The Inversion of Doubt: Public Skepticism Is Concentrated Where Epistemic Warrant Is Highest and Absent Where It Is Lowest

Jeremy McEntire
perardua.dev/research
ORCID: 0009-0006-8008-9587

Working Paper — March 2026

Abstract

Public skepticism is not randomly distributed across scientific domains. This paper argues that it is distributed inversely to where it is epistemically warranted: the scientific claims attracting the most sustained public doubt are among the most empirically robust in the history of inquiry, while claims with documented failure rates of 40 to 97 percent circulate with near-zero friction. This inversion is not the product of individual irrationality. It is structurally produced by three forces: the asymmetry between falsifiability timescales in hard versus soft science, identity-protective cognition that indexes skepticism to perceived threat rather than accuracy, and a selection environment that rewards transmissibility over calibration. The paper synthesizes evidence from Delphi foresight retrospectives, metascientific positive-result analyses, geopolitical forecasting tournaments, clinical reversal epidemiology, and longitudinal public opinion data to document both the inversion and its mechanisms. A secondary finding concerns the measurement infrastructure itself: the absence of dual-cohort survey data measuring public skepticism toward social science claims is not an oversight but a structural consequence of unfalsifiability. Where ground truth is unavailable, the correction machinery never activates, and claims circulate without measurement or challenge. The cross-domain synthesis offered here—arguing that high acceptance rate is evidence of low accuracy above a threshold—has not previously been formalized in the philosophy of science literature, though its component mechanisms have been documented in isolation across metascience, forecasting research, risk psychology, and science communication. A provisional threshold operationalization is offered, proposing that domains meeting two of three measurable criteria—registered report divergence exceeding 20 percentage points, expert calibration gap exceeding 20 percentage points, and falsifiability timescale exceeding a professional generation—warrant presumptive skepticism of public acceptance rates.

Keywords: epistemic warrant, public skepticism, metascience, replication crisis, forecasting accuracy, science communication, motivated reasoning, inversion

1 Introduction

Consider a claim that circulates freely through prestige journalism, documentary media, and policy discourse: that falling birth rates across developed nations represent a civilizational crisis with a straightforward demographic solution. The arithmetic is presented as self-evident. The policy implication follows without argument. No citation is offered. No confidence interval is supplied. The claim is not argued; it is acknowledged, as something any informed person already knows.

The same week, considerable institutional energy is devoted to countering public skepticism about the safety profile of childhood vaccines—a question on which the evidentiary record is among the most extensive in the history of medicine.

This juxtaposition is not unusual. It is the normal condition of public epistemology. The claims attracting the most sustained correction effort, the most active remediation campaigns, and the most voluminous science communication literature are precisely the claims with the strongest empirical foundations: the mechanism of anthropogenic climate change, the fact of biological evolution, the efficacy of vaccines, the heliocentric model of the solar system. Meanwhile, macroeconomic projections, demographic forecasts, sociological predictions about immigration and inequality, and clinical practice guidelines that later prove baseless are absorbed into public discourse and institutional decision-making with minimal friction.

The question this paper asks is whether this pattern is accidental. The evidence suggests it is not. Public skepticism is distributed inversely to epistemic warrant—not randomly, not as a result of ignorance alone, but as a predictable output of structural forces that select for transmissibility over accuracy and that index skepticism to identity threat rather than to the track record of the field making the claim.

This argument requires a preliminary clarification. Epistemic warrant, as used here, does not mean "is this specific claim true." It means: does this field's methodology reliably produce true claims? A domain has high epistemic warrant when its predictions replicate, when its confidence intervals are calibrated, when its expert consensus has a documented track record of being correct, and when errors are visible and correctable on short timescales. By this standard, the hierarchy runs from physics and chemistry at the top through clinical medicine and biology in the middle to macroeconomics, demography, and social forecasting at the bottom. The evidence for this gradient is now extensive and will be reviewed in Section 3.

The cross-domain synthesis offered here—that high acceptance rate can serve as evidence of low accuracy above a threshold—has not previously been formalized in the philosophy of science or science communication literature. The component mechanisms have been documented: Ioannidis on false discovery rates within fields, Fanelli on positive result gradients across the hierarchy, Tetlock on forecasting failure in expert political judgment, Taleb on the insulation of macroe-

conomic experts from accountability, Crichton on the Gell-Mann Amnesia effect, and the affect heuristic literature on inverse risk perception. What has not been done is to synthesize these threads into a single cross-domain argument and draw out its implications for how scientific authority is allocated.

The paper proceeds as follows. Section 2 operationalizes the two key variables—epistemic warrant and public skepticism rate—identifies a structural gap in the measurement of the second, and proposes a threshold operationalization for the central claim. Section 3 documents the epistemic warrant baseline across domains. Section 4 maps the distribution of public skepticism against that baseline and establishes the inversion. Section 5 identifies the three structural forces that produce and sustain it. Section 6 identifies the throwaway claim as the transmission mechanism through which the structural conditions produce high acceptance of low-warrant claims at the level of individual propositions. Section 7 draws out implications for science communication, policy, and the sociology of knowledge. Section 8 addresses limitations and methodological constraints honestly.

2 Operationalizing the Question

Two variables require measurement. The first is epistemic warrant. The second is public skepticism rate. The relationship between them is the central finding.

2.1 Epistemic Warrant

Epistemic warrant, as used here, is the degree to which a claim's acceptance is justified by its field's track record of predictive accuracy, replication rate, and confidence interval calibration. This is a property of methodology, not of individual claims. A domain has high epistemic warrant when independent researchers testing the same hypothesis reach the same result, when expert predictions within the domain are realized at rates substantially better than chance, and when errors are structurally detectable and correctable.

Three tiers are sufficient for the purposes of this analysis.

The first tier covers the hard physical sciences: physics, chemistry, materials science, and related applied engineering domains. These fields are characterized by tight theoretical constraints, low researcher degrees of freedom, high rates of disciplinary consensus, and short feedback loops between prediction and observable outcome. Predictions are falsifiable within years, not decades. The Japan National Institute of Science and Technology Policy (NISTEP) Delphi surveys, which have tracked expert scientific predictions retrospectively since 1971, place overall hard science realization rates at approximately 70 percent, with chemistry and materials science reaching 70 to 80 percent (NISTEP retrospective analyses, surveys 1–5).

