



[RUSSO-UKRAINIAN WAR](#)   [ABOUT](#)  
[SIGN IN](#)   [BECOME A MEMBER](#)   



National security.  
For insiders. By insiders.

# DID SINO-AMERICAN RELATIONS HAVE TO DETERIORATE? A BETTER WAY OF DOING COUNTERFACTUAL THOUGHT EXPERIMENTS

J. PETER SCOBLIC, CHRISTOPHER KARVETSKI, AND PHILIP E. TETLOCK   JULY 16, 2021  
COMMENTARY



## What went wrong with Sino-American relations?

American analysts and policymakers agree that *something* did. As James Steinberg, the former deputy secretary of state, wrote last year, “There are few things that Democrats and Republicans in Washington agree on these days — but policymakers from both parties are virtually unanimous in the view that Sino-American relations have taken a dramatic turn for the worse in recent years.” Quoting President Barack Obama’s and President Donald Trump’s national security strategies, he noted that in just the past decade, a rapidly developing “partnership” has degenerated to “a geopolitical competition between free and repressive visions of world order.”

More recently, the Biden administration’s interim strategic guidance characterized China as a “threat” and described the U.S. relationship with a “more assertive and authoritarian” China as one of “growing rivalry.” Or, as Kurt Campbell, the White House’s top Asia official, put it: “The period that was broadly described as ‘engagement’ has come to an end.” In June, Secretary of Defense Lloyd Austin directed the Pentagon to speed its response to China’s military buildup, and the Senate approved \$250 billion in spending to increase American technological competitiveness. Sen. Chuck Schumer, who co-sponsored the bipartisan bill, said, “If we do nothing, our days as the dominant superpower may be ending.”

BECOME A MEMBER

But when it comes to assigning blame for the deterioration from cooperation to conflict, opinions vary, and determining who is right is difficult. We cannot rerun history to isolate causal variables, as we would in a laboratory experiment. What's more, scholars have shown that foreign policy experts often forecast the future poorly and misjudge what is important in the present. So, how to rate their analyses of the past?

We propose a novel approach to assessing the importance of past events that begins by distilling clashing schools of thought from experts, summarizes their views of the past and future, and then translates those views into a "Bayesian net," an analytic aid we use to quantitatively capture each school's perspective on America's biggest policy blunders and their aftermaths.

Analysts want to do work that's relevant for policymakers, and our approach has the advantage of, among other things, providing probabilistic assessments of the sort that Sherman Kent, the legendary CIA analyst, longed to provide and that officials like Henry Kissinger longed to receive. As national security adviser, Kissinger once reportedly told Andrew Marshall that "top level policy making involves making complicated bets about the future" and that he would prefer intelligence estimates to provide "the betting odds."

Our pilot effort not only provides such odds, but it also reveals surprising consensus across schools of thought that a number of ostensible "turning points" in U.S.-Chinese relations were not turning points at all. Even if certain events had transpired differently, expert assessments of China's

current intentions and capabilities would have changed little, and even when they do, the probabilities they assign to future events (e.g., a Chinese invasion of Taiwan) don't budge much.

The implications are tantalizing, not only for U.S. relations with China, but also for American domestic politics. If taking a different course of action in the past would have barely changed assessments of the present or forecasts of the future, then some policies the United States has pursued may have been needlessly provocative or pointlessly conciliatory. This is not an argument for appeasing or antagonizing China but rather an argument that, before Washington goes to the mat over an issue, it ought to calculate how much it really matters.

The same is true within the Beltway. Even fractious partisans should at least be able to agree on the value of discovering which of their disagreements matter. If opposing policy recommendations lead to roughly the same outcomes, why fight over them — or aggravate an already polarized climate of opinion? Why not refocus the conversation on overlooked areas of agreement? We see the proposed method of doing thought experiments as a powerful tool for disrupting stalemated debates — and as worth institutionalizing whenever lives depend on drawing the right lessons from history.

## **Understanding the Collapse of U.S.-Chinese Relations**

Over the past three decades, China's development has not proceeded as U.S. advocates of engagement once hoped. Political liberalization has not accompanied economic liberalization. Instead, most notably under Xi

Jinping, the Chinese regime has become more authoritarian, suppressing dissent among party members, cracking down on critics in Hong Kong, and detaining Uyghurs in a network of internment camps. Yet, since the end of the Cold War, its economy has surged, growing from approximately 7 percent of U.S. gross domestic product in 1990 to some 70 percent today. Meanwhile, Beijing's defense spending and its military assertiveness in the South China Sea, on its border with India, and even in the Arctic — combined with increasingly advanced technological capabilities — seem threatening to many, if not most, American analysts. Optimism about the near-term future of U.S.-Chinese relations is hard to find.

