How Much Precision Is Really Enough?

A Look at the Precision of Writing in an Age of Unlimited Judgement

Ramin Saadat
Independent Researcher; Founder, British Institute for Ignorance Studies (BIIS) (biis.uk)

AI Disclosure and Translation Statement:
The original manuscript of this work was authored in Farsi (Persian) by the researcher. The English version was produced through AI-assisted translation and linguistic refinement, utilizing Google Translate and Large Language Models. The author has meticulously reviewed, edited, and verified the final English text to ensure that all theoretical nuances, philosophical arguments, and conceptual frameworks accurately reflect the original Persian content. The author remains fully responsible for the integrity and accuracy of the translation. The original Persian manuscript is available from the author upon reasonable request.

Abstract

Large language models have turned the supply of critique from something scarce and costly into something abundant, instant and almost free. As a result, the list of objections these models raise against any text is structurally endless and never reaches saturation. This article argues that the common question, 'How much precision is enough?', is wrongly framed. Contrary to popular assumption, precision and correctness are not actually in dispute. What genuinely needs to be settled is the degree of stringency: the effort a writer must spend to secure those two goals. Drawing on the sceptical tradition (the Agrippan trilemma), the philosophy of science (the Duhem–Quine thesis), the sociology of science (Collins's experimenter's regress), and cognitive economics, the article argues that stopping revision has always been a conventional decision, not the discovery of an objective point. The theoretical core of the article is this claim: the agnotology literature has already documented the demand for unlimited stringency as a recognised technique for producing ignorance, and what happens today is an emergent, agentless form of that same technique. The article closes with a practical four-question criterion and a distinction among five functional domains of writing.

Keywords: stringency in writing; agnotology; ignorance production; large language models; artificial sycophancy; the Agrippan trilemma; cognitive economics; functional domain

---

Introduction

Let me introduce my subject and my problem with a familiar scene.

Almost every day, something like this happens to us. We write a text: a simple email to a friend, a piece of administrative work, or a specialised academic passage. We hand our text to one of the large AI language models and ask it to review the text and point out its flaws. Depending on the model, its capability and the level of precision we have chosen, it returns a list of objections: spelling mistakes, poor word choice, grammar, faulty reasoning, internal contradictions. We correct the text and give it back to the model. A fresh list comes back, not necessarily shorter than the last one. We try a different model; new objections appear, sometimes harsher and more serious. The cycle seems endless, and it continues until we finally tell the model: 'This level of precision is enough. I want nothing more.'

This sentence looks practical and managerial on the surface. In fact it is a full-blown epistemic decision, and this may be the first time in the history of writing that we are forced to make such a hard decision explicitly.

This is exactly the question of this article: on what basis can one say such a sentence?

You might ask why this problem is new. You might object that writers have always stopped revising at some point, and one could always find one more fault...

This objection is partly right, but it misses the main point. In the past, stopping revision was not a conscious decision. It happened only because of the writer's capacity and surrounding circumstances. A deadline arrived, the writer's patience ran out, a competent reader to whom the text could be entrusted was unavailable, or the cost of another round of reading simply made a further round unjustifiable. In other words, the supply of critique was scarce and costly, and this scarcity, without anyone thinking about it, automatically did the work of setting the threshold.

Today, though, the mode of supply has changed. Critique is now abundant, instant and almost free. More important still, it is multivocal: one can summon countless judges, each applying the criteria of a different discipline. As a result, once the old constraints disappeared, that automatic threshold disappeared with them. The task that external circumstances once performed now falls to the writer, and the writer is not prepared for it, because it was never necessary before.

So the problem of this article is neither technical nor about tools. The tool has merely drawn back the curtain and exposed a question that always existed but stayed unanswered: when is a text correct and precise enough? How much precision do we really need?

Before going further into the discussion of judgement, we need to distinguish three related concepts here: correctness, precision and stringency. Everyday language usually treats these as one and the same. Note that these definitions make no claim to any external standard; they are simply three working concepts we use, in this form, to advance the argument:

●Correctness: closeness to fact. That the text is not wrong.

●Precision: the granularity and resolvability of statements. That the text is not vague.

●Stringency: the amount of effort spent securing the two goals above.