The second tier covers clinical medicine and biological science. These fields occupy an intermediate position: more subject to complexity and confounding than physics, but still anchored by randomized controlled trial methodology and the possibility of falsification within human lifespans. As will be discussed in Section 3, the clinical evidence base is substantially less reliable than is commonly assumed, but the correction infrastructure is real and functional.

The third tier covers social science, macroeconomics, demography, and geopolitical forecasting. These fields study open, complex systems where variables cannot be isolated, researcher degrees of freedom are high, feedback loops operate on decadal timescales, and expert consensus has a documented track record of performing near chance on medium-range predictions.

2.2 Public Skepticism Rate

The public skepticism rate is the degree to which a claim is rejected or doubted by the general public, independent of the accuracy of that rejection. This is measurable for hard science claims through longitudinal survey research. The Pew Research Center's 2015 dual-cohort study comparing AAAS scientist beliefs to public beliefs provides the most widely cited empirical anchor: a 33-point gap between scientific consensus and public acceptance of human evolution, a 51-point gap on the safety of genetically modified foods, and a 37-point gap on the mechanism of climate change.

These gaps are well-documented and have generated an extensive science communication literature focused on closing them. What is largely absent from that literature is a parallel measurement for the opposite end of the epistemic warrant spectrum. There are no equivalent dual-cohort surveys asking whether the public accepts macroeconomic growth projections at higher rates than economists believe warranted, or whether demographic collapse narratives are accepted more widely than the evidence supports.

This absence is itself a structural finding and not a coincidence. The survey infrastructure that produces the Pew consensus gap data exists because hard science claims have falsifiable yes-or-no answers that can be posed to respondents and evaluated against expert consensus. Asking a respondent whether they believe humans evolved from earlier species has a correct answer. Asking whether they believe a specific demographic projection will produce economic contraction within thirty years produces a political opinion, not an epistemic measurement, because the claim is not falsifiable on any timescale relevant to a polling instrument.

The measurement infrastructure therefore tracks falsifiability. Where falsifiability is absent, claims circulate without measurement, without correction, and without accountability. This structural asymmetry is one of the mechanisms producing the inversion. It is documented here as a finding rather than treated as a gap to be apologized for.

2.3 The Threshold Problem

The central claim of this paper—that above a threshold, high acceptance rate is evidence of low accuracy—requires at least a provisional operationalization of that threshold. Without one, the claim risks the same unfalsifiability the paper diagnoses in the soft sciences.

The data reviewed in Section 3 provide three convergent indicators that, taken together, identify the boundary between domains where public acceptance tracks accuracy and domains where it decouples.

The first is the registered report divergence. Scheel, Schijen, and Lakens (2021) demonstrated that in psychology, the gap between standard positive result rates (96 percent) and registered report positive result rates (44 percent) is 52 percentage points. This gap measures the magnitude of publication-incentive distortion in a field. In space science, where Fanelli (2010) documents a standard positive result rate of 70.2 percent, the registered report gap is likely small—perhaps 10 to 15 points—because the tight theoretical constraints of the field limit the room for analytic flexibility. A divergence exceeding approximately 20 percentage points indicates that the published literature has substantially departed from ground truth, and that public acceptance of claims from that field carries a correspondingly large risk of accepting distortion as fact.

The second is the expert calibration gap. Moore, Tenney, and Haran (2024) documented a 30-point gap between subjective confidence and empirical accuracy in the Survey of Professional Forecasters. A calibration gap below approximately 10 percentage points—characteristic of weather forecasting and short-range physical predictions—indicates a domain where expert consensus reliably tracks reality. A gap exceeding 20 points indicates a domain where expert confidence has substantially decoupled from predictive accuracy, and where the public's acceptance of expert claims is not justified by the experts' own track record.

The third is the falsifiability timescale itself. When predictions cannot be evaluated against outcomes within a single professional career—roughly 30 years—the error-correction infrastructure that produces calibration in other fields cannot function. Claims from fields operating above this timescale are structurally insulated from the accountability that generates epistemic warrant in fields operating below it.

A domain crosses the threshold when any two of these three conditions are met: a registered report divergence exceeding 20 points, an expert calibration gap exceeding 20 points, or a falsifiability timescale exceeding a professional generation. This is proposed as an operationalization, not a measured constant. It is deliberately conservative: most third-tier fields in the hierarchy described above satisfy all three conditions simultaneously. The threshold is designed to be falsifiable. Any field can demonstrate it operates above the line by conducting registered report studies, submitting to calibration audits, or generating predictions on timescales where outcomes are observable. Fields that have not done so are provisionally unmeasured, and the appropriate epistemic posture

toward their claims is proportional skepticism, not default acceptance.

3 The Epistemic Warrant Baseline

3.1 Hard Science: What the NISTEP Retrospectives Show

Japan's NISTEP Delphi program represents the most sustained longitudinal effort to evaluate the accuracy of expert scientific and technological predictions. Beginning in 1971 and conducted at approximately five-year intervals, the surveys ask panels of experts to forecast when specific scientific and technological developments will be realized. Retrospective evaluations of the first five survey cycles allow direct comparison between predictions and outcomes.

A caveat is necessary before presenting these data. The NISTEP surveys measure technology foresight—when will a capability be realized?—which is a different question from scientific accuracy—is the underlying theory correct? The realization rate is used here as a proxy for hard science reliability, not as a direct measurement of it. The proxy is justified because the hard sciences' direct replication rates are even more favorable: foundational results in physics, chemistry, and materials science replicate at rates approaching unity for well-established findings, and the major theoretical frameworks of these fields (quantum mechanics, general relativity, thermodynamics, molecular biology's central dogma) have survived decades of intensive testing without fundamental revision. The NISTEP data are cited because they offer the only sustained longitudinal dataset with retrospective accuracy evaluation across domains—not because technology foresight is identical to theoretical accuracy, but because it provides a conservative lower bound on hard science epistemic warrant that is still dramatically higher than the corresponding figures for social science.