Although one thorough study suggests that Chinese political liberalization was always an aspiration rather than an assumption — and engagement had its successes (e.g., China's accession to the Nuclear Nonproliferation Treaty) — many experts, both erstwhile proponents of engagement and longtime skeptics, believe that U.S. policy failed. President Bill Clinton's stated belief that economic liberalization would at least hasten political liberalization did not come to pass. High-profile admissions of failure have accompanied a wave of recriminations, from those who decry past naivete and, conversely, those who lament unnecessary provocation.

The implication of many of these critiques is that, if only the United States had done things differently, it would find itself in a better position today. Although some argue that the United States has long overestimated its sway over China and that competition and even conflict were inevitable as China grew stronger, others point to specific decisions: President George H.W. Bush's conciliatory stand after the Tiananmen Square massacre, Congress'

push to grant Taiwanese President Chen Shui-bian a visa in 1995, Clinton's decision to send aircraft carriers to the Taiwan Strait in 1996, U.S. support for China's admission to the World Trade Organization, the Obama administration's "even-handed" treatment of China's dispute with the Philippines over the Scarborough Shoal. The list of ostensible historical hinge points is long.

If policymakers are to improve policy, they need to learn from history. However, the uniqueness of historical events makes it difficult to learn from experience, which is one reason expertise confers disconcertingly little advantage in geopolitical forecasting. Even our ability to identify what is important in the present is limited. A study by Joseph Risi and colleagues found that, in their cables to Washington, U.S. diplomats in the 1970s often missed crucial developments while assigning high importance to events that, with the benefit of hindsight, mattered little. Sometimes it is obvious that we are witnessing history. Often it is not.

Of course, we have more information about the past than we do about the present. Nevertheless, learning from history is difficult because it relies upon our ability to assess counterfactuals. We can ascertain the causal impact of event X only if we explore what the world would look like had event X *not* happened. Unfortunately, we cannot rerun history under different conditions. Instead, we can only infer whether a particular event made a difference by mentally subtracting it from the historical timeline.

The problem is, how do we judge the quality of such reasoning? How do we judge the accuracy of our imaginations?

Because it is not possible to answer such questions directly, we used a proxy to measure the importance of historical events, asking: How much would a change to history affect analysts' assessment of the present? This does not get us to the truth of what would have happened in an alternative reality, but it gets at what policymakers *believe* to be true. Lacking a gold standard for evaluating counterfactuals, we settle for a silver one.