Of these three, correctness and precision are not in dispute in this article, because no one claims that a wrong or vague text is desirable. What is in dispute, and what forms the subject of this article, is stringency: not whether a text should be correct and precise, but how far a writer must go to achieve those two goals. What follows shows why setting this degree has, today, become an explicit and unavoidable decision.

To examine the question 'how much stringency do we need?', the article proceeds through the following sections:

●Section 1 shows why the list of objections never ends.

●Section 2 shows why the model, in practice, always fills this endless potential.

●Section 3 recovers a positive criterion from the philosophical tradition.

●Section 4 turns that criterion into a problem of resource allocation.

●Section 5 formalises the distinction among domains with a two-axis table.

●Section 6 is the theoretical core of the article and presents its central claim: the demand for greater stringency has already been recorded in the agnotology literature as a recognised technique for producing ignorance, and what happens today is an emergent, agentless form of that same technique. Readers who take this article for a note about AI tools should wait until that section.

●Section 7 offers a practical criterion.

●Section 8 answers possible questions and objections.

●Finally, a conclusion follows.

---

1. Unending Judgement

To see why the list of objections never ends, we must first clarify the phenomenon of critique itself, because conflating its different aspects derails the discussion. From the perspective of this article, the phenomenon of critique has three separate features:

●Abundance: critique, especially today, is available in whatever quantity one wants.

●Diversity: critique can come from different standpoints: a journal referee, a supervisor, a general reader, an opposing critic. Each of these judges produces a different list, and the sum of these lists is usually inconsistent. What one calls ambiguity, another calls concision (we have deliberately left language models off this list; as Section 2 shows, this diversity of standpoints is not specific to them, and existed before they did).

●Non-saturation: the list of objections does not shrink with each round of correction. The revised text picks up fresh objections. This is the feature that produces the psychological experience of 'my work is full of faults', because the signal a writer expects (a gradual shrinking of the fault list towards zero) almost never appears.

Of these three, we will examine abundance and diversity in later sections (abundance turns into the problem of resource allocation in Section 4, and diversity finds its empirical evidence in Section 2). What needs an independent argument here, in this section, is only non-saturation, because it is easy to mistake it for a temporary technical limitation rather than a structural feature.

The claim of this article is that non-saturation is a structural feature, not a technical fault that a different or better model can fix. This claim matters because, if the problem were technical, the right expectation would be to wait for the technology to improve and resolve it. If it is structural, however, something must be done about it.

To establish that non-saturation is structural, we offer three arguments, each entering from a different level: the logical, the epistemological and the sociological.

The first argument: the regress of justification

The first reason for non-saturation lies in the very structure of justification and has nothing to do with technology.

This argument traces back to the Greek sceptics. In the Outlines of Pyrrhonism, Sextus Empiricus records Agrippa's five modes, three of which bear directly on our discussion: regress, circularity and hypothesis (Sextus Empiricus, trans. 2000, Book I, §§164–177). The threefold formulation known today as the Agrippan trilemma is a later reading of this fivefold set, which Hans Albert reformulated as the Münchhausen trilemma (Albert, 1985).

We can summarise these three paths as follows. For any claim, one can demand a reason. That reason is itself a claim and demands a reason of its own. Only three paths remain open: an infinite regress, a circle (where the reason loops back to the original claim), or a stop at a point that is itself unjustified.

The point that matters for our discussion is that this argument does not prove scepticism. Rather, the third path, an arbitrary stop, is in practice the only escape from non-saturation, and every text ever written has used it. What we call a well-argued text is simply a text that has stopped giving further reasons at some given point, and no intrinsic feature of that point requires stopping there.

Popper accepts exactly this with exemplary clarity. In The Logic of Scientific Discovery, discussing basic statements, he states plainly that testing stops at the point where we decide to accept a statement 'for the time being', and that this acceptance is conventional, not self-evident (Popper, 1959, ch. V).

His famous image in that same discussion, that science is built on a swamp and we drive piles down only far enough to hold the structure up, not down to bedrock, because there is no bedrock, is exactly what this article says about a text.

This response is no accident. Popper is consciously reformulating the ancient sceptical trilemma, and he presents his own version as Fries's trilemma, a nineteenth-century formulation from the same family as the Agrippan trilemma: dogmatism, infinite regress, or psychologism (Popper, 1959, ch. V).