The aggregate realization rate across all domains in surveys one through five is approximately 70 percent. This figure requires careful interpretation. It does not mean that 30 percent of hard science predictions reflect theoretical errors. The retrospective data distinguishes clearly between different categories of non-realization. In fields closest to fundamental physical limits—elementary particle physics, quantum phenomena, energy infrastructure—the primary causes of failure are technical intractability and cost, not wrong theory. Room-temperature superconductivity was predicted by expert consensus to be realized by 2011. It remains unrealized, and the NISTEP retrospective explicitly attributes the failure to fundamental physical barriers, not to an incorrect theoretical model. Nuclear fusion carries the same signature: the underlying physics is not disputed, but the thermodynamic and capital costs of achieving net positive yield have consistently exceeded expert projections.

This distinction matters for the cross-domain argument. Hard science consensus errors are predominantly errors about timescale and cost, not about the identity of the phenomenon. So-

cial science consensus errors are qualitatively different—errors about whether the predicted effect exists at all.

The realization rate also varies by subdomain in ways that are instructive. Chemistry and materials science, where prediction involves optimizing known physical systems rather than discovering new ones, achieves 70 to 80 percent realization. Applied physics in energy and infrastructure, where thermodynamic and economic constraints interact in ways that expert panels consistently underweight, falls below 60 percent. Technological forecasting also reveals a third failure mode absent from social science: alternative technological paths. Approximately 30 percent of unrealized hard science predictions failed not because the predicted capability was impossible but because a different mechanism achieved it first. The hard science forecaster is often right about the destination and wrong about the road.

3.2 Social Science: The Replication Record

The replication crisis that emerged publicly in the 2010s has produced a detailed empirical record of epistemic warrant across the soft sciences. The picture is not uniform, but the gradient runs consistently in one direction.

In psychology, the Open Science Collaboration's Reproducibility Project attempted to replicate 100 studies published in premier journals. Statistically significant results held in approximately 36 percent of replications, with mean effect sizes roughly half those originally reported. The Many Labs 2 project, testing 28 classic and contemporary findings across 125 samples, produced successful replication in approximately 50 percent of cases. These are the replication rates for published findings that had already survived peer review and editorial selection—the best the literature offers.

In economics, a distinctive signature of data manipulation has been documented at scale. Brodeur, Lé, Sangnier, and Zylberberg analyzed the distribution of 50,078 test statistics from publications in the *American Economic Review*, the *Quarterly Journal of Economics*, and the *Journal of Political Economy* between 2005 and 2011. The distribution should, under honest reporting conditions, follow a smooth continuous curve. Instead, the data reveals a pronounced depression in the region just above the conventional significance threshold of $p = 0.05$, paired with an anomalous spike just below it. This two-peaked distribution is the statistical signature of p-hacking: researchers encountering marginally non-significant results and adjusting their analyses until significance is achieved. The authors estimate that between 10 and 20 percent of marginally rejected tests in elite economics journals are false positives produced by this process.

The macroeconomic forecasting record is available through the Survey of Professional Forecasters, maintained by the Federal Reserve Bank of Philadelphia since 1968. Analysis of this dataset by Moore, Tenney, and Haran (2024) reveals that professional economists reported an av-

erage confidence level of 53 percent in their predictions while achieving correct outcomes only 23 percent of the time—a 30-point gap between subjective certainty and empirical accuracy that persists across decades of measurement.

Philip Tetlock's Expert Political Judgment project tracked 284 recognized professionals across political science, economics, sociology, and intelligence analysis, collecting tens of thousands of quantifiable probabilistic forecasts over approximately fifteen years (1988–2003). Assessed using the Brier score—a strictly proper scoring rule where 0.5 represents random chance performance—the aggregate expert population barely outperformed the dart-throwing chimpanzee baseline. More consequentially, accuracy decays rapidly with time horizon: expert predictions merge statistically with random guessing at approximately the three-to-five year mark and show no evidence of meaningful foresight beyond that.

The Tetlock data also produced a finding directly relevant to the inversion argument: expert media prominence is negatively correlated with predictive accuracy. The correlation between a forecaster's self-reported media contact frequency and their calibration score was $r = -0.12$, and broader fame proxies such as citation impact indices show stronger negative correlations with accuracy (Tetlock, 2005). The forecasters most likely to appear on television and shape public discourse are measurably the least accurate. This is not a coincidence. Media selection operates on narrative clarity and projective confidence, not on calibration track records. Tetlock's fox-hedgehog distinction quantifies the mechanism: the cognitive style associated with confident, narratively coherent prediction (the hedgehog) shows a Cohen's d of 0.42 in the direction of worse calibration compared to the integrative, uncertainty-acknowledging style (the fox). The structural consequence is that the public sphere is systematically populated by the least reliable epistemic actors in any given domain—not because media organizations are malicious, but because the properties that make a forecaster telegenic are the same properties that make a forecaster inaccurate.

3.3 Clinical Medicine: Consensus and Its Reversals

Clinical medicine occupies a complicated position in the epistemic warrant hierarchy. It has the strongest correction infrastructure of any applied field—randomized controlled trials, systematic reviews, the Cochrane Collaboration—but the baseline rate of expert consensus later shown to be wrong is higher than is generally acknowledged.

Herrera-Perez, Haslam, Crain, and colleagues conducted a meta-epidemiological review, published in *eLife* in 2019, that systematically evaluated a decade and a half of publications across three flagship medical journals: *JAMA*, *NEJM*, and *The Lancet*. Of the original articles reviewed, 3,017 reported results of randomized controlled trials that tested an established clinical practice. Of those, 396—approximately 13 percent—reversed the practice under review, demonstrating through

rigorous methodology that an established standard of care was no better than the control condition, or was actively harmful. Cardiovascular medicine accounted for the largest share of reversals at 20 percent of the total, followed by public health interventions at 12 percent and critical care at 11 percent.