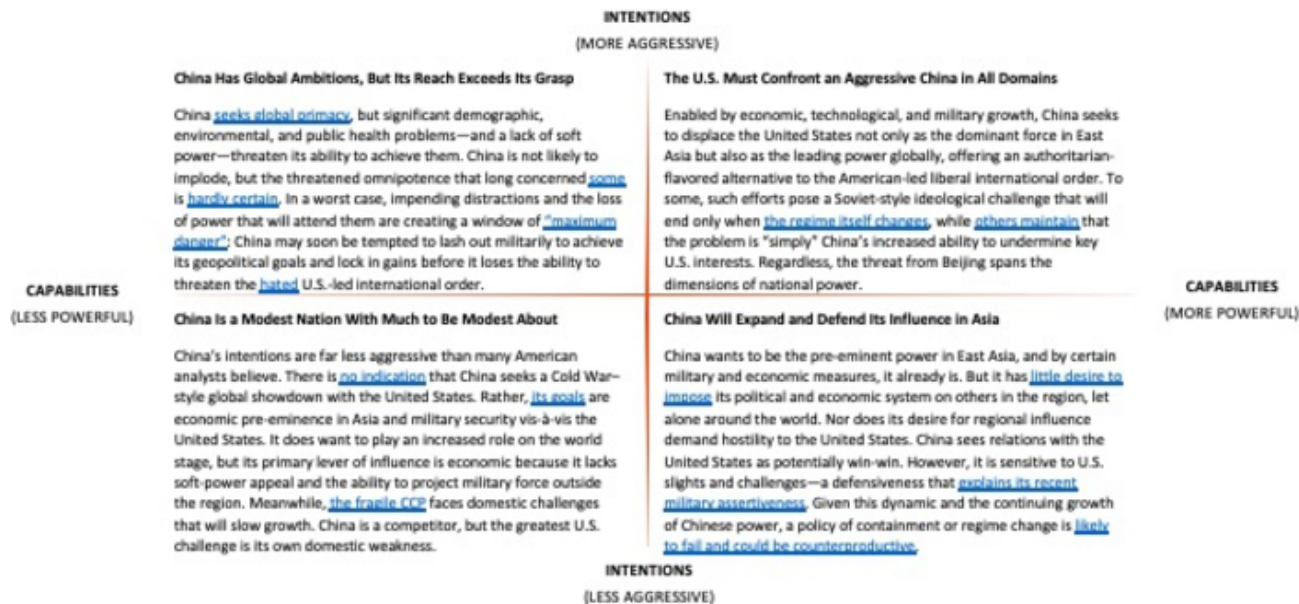
## Understanding the Landscape

To assess how policy analysts update their beliefs in the face of counterfactual changes, we first had to establish what those beliefs are — and they are many and varied. Although few see China as a status quo power, there is disagreement about the audacity of its intentions, the strength of its capabilities, and the inevitability of its rise. In addition, one's general foreign policy orientation may not predict one's stance toward China. The range of positions is too nuanced to be accommodated by a crude hawk/dove divide. We required a more fulsome ornithology. At the same time, any representation would involve simplification.

So, we began with a perspective-taking exercise. Setting aside our own beliefs about China, we interviewed 15 experts on U.S.-Chinese relations and canvassed the research and writing of many more. Although there are certainly other ways to capture the universe of perspectives on bilateral ties, our exploratory work suggested that the best approach was to construct a 2×2 matrix oriented around the classic variables of intentions and capabilities. This allowed us to distill a wide swath of opinion into four schools, which we did — first into 800-word essays and then into 80-word

gists of those essays. This exercise was useful in and of itself, mapping perspectives and disaggregating differences, while also identifying points of agreement.

*Image 1: U.S. Perspectives on China*



*Source: Image generated by the authors. This model is not a taxonomy of preferred U.S. policies toward China because, within each school of thought, we find proponents of distinct blends of confrontation, competition, and cooperation. Policy variance seems to depend on functional priority (e.g., military, economic, environmental), philosophical orientation (e.g., zero- vs. positive-sum), and organizational role (i.e., being “tough on China” can be politically advantageous). China may seem more menacing than it did in the 1990s, but there is no consensus that the correct approach is adversarial. What’s more, this typology is both U.S.-centric and U.S.-*



*ignorant: It characterizes only Washington's views of what is, after all, a bilateral relationship, and it does not fully account for the influence that Washington has over the relationship.*

Next, we took a methodological leap, translating the qualitative beliefs of each perspective into quantitative estimates of the confidence that analysts have in their opinions. For example, we estimated that members of the upper-right quadrant see an 80 percent probability that their view of China as “more aggressive” is correct. By contrast, members of the lower-left quadrant think there is only a 30 percent chance that China is “more aggressive.” A more nuanced interpretation is to view the matrix as a continuous, two-dimensional space — as opposed to a set of four discrete scenarios that are logically exhaustive and mutually exclusive — such that the average member of the upper-right quadrant, asked to rate China’s aggressiveness on a scale of zero to 10, would assign China an “eight,” whereas the average member of the lower right would give it only a “three.”

After quantifying the beliefs of experts in each quadrant, we looked at five possible futures, reflecting scenarios that came up in our research: China lashes out militarily in the next 10 years; China creates a gaping technological advantage over the United States; China’s military strength equals that of the United States; China experiences unexpected economic problems; and China and the United States cooperate on shared interests to their mutual benefit. We then estimated the probabilities that members of each school of thought would assign these futures. So, for example,

members of the lower-left quadrant would put only a 10 percent probability on China achieving military parity, a sign that they perceive less threat from Chinese capabilities, while members of the upper-right quadrant would assign that future a 70 percent probability, reflecting their emphasis on China's growing investment in its armed forces.

Finally, we examined four counterfactual scenarios that interviewees raised: What would have happened if the United States had more consistently pressured Taiwan to increase its defense preparedness? What would have happened if the United States had taken a harder line with China during the Scarborough Shoal controversy? What would have happened if the United States had not withdrawn from the Trans-Pacific Partnership in 2017? What would have happened if the United States had not invaded Afghanistan and Iraq in the wake of the Sept. 11 attacks (a change that would have enabled an earlier and more substantial pivot to Asia)? We then revised our estimates of each school of thought's beliefs about China's present intentions and capabilities and calculated the effect those changes would have on experts' forecasts of the future.