We therefore see that non-saturation is not a property of the model. It is a fundamental, structural feature of demanding reasons at all. In our era, the model has merely reduced the cost of continuing the regress to zero, and as a result has left undetermined the stopping point that cost used to fix.

The second argument: underdetermination

The second reason comes from the philosophy of science and explains why one can always manufacture an objection, even against a wholly correct text.

The Duhem–Quine thesis holds that no hypothesis is tested on its own. What gets tested is the hypothesis plus a set of auxiliary assumptions (Duhem, 1906/1954). In 'Two Dogmas of Empiricism', Quine pushes this to its radical form: any statement can be held true in the face of any evidence, provided we make sufficient adjustments elsewhere in the web of belief, and conversely, no statement is immune to revision (Quine, 1951).

Duhem and Quine formulated this thesis about scientific theories, but the structure of the argument depends on no property specific to science. Wherever a claim rests on unstated assumptions, the same structure holds, and a written text is exactly such a place. Every claim in a text rests on a set of unstated assumptions: about the meaning of its terms, about what needs no proof, about what the reader already knows. A critic can target any one of these assumptions, and doing so is legitimate, because the assumption really has gone unstated. But the number of such assumptions is not finite, so the list of legitimate objections is not finite either.

A conclusion follows that looks paradoxical at first and is worth stressing. 'This text has many objections' and 'this text is defensible' can both be true at once. Their apparent inconsistency comes from the false assumption that the number of objections indicates quality. It does not. In fact, the number of objections indicates only how much effort has gone into finding them.

The third argument: the experimenter's regress

The previous two arguments were theoretical. The third is empirical and comes from the sociology of science.

In his study of laboratory efforts to detect gravitational waves, Harry Collins formulated a problem he called the experimenter's regress (Collins, 1985). The problem runs as follows. To know that an experiment was carried out correctly, we must know what its result should be. But the result is exactly what we wanted to find out. So no independent criterion exists for 'well enough done'.

Collins found that, in practice, social consensus breaks this circle, not the arrival at an objective threshold. The research community declares the debate closed at some point, and factors that determine that point are not internal to the experiment itself: the standing of individuals, fatigue, the allocation of resources to other problems.

Collins documented this circle in a physics laboratory, but the structure of the problem depends on no property specific to experimentation. Wherever the criterion of 'good enough' cannot be drawn from anything but the very work under judgement, the same circle recurs, and revising a text is exactly such a situation. No intrinsic feature of a text says when revision has finished. What used to break this circle was external: deadlines, editors, patience. That is exactly what has vanished in the new situation. A writer who works alone with a model has fallen outside the constraining social consensus and cycles through correction, revision and judgement without limit.

Taken together, then, all three arguments (logical, epistemological and sociological) reach the same conclusion: stopping optimisation is not a discovery, it is a decision.

If this is right, the question 'how much stringency is needed?' appears to have been wrongly framed from the outset. The right question is not where the threshold lies, as if a threshold of perfection were something to be discovered, but on what basis to set it. The difference between these two framings is the difference between search and judgement, and that difference is the central subject of the rest of this article.

Yet even though we have seen that the list of objections is inherently endless, that is not the whole story. The model does not merely have the capacity to extend this list indefinitely; in practice, it always does so, and a separate mechanism must explain why.

In other words, the remaining question is why the model always fills the list, and why it never tells a good text that it has no serious faults. The answer, it seems, lies not in the limits of the model's ability but in its motivational mechanism.

---

2. Critique as Performance

The previous section showed that, structurally, the list of objections can never saturate. We said the answer to why lies not in the model's limited ability but in its motivational mechanism, and that mechanism is the subject of this section.

The fact is that when we ask a model to write 'as a strict judge', what we get is not the output of an assessment; it is the performance of a role. The model plays a role whose defining feature is finding fault, and no feedback exists to correct that role. A real referee who is excessively strict damages their standing with an editor, but a simulated machine referee pays no cost for excessive strictness.

Goffman's analysis of role performance in The Presentation of Self in Everyday Life gives us a precise conceptual tool for this situation, particularly his distinction between front stage and back

stage (Goffman, 1959). In humans, performing a role always has a back stage, a place where the actor steps back from the role. In a model, no back stage exists. The performance is the whole of what is, and all of what is seen.

This is the same structure raised in discussions of honesty in models trained with human feedback: the distinction between being honest and appearing honest collapses in practice, where only the output is observable. Our problem is the same problem with a different variable: the distinction between genuine stringency and performed stringency.