The oncology evidence base presents the sharpest example of institutionally protected expert consensus in the absence of supporting data. The National Comprehensive Cancer Network guidelines govern the standard of care for more than 97 percent of cancer patients in the United States. Federal reimbursement law mandates adherence to these guidelines by Medicare and most private insurers. Yet comprehensive audits of the guidelines find that only approximately 6 to 8 percent of therapeutic recommendations are supported by Category 1 evidence—defined as at least one randomized phase 3 trial demonstrating efficacy with uniform panel consensus—depending on the audit and the cancer types included (Poonacha and Go, 2011). For uterine cancer and pancreatic cancer guidelines, the proportion of recommendations backed by Category 1 evidence is zero percent. Every therapeutic pathway recommended for patients with these diseases rests on expert panel consensus applied to lower-tier observational evidence, yet carries the force of federal reimbursement law and the cultural authority of institutional medical consensus.

3.4 The Positive Result Gradient

Fanelli's 2010 analysis of 2,434 published papers across 20 disciplines provides the clearest single snapshot of how epistemic warrant varies across the hierarchy of sciences. Across all disciplines, 84 percent of papers reported positive, hypothesis-confirming results. But the distribution was not uniform and followed the hierarchy with high statistical reliability: Space Science reported positive results at a rate of 70.2 percent, while Psychology and Psychiatry reported positive results at a rate of 91.5 percent. The odds of reporting a positive result were approximately five times higher in psychology than in space science.

The mechanism is researcher degrees of freedom. In fields with tight theoretical constraints and physically determinate outcomes, researchers cannot adjust their analyses until a desired result appears without the adjustment being detectable. In fields where the choice of covariates, sample composition, outcome measures, and analytical models is genuinely discretionary, the same underlying data can be analyzed in dozens of ways, and the version that achieves significance is the version that gets published.

The magnitude of this distortion was directly measured by Scheel, Schijen, and Lakens in 2021, using a natural experiment. Registered Reports are a publication format in which the study design is peer-reviewed and accepted before any data is collected, with the journal committing to publish the results regardless of their direction. In psychology, where standard publication rates show

roughly 96 percent positive results among hypothesis-testing studies, Registered Reports produce positive results at a rate of approximately 44 percent. The 52-point gap between these figures represents the measured size of the distortion introduced by publication incentives in hypothesis-testing psychology specifically. It is not a small effect. It means that the published psychological literature, under standard conditions, is roughly twice as optimistic about the rate of true effects as the underlying phenomena warrant.

The difference between Fanelli's 91.5 percent and Scheel and colleagues' 96 percent reflects methodological scope: Fanelli sampled broadly across paper types within psychology, while Scheel and colleagues restricted their comparison to hypothesis-testing studies, where the distortion is most concentrated. Both figures confirm the same directional finding: the gap between published positive results and actual positive results widens as one moves down the hierarchy of sciences.

4 The Skepticism Distribution

4.1 Where Skepticism Concentrates

Public skepticism in the United States and across Western democracies is concentrated in hard science domains with strong epistemic warrant. The National Science Foundation's Science and Engineering Indicators have documented for decades that approximately one quarter of American adults cannot correctly identify that the Earth orbits the Sun. Pew Research Center data from 2015 shows that 65 percent of the public accepts human evolution, against 98 percent of surveyed AAAS scientists—a 33-point gap. Acceptance of anthropogenic climate change sits at 50 percent among the public against 87 percent among climate scientists. Vaccine confidence shows documented patterns of refusal concentrated in communities with high educational attainment and strong identity-based skepticism of institutional medicine.

These are precisely the domains where the epistemic warrant evidence is strongest. Physics, evolutionary biology, epidemiology, and atmospheric chemistry operate under the first tier described in Section 3: tight theoretical constraints, short feedback loops, high replication rates, and a track record of prediction that spans decades. The historical success of these disciplines in predicting phenomena—from the existence of particles to the trajectory of climate forcing—is not in serious dispute among domain experts.

Yet these are the domains generating active remediation efforts, science communication campaigns, media coverage of denial, and institutional concern about public credulity. The infrastructure for measuring and correcting public skepticism is almost entirely oriented toward them.

4.2 Where Skepticism Is Absent

No equivalent measurement infrastructure exists for public acceptance of social science claims. There are no dual-cohort surveys comparing public acceptance of demographic projections to demographer consensus. There are no polling instruments measuring whether consumers of macroeconomic forecasting believe those forecasts at higher rates than the forecasters' track records would justify. There is no institutional campaign correcting overcredulous acceptance of sociological narratives about immigration causation, inequality dynamics, or population decline.

This is not because such overcredulous acceptance does not exist. Research on science polarization demonstrates that acceptance of macroeconomic and sociological claims is heavily and symmetrically driven by political identity, not by epistemic evaluation (Rekker, 2021; Kahan, Peters, Dawson, & Slovic, 2017). Both conservative and liberal publics accept social science consensus in domains where it aligns with political priors and reject it where it conflicts. Professional economic forecasters—people with advanced training, access to sophisticated models, and professional reputations on the line—demonstrate the same partisan skew in GDP predictions, according to analysis linking *Wall Street Journal* forecasting panel submissions to Federal Election Commission donation records.

The difference is that acceptance of high-warrant science in the face of political inconvenience is treated as a virtue to be cultivated, while acceptance of low-warrant science in the service of political identity is not recognized as an epistemic failure at all. The framing that dominates science communication concerns the public's failure to accept what science has established. The framing that is almost entirely absent concerns the public's failure to doubt what social science has asserted.

4.3 The Inversion Stated Plainly

The hard sciences achieve realization rates of 70 percent or better, replication rates in the range of 70 to 100 percent for well-established findings, and calibrated forecasting accuracy on short-range predictions. They attract active public skepticism, institutional remediation campaigns, and sustained concern about science denial.

The soft sciences achieve replication rates of 36 to 50 percent, positive result rates driven by publication incentives to 91 to 97 percent, forecasting accuracy that merges with random chance at the 3-to-5 year horizon, and macroeconomic prediction confidence intervals that are off by 30 percentage points in systematic, not random, directions. They attract near-zero organized public skepticism and are routinely treated as authoritative sources for policy decisions affecting hundreds of millions of people.