Statistically savvy readers will recognize this as an exercise in Bayesian reasoning: We updated the probabilities accorded future events in response to learning that history unfolded differently from how it actually did. By systematically mapping each school's views of China in the Intentions x Capabilities space to account for each counterfactual, we were able to estimate the current and future ramifications of going down alternative policy paths in the past.

## What We Found and What It Means

We estimated five forecasts for four schools of thought and then reconsidered those initial 20 probabilities in light of four different counterfactuals and their impact on how the schools of thought estimated China's intentions and capabilities, producing 80 new probabilistic forecasts to compare with the baseline world. There were significant differences in what each school forecast based on conditions today, but the differences were most significant with respect to military issues. There was greater agreement on China's economic and technological aims.

Changing the starting conditions almost always resulted in *some* movement in expected assessments of China's intentions and capabilities, but even in extreme circumstances those changes did not dramatically affect forecasts of the future. For example, had the United States not invaded Iraq and Afghanistan, we estimate that U.S. policymakers worried about Chinese aggression now would be less concerned about China's relative military power. However, this was the counterfactual requiring the most significant change to history, and even it had only modest effects on projections of, say, future bilateral cooperation. In sum, the logical implications remained largely the same — no likely event (one in which the probability was way above 50 percent) became unlikely (in which the probability was way below 50 percent) — and any change induced by the counterfactuals was almost always within the margin of error of the original forecasts.

Perhaps the most counterintuitive result of this effort was finding that perspective-taking reveals things about clashing belief systems that may not have been fully apparent to the believers themselves. Quantifying qualitative statements of perspectives yielded surprising results. Certain widely cited pivots in the relationship had decidedly less impact than usually believed.

A particularly thought-provoking example given recent attention to the possibility of a Chinese attempt to seize the island is the degree to which greater Taiwanese defense efforts would have influenced the present. On the one hand, we find that members of each of the top two quadrants — those most worried about China's intentions — would, in fact, be placated to some degree. On the other hand, the change would not be dramatic enough to convince them that China's ambitions are merely defensive, and it would barely budge their assessment of China's capabilities.

This is unexpected, but even more surprising is what happened — or, rather, what didn't — when we applied that counterfactual to the most germane future scenario: China lashing out militarily. Even if we increased the resources Taiwan had spent preparing to defend against a Chinese attack, forecasts that China might lash out militarily in the next 10 years did not change markedly. Even among members of the upper-right quadrant, who are the staunchest proponents of greater Taiwanese military spending and U.S. support for the island, the prediction that China might lash out militarily decreased merely 7 percentage points — from 39 percent to 32 percent. In fact, even if the United States had not fought the wars in Afghanistan and Iraq — thereby diverting resources from the Indo-Pacific

— the estimated likelihood of China attempting to seize Taiwan dropped only 3 points, highlighting a dynamic familiar to rigorous Bayesians: Seemingly dispositive changes to the current environment do not necessarily change conditional forecasts as dramatically as one might expect. Interestingly, the past that hawks would have preferred does little to change their beliefs about the future.

*Image 2: Will China Lash Out?*

*Source: Image generated by the authors.*

These results — and we should stress that this effort is only a pilot program — are counterintuitive. So, we must ask what is happening. We see five possibilities:

First, the method itself has no merit. It is meaningless or, worse, misleading to attach quantitative probabilities to qualitative opinions and to carve overlapping, fuzzy-set schools of thought into exclusive and exhaustive categories. To these critics, we concede there will always be a tension between the stark simplifications demanded by the probability calculus and the messiness of political debate. Purists will always have the option of vetoing efforts such as ours. But purism carries opportunity costs. Imperfect, rough-and-ready quantification delivers something that policymakers value: better odds estimates. For instance, in-depth research on geopolitical forecasting that Philip Tetlock and Barbara Mellers led during the multiyear tournaments sponsored by the Intelligence Advanced Research Projects Activity showed that top forecasters could place surprisingly well calibrated probabilities on future events. And more recent research on forecasting the impact of counterfactuals in simulated worlds — in which we really can rerun history and assess probability distributions of possible worlds — has reinforced that conclusion.