The key point here is that existing research on the sycophancy of language models shows that models trained with human feedback have a systematic tendency to align with the user's position, even at the expense of correctness (Sharma et al., 2023). A larger-scale study (Cheng et al., 2026), with more than two thousand participants, confirms and sharpens this finding: sycophantic responses are still rated as more trustworthy even when they distort the user's moral judgement. That is, the training process cannot distinguish a signal that merely raises user satisfaction from a genuine signal. The structural explanation is easy to grasp once we recall that human raters prefer agreeable answers, and this preference settles into the reward function as well.

This finding is usually understood to mean that the model over-approves. That understanding is incomplete, and in fact misleading. What the model aligns with is not the content of the user's position; it is the user's expectation. If the user expects approval, alignment means approval. If the user, by saying 'criticise this harshly', expects harsh criticism, alignment means producing harsh criticism.

The 'merciless referee' and the 'undeserved flatterer' are therefore not two different phenomena, one desirable and the other not. They are two forms of a single mechanism, and both are, at least in part, a function of the prompt, not just of the text. This claim should not be read as absolute; the model is also sensitive to the content of the text, or no criticism would ever land. But the prompt's share in setting the intensity and direction of criticism is large enough that the length or severity of the list, on its own and without controlling for the prompt, is not a reliable measure of the text's quality.

The practical consequence of this claim should be stated plainly: a writer who asks a model for harsh criticism, to escape its sycophancy, has not escaped sycophancy at all. They have only changed its direction.

To keep this argument from collapsing into technophobia, we need to show that this pattern is documented in human judgement too, and that the model has merely intensified it. Here we can point to at least two classic studies.

Peters and Ceci resubmitted twelve articles, already published in reputable psychology journals, to those same journals under changed author names and institutions. Most of the surviving submissions were rejected, and the stated reason for rejection was methodological flaws (Peters & Ceci, 1982). The point of the study was not to confirm the seriousness of those flaws. The point was that the texts had not changed; only the reviewing situation had changed. In other words, the same flaws had gone unseen the first time round.

In another study, Cole, Cole and Simon showed that, in the assessment of research proposals to the US National Science Foundation, most of the variance in outcome was explained by disagreement among reviewers (Cole, Cole, & Simon, 1981). Their own conclusion was not that the quality of a proposal has no effect, but that this variance reflects the absence of real consensus

in the scientific community, a finding that is just as unsettling, though more nuanced than saying ‘quality plays no role’.

Both studies show one thing in common: the volume of objections found is a function of the reviewer’s situation, not only of the text itself. The academic slang term ‘Reviewer 2’ is the informal name for exactly this finding.

This clarifies why the list of objections can be endless and why, in practice, it is endless. These are two distinct matters, and the difference matters greatly: the first is a possibility, the second a mechanism. Combining the two yields a conclusion that grounds the sections to follow: the list of objections returned by a model is not a pure signal about the text. It is a mixture of the text’s signal and the prompt’s noise, and without controlling the prompt one cannot separate the two. Any decision a writer makes based solely on the length or severity of that list, without regard to which prompt produced it, rests on data that cannot be disentangled.

---

3. Precision Fitted to the Subject

The previous two sections were negative in character. They showed why the unlimited demand for stringency never ends and why we should not take it seriously. No positive criterion has yet emerged, however. This section turns to a criterion from the philosophical tradition, one far older than our problem, which is itself evidence that the problem is not technological. What that tradition usually discusses under the heading of precision is what this article calls stringency: the amount of effort one must spend to secure correctness and precision. To stay faithful to the language of these texts, the quotations below keep the word precision, but each conclusion bears directly on stringency.

In the Nicomachean Ethics, Book I, Chapter 3, Aristotle states a principle that is, in effect, the thesis of this entire article:

‘It is the mark of an educated person to look for precision in each class of things just so far as the nature of the subject admits’ (Aristotle, trans. 2009, Book I, Ch. 3). He then names two symmetrical errors: accepting merely probable reasoning from a mathematician, and demanding rigorous proof from an orator. He judges both equally unreasonable.

From the perspective of this article, three points can be drawn from Aristotle’s principle:

●First, symmetry: Aristotle does not say that less stringency is better. He says that mismatched stringency is an error, in either direction.