This is the inversion. Claims with the strongest evidence base attract the most doubt. Claims with the weakest evidence base attract the least.

5 Mechanisms

Three structural forces produce this inversion. They are not independent, but they are analytically separable.

5.1 Falsifiability Timescale

Hard science predictions fail visibly, quickly, and in ways that are publicly observable. A drug trial is completed. A bridge holds or it does not. An atmospheric model's ten-year prediction arrives. The feedback loop between prediction and observable outcome is short enough that public error correction is possible in principle, and the correction infrastructure—retraction systems, replications, institutional review—activates within timescales that practitioners can experience in their careers.

Social science predictions fail on timescales that systematically exceed human attention spans and professional accountability windows. A macroeconomic forecast for 2035 will be evaluated, if at all, by people who may not remember making the forecast, using methods that cannot be straightforwardly applied to complex post-hoc outcomes, by institutions that have no formal accountability mechanism for predictive failure. Tetlock documented this dynamic in detail: expert forecasters routinely defend failed predictions with arguments about timing (the prediction was right, just early), counterfactuals (the outcome would have been worse without the intervention), and near-misses (the mechanism was correct even if the specific outcome differed). These defenses are structurally available for any social science prediction precisely because the systems being predicted are complex enough that some supporting interpretation can always be constructed.

The practical consequence is that public doubt is generated primarily by observable failure. Where failures are invisible, delayed, or interpretable away, doubt does not accumulate. Hard science operates in the domain of visible failure. Social science operates in the domain of interpretable failure. The skepticism follows accordingly.

5.2 Identity-Protective Cognition

The psychology literature on motivated reasoning provides the second mechanism. Rekker (2021) establishes that rejection of scientific claims occurs at multiple distinct levels: rejection of specific findings (Level 1), rejection of an entire research field (Level 2), distrust of science as a whole (Level 3), and distrust of the system and elite more broadly (Level 4). Trust travels readily between levels: a person who distrusts the political establishment (Level 4) may use that general distrust as a heuristic to reject specific scientific claims (Level 1), even absent any engagement with the underlying evidence.

Kahan's cultural cognition research program provides the most rigorous empirical foundation for understanding how this operates. Kahan, Peters, Dawson, and Slovic (2017) designed an experiment that cleanly separates numeracy (the ability to reason quantitatively) from motivated reasoning (the tendency to reason in identity-consistent directions). Subjects were presented with a data interpretation task—a 2×2 contingency table requiring covariance detection—in two conditions. In the control condition, the data concerned the effectiveness of a skin cream. In the experimental condition, the identical numerical structure was reframed as data about gun control's effect on crime. In the skin cream condition, higher numeracy predicted more accurate interpretation, as expected. In the gun control condition, higher numeracy predicted *more* partisan bias, not less. Subjects with the strongest quantitative skills were the most likely to arrive at the interpretation that confirmed their prior political identity, regardless of what the data actually showed.

This finding directly undermines the science comprehension thesis—the assumption that scientific literacy reduces rejection of well-supported claims. Kahan's data shows that scientific literacy increases the *efficiency* of motivated reasoning rather than correcting it. The mechanism is what Kahan terms identity-protective cognition: individuals unconsciously process information in whatever way minimizes the threat to their standing within identity-defining groups. For a person whose community identity is bound up with skepticism of institutional authority, accepting the scientific consensus on climate change carries a social cost that no amount of data presentation can offset. The reasoning is not irrational; it is rational with respect to a different objective function—one that optimizes for social belonging rather than epistemic accuracy.

This has a direct and asymmetric implication for the inversion. The claims attracting skepticism from conservative publics—evolution, anthropogenic climate change, vaccine efficacy—are hard science claims that conflict with specific religious or cultural identity structures. The claims attracting skepticism from liberal publics—immigration crime statistics, macroeconomic costs of certain redistributive policies, biological contributions to behavioral sex differences—are soft science claims. Both forms of skepticism are identity-driven rather than evidence-driven. Kahan's data confirms the symmetry: neither side of the political spectrum is more or less prone to motivated reasoning in the aggregate. But the *consequences* of this symmetry are asymmetric in a way that produces the inversion. Conservative identity-protective cognition is directed at high-warrant claims and is therefore recognized as a science communication problem. Liberal identity-protective cognition is directed at low-warrant claims—and because those claims lack the correction infrastructure described in Section 5.1, the motivated acceptance is never measured, never challenged, and never recognized as an epistemic failure.

Kahan's research also identifies a partial exception that proves the rule. Scientific *curiosity*—measured by a validated scale distinct from scientific knowledge or numeracy—does reduce polarization. Curious individuals seek out and engage with information that challenges their prior

beliefs, even when that information is politically inconvenient. But curiosity is a personality trait, not a teachable skill, and its distribution is independent of educational attainment. This means that the standard prescription of the science communication field—more science education—addresses a variable (knowledge) that amplifies motivated reasoning while neglecting a variable (curiosity) that actually reduces it. The prescription is not merely ineffective; it is counterproductive with respect to the inversion, because it increases the sophistication with which both overdoubting and underdoubting are conducted.

5.3 The Selection Environment

The third mechanism operates at the level of the information channel rather than the individual. Channels select for transmissible claims. A transmissible claim in the context of mass media is one that is narratively coherent, emotionally engaging, practically actionable, and delivered with confidence. These properties are orthogonal to epistemic warrant.

The Tetlock finding that media prominence is negatively correlated with forecasting accuracy describes this selection process in quantitative terms. Media organizations do not select for accurate forecasters; they select for confident, narratively compelling ones. The forecaster who offers calibrated uncertainty—"the evidence suggests a 60 to 70 percent probability of recession within 18 months, with a wide confidence interval due to the interaction of several independent variables"—produces content that is less engaging and less shareable than the one who says "recession is coming and here is exactly why." The accurate fox produces what Tetlock calls a miasma of nuance. The inaccurate hedgehog produces a satisfying story.