Second, the method has merit, but the exercise requires doing something difficult: assigning quantified values to qualitative opinions. Remember that we are not offering our own forecasts but estimating the beliefs of experts from their statements and writings. We may have misinterpreted the probabilities implicit in their analyses. Scholars have repeatedly established that people mean different things by different probabilistic

phrases (e.g., “likely,” “almost certain,” or “nearly impossible”) so it is reasonable to believe we made some mistakes. For example, take Oriana Skylar Mastro’s forecast about the possibility of China lashing out: “Although a Chinese invasion of Taiwan may not be imminent, for the first time in three decades, it is time to take seriously the possibility that China could soon use force to end its almost century-long civil war.” How should we quantify the probabilities implicit in such a statement? The remedy here is closer collaboration with proponents of the perspectives (if we want to know what they truly think) and recruiting larger panels of readers (if we want a less noisy gauge of what the authors want us to think). This objection is not easy to resolve, but it is resolvable.

Third, the method has merit, but we tested the “wrong” counterfactuals. In other words, we did not examine those events that had the most impact on Sino-American relations, from the 1996 military standoff in the Taiwan Strait, to Chinese accession to the World Trade Organization, to the 1999 U.S. bombing of the Chinese Embassy in Belgrade, to, more recently, the trade war that Trump initiated. In this view, we obtained the results we did because we chose only milquetoast points in history. We wholeheartedly agree that we should test additional counterfactuals, including by adding events that did not happen and not just subtracting events that did. We chose the counterfactuals we did because our experts had focused on them. But if this critique is valid — if the counterfactuals we investigated were obviously non-determinative — then why was there any dispute over them? If, for example, no one expected U.S. withdrawal from the Trans-

Pacific Partnership to impact the U.S. position vis-à-vis China, then why the hubbub about ostensible benefits to China? To the extent that there was controversy over issues that mattered little, our point is only stronger.

Fourth, the method has merit and the model is accurate, but analysts might reject its implications because — odd though it sounds — they disagree with themselves. Almost seven decades of research, dating back to an influential book by psychologist Paul Meehl in 1954, has repeatedly demonstrated that simple statistical models can outperform expert judgment. We should not therefore rule out the possibility that our simple Bayesian models of schools of thought may produce better judgments than can the experts themselves. One reason is that expert judgments are often noisy and contain a lot of random variation whereas computers are at least relentlessly consistent. They never get tired or moody or euphoric or drunk. Another reason is that people are not natural statisticians, and their judgments of how much belief updating is appropriate in response to what-if scenarios are systematically biased to justify current policy stands and their positions in social networks.

In this light, we should raise the final possibility: Our preliminary results contain a considerable kernel of truth. Oft-considered “pivots” are just not that pivotal. In their analyses of “what went wrong?” experts on Sino-American relations have been considering not hinge points in history but rather “red herring hinges” — ostensibly valuable clues that actually point in the wrong direction. The question then would become why are so many



analysts focused on the wrong things? We would look for answers by turning to two of the greatest psychologists of the last century: Daniel Kahneman and Leon Festinger. And the answers are sobering.

Kahneman maintains that our imaginations are biased. Asked to conjure counterfactual, what-if scenarios, people are prone to change abnormal states to normality: to undo blunders we could have avoided if only those in charge had been sensible enough to listen to them. By that psycho-logic, experts predictably looked to episodes in the history of U.S.-Chinese relations for reasons as to why relations soured. That does not mean that there were not pivotal historical moments, but these “hidden hinges” may be harder to find, lurking either in less proximate phenomena or in seemingly mundane changes whose effects ripple outward in ways that obscure causality.

The answer that emerges from Festinger’s theory of cognitive dissonance is that we are wired to prioritize justifying our public attitudes over discovering the truth. The more one favors a hawkish stance toward China, the greater the dissonance-reduction rewards for endorsing counterfactuals in which more assertive policy responses at key junctures rerouted history onto a more positive trajectory. And, of course, the flipside proposition applies to dovish observers who favor accommodative policies. The more explicitly one lays out the logic of one’s causal claims before the fact, the harder it becomes to rationalize away awkward facts that arise later. It reveals what proponents of clashing schools of thought really think, not what they want us to think.