●Second, the nature of the subject: the criterion is not the writer’s taste or ease of work. It is the nature of what is being discussed. This keeps the criterion objective and separates it from relativism.

●Third, education: Aristotle attributes the judgement of the right degree not to a rule but to a person’s formation. This point is uncomfortable for our article, because it implies that the criterion we propose in Section 7 cannot be taught to, or generalised for, someone who lacks this Aristotelian formation.

Aristotle is right that the degree of precision depends on the subject, but Wittgenstein goes a step further and says that the very concept of precision is meaningless without specifying a purpose. In Philosophical Investigations, §88, he explains this with an example:

If I say ‘stand roughly there’, is this order vague? One might object that exact boundaries are not fixed. But the order works: the listener stops, and stops where they were meant to stop. So the order is entirely precise for its purpose. Demanding sharper boundaries does not increase precision; it only changes the purpose (Wittgenstein, trans. 2009, §88).

The conclusion Wittgenstein draws from this is clear and forceful: no single ideal of precision exists. ‘Precise’ is not a property a text either has or lacks absolutely; it is always precise for some purpose.

This reading matters greatly for our article, because it shows that the question ‘is this text precise enough?’, without specifying its purpose and its audience, is a question that has no answer. Not because the answer is difficult, but because the question itself is incomplete. The model answers exactly this incomplete question, and it has no choice but to complete it with assumptions of its own that the writer never stated.

Borges’s one-paragraph fable ‘On Exactitude in Science’ casts the same conclusion in an image: an empire in which the art of cartography reaches such perfection that it finally produces a map on a scale of one to one, exactly matching the territory. Later generations find it useless and abandon it to sun and winter (Borges, trans. 1998).

What this fable adds to the two previous arguments is a further step. Aristotle said mismatched stringency is unreasonable. Wittgenstein said it is meaningless. Borges says that, past a certain point, it is destructive. The one-to-one map is not useless for being too small; it is useless because it has left nothing out, and leaving things out is the whole job of a map. This third step is essential for our article, because only with it can we move from a purely economic position (that more stringency costs more) to an epistemic one: past a certain point, more stringency makes a text worse. Section 7, in discussing defensive writing, applies exactly this step.

These three judgements together say that mismatched stringency is unreasonable, meaningless and destructive. None of them, however, tells the writer how to decide, faced with a specific objection, whether fixing it is necessary. We must draw that practical test from elsewhere: from Bateson’s definition of information as ‘a difference that makes a difference’ (Bateson, 1972). In his view, the world contains countless differences, but only those that affect a system count as information. This definition turns the three preceding philosophical judgements into a workable test, as we shall see in the fourth question of Section 7.

So we have four sources, four independent judgements, stacking on top of one another: stringency must fit the nature of the subject (Aristotle); precision is an incomplete concept without a stated purpose (Wittgenstein); past a certain point, stringency destroys function (Borges); and an objection with no effect on the outcome is not information (Bateson). The philosophical tradition, then, already holds our answer. What it lacks is a procedure for applying that answer to a specific list of objections, and filling that gap is the task of Section 7.

But before that, we need to see what exactly the word fitting means in Aristotle's principle, and to answer this we must turn to another field: cognitive economics.

---

4. Cognitive Economy

Section 3 showed that stringency must be proportionate, but the very concept of proportionate needs a definition. This section formulates proportionality in terms of resource allocation, and by

doing so moves the article's question from the domain of pure epistemology into a domain where an answer is determinable.

In a note on the economy of research, Charles Sanders Peirce was probably the first to state explicitly that research resources, time, attention and budget, are limited and must be divided among competing questions. Pursuing a single problem to the point of perfection is therefore equivalent to abandoning other problems (Peirce, 1879). A subtle point Peirce added, and one often overlooked, is his belief that this constraint is not an external obstacle to epistemic work; it is part of the method itself. A researcher who fails to weigh opportunity cost when choosing a problem is doing bad research, not merely slow research.

On this approach, the tenth round of revising one text competes with writing the next one. A writer who does not see this thinks they are choosing between a more precise text and a less precise one. In reality, they are choosing between this text being slightly better and another text not existing at all.

Peirce, though, only says that resources must be divided; he does not say how an agent, in practice, recognises that it is time to move on to the next task. Herbert Simon answers this question by arguing that a real agent does not choose the optimal option. It accepts the first option that clears its aspiration level, a behaviour he called satisficing (Simon, 1955).