This selection dynamic means that the public sphere is systematically populated by the epistemic actors least likely to be correct in domains where accuracy is difficult to verify. In hard science, this dynamic is partially checked by the existence of observable predictions that eventually arrive: the climate model, the particle discovery, the drug outcome. The forecaster who was wrong about measurable phenomena faces at least the possibility of public accountability. In soft science, where predictions operate on long timescales, involve complex counterfactuals, and can always be partially rescued by interpretive adjustments, the channel selects for confident narrators who face minimal accountability for the accuracy of their claims.

The transition from editorial to algorithmic content curation has altered the selection environment without clearly changing its direction. Social media recommendation systems optimize for engagement—a proxy for emotional arousal and identity-relevant signaling—which intensifies at least two of the three structural mechanisms described here. Echo chamber dynamics make credibility assessment identity-based rather than methodology-based: studies of bot misidentification show that users classify automated accounts as human when the account shares their political ori-

entation, and classify genuine human users as bots when they do not. Interventions designed to reduce partisan animosity and promote objective information processing show only modest effects—approximately 5.4 points on a 101-point scale—that decay within two weeks as subjects re-enter their polarized information environments. Algorithmic amplification of minority positions also produces pluralistic ignorance: global surveys on climate action reveal that while large majorities are willing to make financial sacrifices, they systematically underestimate the willingness of their fellow citizens by wide margins, a distortion attributable to the disproportionate amplification of skeptical voices.

At the same time, social media has partially disintermediated the metascientific discourse that made the replication crisis publicly visible. Blog posts, preprint discussions, and open peer commentary by methodologists—much of it conducted on social media platforms—played a material role in bringing the replication data reviewed in Section 3.2 into public awareness. This suggests a partial counterweight: the same platforms that intensify identity-protective cognition and select for confident narrators also create channels through which correction signals can bypass editorial gatekeeping.

The net effect of algorithmic curation on the inversion has not been directly measured. No existing study tests whether the specific pattern documented here—high-warrant claims attracting disproportionate doubt while low-warrant claims circulate with minimal friction—has accelerated, decelerated, or remained stable since the transition from editorial to algorithmic media dominance. This is an empirical question that the existing evidence does not resolve, and it is noted here as an open problem rather than addressed with speculation.

The aggregate effect of these three mechanisms—asymmetric falsifiability timescales, identity-protective cognition operating differently across domains, and channel selection for confidence over calibration—is the inversion described in Section 4. It is not a failure of individual intelligence. It is the expected output of a system with these properties.

6 The Throwaway Claim as Transmission Mechanism

The three forces described in Section 5 create the structural conditions for the inversion: asymmetric falsifiability ensures that low-warrant claims avoid correction, identity-protective cognition ensures that acceptance and rejection track group identity rather than evidence quality, and the selection environment ensures that the most confident and least accurate voices dominate the channels. But these forces require a transmission mechanism—a process at the level of individual propositions through which low-warrant claims achieve high acceptance in the course of normal discourse, without ever being evaluated as claims at all.

Consider a genre of factual assertion that appears routinely in prestige journalism and docu-

mentary media: the unattributed, uncited statement made in passing, embedded as shared context rather than argued as a proposition. A commentator observing that restrictive immigration policy in certain East Asian nations reflects the same racial purity ideology visible in rising right-wing politics globally is not presenting an argument. The sentence is constructed as shared context—something the viewer already knows and the speaker is merely acknowledging. No evidence is offered because none is required. The claim functions as scaffolding, not as content.

This rhetorical structure—what discourse semantics calls the presuppositional encoding of contested propositions—exploits a fundamental feature of cooperative communication. When a speaker embeds a proposition as assumed background, the cooperative listener accepts the presupposition in order to process the main point. The cognitive resources devoted to evaluating the claim are minimal because the claim is not presented as something that requires evaluation. The claim is presented as something that requires recognition.

The mechanism is documented extensively in communication theory and cognitive psychology. Claims presented as shared context are processed with lower cognitive friction than argued claims, which activates the illusory truth effect: repeated exposure to a proposition, regardless of its truth value, increases rated probability of truth. The relationship between repetition and belief is logarithmic, not linear: the most dramatic cognitive shift occurs upon the second encounter with a claim, with truth ratings approaching a practical asymptote after approximately nine exposures. Factual statements are significantly more prone to source-monitoring errors than evaluative or belief-based content—the listener forgets *where* the claim originated while retaining the claim itself. By the third or fourth encounter across independent media sources, the proposition has achieved the phenomenology of established fact for most consumers, regardless of its actual evidentiary status.

The correction infrastructure for presuppositionally embedded claims is weaker than for argued claims for the same reason it is weaker for social science generally: the feedback loop is broken. Argued claims generate rebuttals; presupposed claims generate neither assent nor dissent but simply accumulate in the background of shared discourse. When a presupposed claim is wrong and the error is eventually identified, the correction faces the documented asymmetry between original propagation speed and correction propagation speed. The continued influence effect—the empirically documented persistence of false beliefs after credible retraction—is most pronounced for claims that were initially encoded as background context rather than as argued propositions, because the mental model constructed around the claim has no obvious gap that the correction fills. The correction effort also faces a paradox: by explicitly addressing and repeating the throwaway claim in order to debunk it, the correction necessarily adds to the frequency of repetition. Because processing fluency increases logarithmically with repetition regardless of truth value, the corrective effort can actively backfire, increasing the perceived truth of the claim it intends to retract.

This mechanism is a testable hypothesis rather than a finding of the present paper. A corpus study of prestige journalism, measuring the domain distribution and propagation rates of syntactically backgrounded claims against argued claims, would provide direct evidence for or against the prediction that throwaway claims cluster in low-warrant domains and propagate faster than their argued counterparts. Such a study is technically feasible: existing NLP pipelines for claim extraction (e.g., NEWSCLAIMS, ClaimBuster) achieve high recall at the candidate extraction stage, and presupposition detection classifiers report F1 scores above 0.65 for identifying not-at-issue content. However, no standardized end-to-end task for backgrounded claim detection in news text currently exists, and existing check-worthiness systems actively down-rank the class of unattributed factual assertions that constitute throwaway claims—treating as unimportant precisely the propositions whose propagation dynamics this paper identifies as consequential. A companion research program addressing this gap is underway. For the purposes of this paper, the throwaway claim mechanism is presented as the transmission process through which the structural conditions documented in Sections 3 through 5 are realized in individual acts of communication. The three structural forces create the environment; the throwaway claim is the vehicle. Asymmetric falsifiability means no correction arrives. Identity-protective cognition means the listener is predisposed to accept. Channel selection means the speaker is confident and uncalibrated. The throwaway claim completes the circuit by encoding the low-warrant proposition as background rather than argument, bypassing the listener's remaining evaluative capacity entirely.

