzer) share institutional frameworks, so their opinions are partially redundant. We penalized this overlap using a correlation

matrix, reducing the effective number of independent voices from fifteen to about ten.

* * *

VII. The Verdict

The model’s output is stark. Across 100,000 simulations, weighting each expert by their empirical track record and correcting for institutional correlations, the consensus forecast is:

Scenario	Probability	80% Credible Interval	95% Credible Interval
Prolonged Quagmire	52.3%	45.1% – 59.7%	41.2% – 63.4%
Stalemate / De-escalation	21.9%	15.7% – 27.2%	13.5% – 31.1%
Catastrophic Spread	16.6%	11.2% – 21.2%	9.2% – 24.5%
Quick Win / Regime Change	9.2%	5.4% – 12.2%	4.3% – 14.8%

A prolonged quagmire is the most likely outcome, with a probability of 52.3%. This is true not because we stacked the panel with doves, but because even the most accurate hawks agree on the underlying dynamics. Takeyh, the most sophisticated hawkish analyst in our panel, explicitly argues that the Iranian regime is too multi-layered to be toppled by air power alone. Ben Taleblu, the highest-scoring hawk, has spent years documenting Iran’s missile capabilities and drone programs—the very systems that would enable prolonged retaliation.

Quick regime change has only a 9.2% probability. This is the finding that may surprise some readers. Even with six hawkish analysts added specifically to ensure ideological balance, the model assigns less than a 10% chance to the scenario where strikes succeed quickly and the regime falls. The reason is simple: the experts who assign the highest probabilities to quick victory (Dubowitz at 30%, Gerecht at 25%) also have lower track records, so their views get discounted. Meanwhile, the experts with the best track records—including hawks like Takeyh and Ben Taleblu—are explicit that bombing alone won’t do it.

There is a roughly 1-in-5 chance (21.9%) that the conflict reaches some kind of stalemate or de-escalation—a scenario where both sides find an off-ramp, perhaps through back-channel diplomacy. And there is a roughly 1-in-6 chance (16.6%) of catastrophic spread into a wider regional war.

* * *

VIII. Why Quagmire Dominates

The convergence across ideological lines on the quagmire scenario is the most important finding in this analysis. It is driven by five structural factors that both hawks and doves acknowledge, even if they disagree on what to do about them.

Geography. Iran is 1.6 million square kilometers—roughly four times the size of Iraq—with 87 million people and terrain ranging from mountain ranges to urban megacities. George Friedman of Geopolitical Futures has long argued that Iran’s physical geography makes sustained military operations extraordinarily difficult.

Regime depth. Both Takeyh and Narges Bajoghli, a Johns Hopkins ethnographer who has conducted years of fieldwork inside Iran, describe a regime that is multi-layered and deeply embedded. The IRGC is not just a military organization; it controls perhaps a third of Iran’s economy and has penetrated every institution. Killing Khamenei—as Takeyh has argued—would create confusion but not necessarily collapse, because succession mechanisms are already in place.

Historical pattern. Mearsheimer, Sachs, and Kinzer all point to the same template: Iraq and Afghanistan. In both cases, initial military operations succeeded quickly, but the aftermath devolved into years-long insurgency and state failure. Mearsheimer has argued that the early Afghan campaign created a dangerous illusion—that air power and special forces could topple regimes cheaply—which directly led to the disastrous decision to invade Iraq.

Source: *Mearsheimer, J.J., lecture at Yale’s MacMillan Center, “Liberal Ideals and International Realities,” 2017. Also: Kinzer, S., Overthrow: America’s Century of Regime Change, Times Books, 2006.*

Proxy networks. Jonathan Schanzer of FDD has documented Iran’s “ring of fire” strategy—using Hezbollah, the Houthis, Iraqi militias, and other proxies to create pressure on Israel and US interests across multiple fronts simultaneously. Although Israeli strikes in 2024 significantly degraded this network (Schanzer himself

acknowledges the axis of resistance is “weakened but not eliminated”), the infrastructure for proxy retaliation still exists.

Economic resilience. Both Takeyh and, surprisingly, Dubowitz agree that Iran’s “resistance economy” has proven more durable than expected. Despite crushing sanctions, Iran has found workarounds through Chinese oil purchases, cryptocurrency, and intermediary networks. The regime’s ability to absorb economic punishment without collapsing is a key factor in the quagmire scenario.

* * *

IX. What This Means

This analysis does not predict the future with certainty. No model can. What it does is aggregate the best available expert judgment, weighted by empirical track records rather than credentials, ideology, or media prominence. The result should give pause to anyone arguing confidently that military action against Iran will produce a quick, clean outcome.

The most striking feature of the data is how robust the quagmire finding is across model specifications. When we started with just four experts, quagmire was the modal outcome at 53%. When we expanded to ten experts (all on the dovish side), it rose to 57%. When we added six hawks specifically to challenge this result, it settled at 52%. The number barely moved—because even the hawks, when you read their detailed analysis rather than their op-ed headlines, acknowledge the same structural dynamics.

If anything, the hawks have made the model’s uncertainty ranges more useful. They shifted the quick-win probability from 5% up to 9%—still low, but no longer negligible. They reduced the catastrophe estimate from 19% to 17%. And they boosted the stalemate/de-escalation scenario from 19% to 22%, reflecting their view that a negotiated off-ramp may be more available than doves assume.

The lesson here is not that doves are right and hawks are wrong. It is that *the analysts with the best track records—regardless of ideology—converge on a common picture*: military action against Iran is far more likely to produce a prolonged, costly entanglement than a swift resolution. The disagreement is about whether that risk is worth taking, not about whether the risk exists.

* * *

X. Methodology Note

This analysis scored 113 predictions across 15 experts. Predictions were graded TRUE (1.0), MOSTLY TRUE (0.75), PARTIAL (0.5), MOSTLY FALSE (0.25), or FALSE (0.0) based on publicly verifiable outcomes. Reliability posteriors used Jeffreys prior Beta(0.5, 0.5). Scenario probabilities were modeled as Dirichlet distributions with concentration parameters reflecting expert specificity.

Aggregation used three methods: linear pool (reliability-weighted average), logarithmic pool (geometric weighted average), and an extremized correlation-adjusted estimate following the methodology of Satopaa et al. (2014). The reported results use the extremized method. A 15×15 correlation matrix captured institutional overlaps (FDD cluster: 0.45–0.60; realist cluster: 0.40–0.50). Eigenvalue decomposition yielded 10.13 effective independent experts and an extremizing factor of $d = 0.675$.

All simulations used $N = 100,000$ Monte Carlo draws with random seed 42 for reproducibility. Uncertainty intervals are highest density intervals (HDIs) computed directly from the posterior distribution, not normal approximations.

Limitations: Prediction scoring involves subjective judgment. Scenario probabilities are inferred from published positions, not directly elicited. The correlation matrix is qualitatively constructed. Several experts' track records are dominated by Iraq-era predictions, which may not reflect their current analytical quality. This is a living-conflict analysis; ground truth has not been determined.