## Implications

We see perspective-taking — the initial part of the exercise — as having value in itself. Charting the dimensions of opinion provides a path to cognitive empathy, to making the beliefs of one's opponents more understandable, if not necessarily more palatable. To put it in more Machiavellian terms, the better one understands one's opponents, the better one can anticipate their actions. Empathy is therefore strategically useful. Relatedly, research shows that one of the most reliable indicators of geopolitical forecasting accuracy is active open-mindedness, which is another way of saying that the ability to entertain multiple points of view improves judgment. So, although we happen to believe that developing a deeper and richer understanding of the world has intrinsic value, we can also say that it has functional utility.

More significantly, however, the implications of the Bayesian net are far-reaching — not only to historians but to policymakers and politicians. When we lay out our beliefs so unusually openly, we may discover policy debates in which we are arguing about the wrong things — indeed, we may not need to argue at all! If even a major change to the past, such as the subtraction of the 9/11 attacks, does little to impact expectations of the future, one has to ask which, if any, policy decisions matter. At the very least, the bar for what constitutes “significant” may be much higher than normally assumed. As Steinberg summarized his article on the collapse of relations between Beijing and Washington: “My point is that even if

decisions were wrong, were they consequentially wrong? That is, would a different decision likely have made a significant impact on the overall trajectory of the relationship? Not really.”

Our preliminary results support this observation. Historical-counterfactual interventions explored thus far do not have big effects on the Bayesian-net predictions of the major schools of thought on Sino-American ties. At least within the last 30 years, there is considerable agreement across clashing schools of thought that it is hard to find politically plausible U.S. policy interventions that would have altered the trajectory of the relationship.

Among other things, this suggests that political polarization might be reduced because many past disagreements may have been unnaturally and unnecessarily exaggerated. Policy differences may not have been as determinative as assumed — and therefore need not have been as divisive. Going forward, we can head off this dynamic through a fairly simple iterative process: People input their beliefs, machines output the logical implications of those beliefs, and people then decide whether to accept those implications or rethink their belief inputs.

Of course, political polarization is propelled by more than policy differences. The primordial forces of identity politics also come into play. Polarization is affective as well as cognitive — opinions are deeply felt as well as strongly argued — and we are not so naïve as to suppose that mere exposure to different viewpoints can erase it. At the same time, we are not so jaded as to abandon all meliorism and conclude that identifying areas of agreement is useless.

This hypothesis remains to be tested, but recent scholarship provides reason for hope. Researchers have shown that political in-groups tend to exaggerate the negativity that out-groups feel toward them (e.g., Democrats believe that Republicans think worse of them than Republicans actually do). Greater inaccuracy is associated with greater affective polarization toward the out-group, but, importantly, informing people of these inaccurate beliefs *reduces* affective polarization (a finding that generalizes to 25 countries). In short, accurately understanding what the other side believes matters. It is depolarizing. Perhaps helping each side understand how the other draws lessons from history will matter, too. We can make conversations smarter, faster if we put our counterfactual cards on the table — and explore the arguments that unite or divide us.

## BECOME A MEMBER

*J. Peter Scoblic is co-founder of Event Horizon Strategies, a senior fellow in the International Security program at New America, and a fellow at Harvard's Kennedy School. Christopher Karvetski is a senior data and decision scientist at Good Judgment Inc. Philip E. Tetlock is Leonore Annenberg University Professor at the University of Pennsylvania, co-founder of Good Judgment Inc, and co-author of Superforecasting: The Art and Science of Prediction.*

Image: Xinhua (Photo by Bao Dandan)

COMMENTARY

# SIGN UP FOR OUR NEWSLETTER

*youremail@domain.com*

**SUBMIT**

**ABOUT**

**MEMBERS**

**PODCASTS**

**GET MORE WAR ON THE ROCKS**

**BECOME A MEMBER**

**MISSION**

**JOIN**

**WOTR**

SUPPORT OUR MISSION AND GET EXCLUSIVE CONTENT

**PEOPLE**

**WAR HALL**

**NET ASSESSMENT**

**FOUNDER'S CLUB**

**JAW-JAW**

**FOLLOW US**

**CONTACT**

**HORNS OF A DILEMMA**



**NEWSLETTER**

*youremail@domain.com*

**SUBSCRIBE**

SIGNING UP FOR THIS NEWSLETTER MEANS YOU AGREE TO OUR DATA POLICY

[PRIVACY POLICY](#) | [TERMS & CONDITIONS](#) | [SITEMAP](#) |

COPYRIGHT © 2023 METAMORPHIC MEDIA. ALL RIGHTS RESERVED.