7 Implications

7.1 For Science Communication

The science communication literature is organized almost entirely around a deficit model: the public lacks accurate information about high-warrant science, and the solution is more and better information provision. The evidence reviewed here suggests that this model addresses, at best, half of the problem and may actively obscure the other half.

If the public systematically underdoubts low-warrant claims at the same time as it overdoubts high-warrant ones, then successful campaigns to increase acceptance of climate science, vaccine safety, and evolutionary biology do not correct the underlying epistemic misalignment. They shift the balance without addressing its cause. A public that accepts the heliocentric model and the mechanism of vaccination while simultaneously accepting without evaluation a macroeconomic projection with 23 percent historical accuracy has not become more epistemically reliable. It has become differentially compliant.

The science communication field has invested heavily in understanding science denial. It has

invested almost nothing in understanding science overcredulity—the systematic acceptance of institutionally branded claims that do not carry the epistemic warrant implied by the institutional branding. This is the other half of the inversion, and it is at least as consequential for the quality of public reasoning and policy formation.

7.2 For Policy

Policies built on low-warrant social science claims are not constructed with lower confidence than policies built on high-warrant hard science claims. They are often constructed with higher confidence, because the selection environment described in Section 5.3 rewards confident assertion regardless of underlying accuracy, and because the absence of short-feedback-loop accountability means that confident but inaccurate claims are not rapidly penalized. The practical result is bold policy on weak foundations coupled with institutional resistance to policy on strong foundations.

The clinical medicine example is instructive because it is the domain where the failure is most directly measurable. NCCN oncology guidelines carry the force of federal reimbursement law and are implemented by oncologists as the authoritative standard of care. Approximately 6 to 8 percent of the therapeutic recommendations in those guidelines are supported by randomized controlled trial evidence. For some cancer types, the proportion is zero. A healthcare system allocating hundreds of billions of dollars annually to cancer treatment on the basis of expert panel consensus unsupported by high-level evidence has constructed one of the largest policy programs in human history on the weakest available evidentiary foundation—and this is considered the normal condition of evidence-based medicine.

This is not unique to oncology. It is the expected outcome when expert authority is institutionally substituted for empirical evidence across any domain where the evidence is unavailable, inconvenient, or simply has not been sought.

7.3 For the Sociology of Knowledge

The Positive Predictive Value of a published finding in a low-consensus field—the probability that a positive result reflects a true effect rather than a false positive—approaches the base rate of true hypotheses in that field. Ioannidis demonstrated in 2005 that under typical study parameters in the biomedical literature, including small samples, low prior probability, and researcher degrees of freedom, the majority of published findings may be false. This calculation has been contested in its particulars but its basic logic has not been successfully refuted, and the replication data reviewed in Section 3 is broadly consistent with it.

The implication for the sociology of knowledge is that accepting institutional consensus in a low-warrant field is, under some conditions, epistemically worse than coin-flipping. The consensus

carries a base rate of accuracy that is documented to be below 50 percent, but it carries institutional authority that makes it much harder to doubt than a random proposition would be. The authority is real in the sense that it influences behavior and shapes discourse. It is not real in the sense of being proportional to the underlying evidence. The gap between institutional authority and epistemic warrant is exactly what the inversion exploits.

8 Limitations and Methodological Constraints

Several limitations of this analysis deserve direct acknowledgment.

The most significant constraint is that the inversion is demonstrated inferentially rather than directly. The measurement infrastructure for public skepticism, as documented in Section 2.2, covers hard science domains through dual-cohort survey methodology and does not cover social science claims. The claim that public skepticism is absent toward low-warrant social science is supported by the absence of corrective campaigns, the absence of equivalent polling infrastructure, and the political science research showing that acceptance of social science consensus is driven by partisan identity rather than epistemic evaluation. But a direct measurement—a Pew-style survey asking the public to rate their confidence in macroeconomic projections against a calibrated expert consensus—does not exist. This paper calls explicitly for the creation of such measurement tools.

Second, the selection of examples for the hard science side of the inversion carries a risk of motivated presentation. The hard sciences have been wrong at the consensus level on significant questions: continental drift was denied by geological consensus for decades after Wegener's evidence was available; peptic ulcer disease was believed to be a product of stress until Barry Marshall's *Helicobacter pylori* demonstration; the low-fat dietary guidelines that shaped decades of American nutritional policy rested on observational evidence that was later contradicted. These errors are acknowledged here not as defeaters of the argument but as evidence that the epistemic warrant claim is probabilistic and comparative rather than absolute. Hard science consensus is substantially more reliable than soft science consensus; it is not infallible.

Third, the psychology exception documented in the literature requires acknowledgment and resolution. The public does express skepticism toward academic psychology when it is explicitly presented as science—dismissing it as "just common sense" or questioning its scientific rigor. This appears to contradict the inversion argument. The resolution is that the public's skepticism toward laboratory psychology and its acceptance of social science narratives are not directed at the same object. The public is skeptical of the scientific credentials of psychology as an academic enterprise. It is not skeptical of the narratives that psychological and sociological frameworks produce when they are delivered as confident sense-making by authoritative voices in media and policy contexts. The inversion operates on the rhetorical register of the claim—argued science

versus ambient narrative—not only on the domain label.

Fourth, the mechanism proposed in Section 6 is presented as a testable hypothesis rather than a demonstrated finding. The corpus study required to directly measure throwaway claim propagation rates and domain distribution is the subject of a companion research program and is not reported here.

Fifth, this analysis does not engage with the question of whether some degree of epistemic inversion is adaptive. Identity-protective cognition exists because it serves social coordination functions that may outweigh individual epistemic accuracy in evolutionary fitness terms. A society that accepted every scientific finding regardless of its implications for social cohesion would face coordination problems that a society with some motivated reasoning does not. This tradeoff is real and is not addressed here because it concerns whether the inversion should be corrected, not whether it exists.

9 Conclusion

The distribution of doubt in public epistemology is not random, not primarily driven by ignorance, and not adequately described by the deficit model that has organized science communication for three decades. It is structured by the interaction of falsifiability timescales, identity-protective cognition, and channel selection pressures that together produce a consistent inversion: the most reliable knowledge attracts the most doubt, and the least reliable knowledge circulates with the least friction.

The evidence for this pattern runs from the NISTEP Delphi retrospectives showing 70 to 80 percent hard science realization rates through Tetlock's documentation of near-random expert forecasting in political and economic domains through Herrera-Perez and colleagues' 13 percent clinical reversal rate through Fanelli's measurement of a positive result gradient tracking the hierarchy of sciences and Scheel and colleagues' 52-point distortion between standard and registered report positive result rates in psychology. Across every domain where the comparison can be made, the gradient runs in the same direction: fields with tight theoretical constraints, short feedback loops, and high replication rates attract disproportionate public skepticism, while fields with loose theoretical constraints, long feedback loops, and low replication rates attract disproportionate public acceptance.

This cross-domain synthesis—that above a threshold, high acceptance rate is evidence of low accuracy—has not previously been formalized in the philosophy of science literature. The component mechanisms are documented across metascience, forecasting research, risk psychology, and science communication, but they have not been assembled into a unified argument. That assembly is the primary contribution of this paper.

The implications are uncomfortable in multiple directions. For science communication, they suggest that the field's focus on science denial is addressing the less consequential half of a two-sided problem. For policy, they suggest that institutional confidence in expert consensus is systematically miscalibrated in favor of domains where the accuracy track record is worst. For the sociology of knowledge, they raise the possibility that the authority structures through which scientific expertise is transmitted to public discourse are selecting, under normal operating conditions, for the least reliable epistemic actors in every domain where direct measurement of forecasting accuracy is avoided.

The cage that constrains public epistemology is not built from lies. It is built from claims that are accurate enough to transmit, confident enough to act on, and wrong often enough to produce outcomes that nobody chose and nobody planned.

References

Brodeur, A., Lé, M., Sangnier, M., & Zylberberg, Y. (2016). Star Wars: The empirics strike back. *American Economic Journal: Applied Economics*, 8(1), 1–32.

Crichton, M. (2002, April 26). Why speculate? [Address]. International Leadership Forum, La Jolla, CA.

Fanelli, D. (2010). "Positive" results increase down the hierarchy of the sciences. *PLOS ONE*, 5(4), e10068.

Fanelli, D. (2012). Negative results are disappearing from most disciplines and countries. *Scientometrics*, 90(3), 891–904.

Herrera-Perez, D., Haslam, A., Crain, T., Gill, J., Livingston, C., Kaestner, V., Hayes, M., Morgan, D., Cifu, A. S., & Prasad, V. (2019). A comprehensive review of randomized clinical trials in three medical journals reveals 396 medical reversals. *eLife*, 8, e45183.

Ioannidis, J. P. A. (2005). Why most published research findings are false. *PLOS Medicine*, 2(8), e124.

Kahan, D. M., Landrum, A., Carpenter, K., Helft, L., & Hall Jamieson, K. (2017). Science curiosity and political information processing. *Advances in Political Psychology*, 38(S1), 179–199.

Kahan, D. M., Peters, E., Dawson, E. C., & Slovic, P. (2017). Motivated numeracy and enlightened self-government. *Behavioural Public Policy*, 1(1), 54–86.

Moore, D. A., Tenney, E. R., & Haran, U. (2024). Overprecision in the Survey of Professional Forecasters. *Collabra: Psychology*, 10(1), 92953.

NISTEP (National Institute of Science and Technology Policy). (2005). *The 8th Science and Technology Foresight Survey: Delphi Analysis.* Science and Technology Foresight Center.

Open Science Collaboration. (2015). Estimating the replicability of psychological science. *Science*, 349(6251), aac4716.

Pew Research Center. (2015). *Public and scientists' views on science and society.* https://www.pewresearch.org/science/2015/01/29/public-and-scientists-views-on-science-and-society/

Poonacha, T. K., & Go, R. S. (2011). Level of scientific evidence underlying recommendations arising from the National Comprehensive Cancer Network clinical practice guidelines. *Journal of Clinical Oncology*, 29(2), 186–191.

Prasad, V., Vandross, A., Toomey, C., Cheung, M., Rho, J., Quinn, S., Chacko, S. J., Borkar, D., Gall, V., Selvaraj, S., Ho, N., & Cifu, A. (2013). A decade of reversal: An analysis of 146 contradicted medical practices. *Mayo Clinic Proceedings*, 88(8), 790–798.

Rekker, R. (2021). The nature and origins of political polarization over science. *Public Understanding of Science*, 30(4), 352–368.

Scheel, A. M., Schijen, M. R. M. J., & Lakens, D. (2021). An excess of positive results: Comparing the standard psychology literature with registered reports. *Advances in Methods and Practices in Psychological Science*, 4(2).

Sterling, T. D. (1959). Publication decisions and their possible effects on inferences drawn from tests of significance—or vice versa. *Journal of the American Statistical Association*, 54(285), 30–34.

Survey of Professional Forecasters. Federal Reserve Bank of Philadelphia. https://www.philadelphiafed.org/surveys-and-data/real-time-data-research/survey-of-professional-forecasters

Taleb, N. N. (2007). *The Black Swan: The Impact of the Highly Improbable.* Random House.

Tetlock, P. E. (2005). *Expert political judgment: How good is it? How can we know?* Princeton University Press.

Tetlock, P. E., & Gardner, D. (2015). *Superforecasting: The art and science of prediction.* Crown Publishers.

Correspondence: Jeremy McEntire, perardua.dev/research. ORCID: 0009-0006-8008-9587.

The author declares no competing interests. No external funding was received for this work.