One might read this behaviour as resignation to imperfection, but Simon argues the opposite. Because searching itself has a cost, satisficing is, in real conditions, a rational behaviour, and pursuing maximal optimisation, in a world of limited resources, is irrational.

Simon later refined this view, arguing that, in an information-rich environment, the scarce resource is not information but attention, and an abundance of information makes attention scarcer still, precisely because it consumes more of it (Simon, 1971). This point maps precisely onto our earlier example and our present circumstances: the problem facing a writer who receives a thousand objections is not a shortage of feedback. It is a shortage of the attention that feedback consumes, and the Simonian solution is not to add more feedback to sift through. It is to declare an aspiration level.

Even this criterion, however, seems insufficient. Simon tells us where to stop, the aspiration level, but he does not chart the utility curve leading up to that point. Nicholas Rescher did this, pushing Simon's criterion towards a more quantitative model and showing that research has diminishing returns: easy findings come first, and each additional unit of certainty costs more than the one before it (Rescher, 1989).

In other words, on Rescher's view, the first round of revision catches gross errors, the second catches subtler ones, and from the fourth or fifth round onward, what one finds is mostly a matter of taste that shifts with every rereading. In practical terms, we can say that once the current round's objections conflict with the previous round's, that is, something already corrected is now flagged as a fault, we have entered the zone of taste, and continuing is no longer improvement. It is wandering.

What Peirce, Simon and Rescher share is that they treat the problem purely as economic, a matter of cognition and attention: resources are limited, so they must be divided. William James, however, showed that the danger of endless revision runs deeper still. It is not merely inefficient; it is an epistemic failure.

In 'The Will to Believe', James draws a precise distinction. We have two epistemic commands, he says, usually taken as one but in fact in tension: seek truth, and avoid error (James, 1896). If we

optimise only for the second command, avoiding error, an easy path lies open: believe nothing. Taking this path, we will never err, but we will also never know anything and never reach the truth. James shows that this is not epistemic caution; it is epistemic failure, because it sacrifices one goal entirely for the other.

In other words, a text that is never finished and never published contains no errors, so it fully satisfies the second command. But it also fails entirely to meet the first. A writer who revises without end believes they are defending precision, but in fact they are abandoning one of the two commands altogether.

We can see, then, that the question 'how much stringency is needed?' is not an epistemological question with an absolute answer. It is a problem of resource allocation, and its answer is a function of three variables:

●The cost of error: what happens if this text is wrong somewhere? The answer differs for a medication leaflet and for a philosophical essay.

●The audience: who reads it, and on what basis do they judge it?

●Opportunity cost: what other work does this round of revision displace?

These three variables are the three inputs to the threshold, and note that none of the three lies inside the text. The threshold cannot be derived from the text itself, which is the same conclusion as Section 1, reached this time by a different route.

What matters is that these three variables are not the same from one functional domain to another, and that difference is the subject of the next section.

---

5. Distinguishing the Domains

To show how the three variables named earlier, the cost of error, the audience and opportunity cost, combine differently across functional domains, we first need a sharper tool than the word error itself. In everyday language, error hides two different things under one name: being wrong and being imprecise, and these need separating.

Correctness and precision are independent concepts. A text can be highly precise and still wrong, many decimal places around a false number, or correct but imprecise, stating a country's population in millions rather than the exact census figure. This distinction is not merely terminological; it has direct practical use, because much of what models return as an objection is a request to increase precision, not to correct a wrongness, and conflating the two is the source of much fruitless rewriting.

Returning to the distinction among domains, for that distinction to be of practical use, we need to know which domain the text in question belongs to, because the criterion for correct enough and precise enough shifts from one domain to another. We separate five domains, not by publication format but by the kind of claim the text makes:

●Scientific-empirical (a testable truth claim about the world);

●Philosophical-analytical (a conceptual, argumentative claim, independent of empirical testing);

●Public intellectual (a combination of both, for a non-specialist audience);

●Literary-fictional (no truth claim or formal argument; value lies in the reading experience);

●Visual-artistic (where form and work are one, and the content/form distinction dissolves).

On this classification, a text standing at the boundary of two domains is bound by the more stringent criterion of whichever domain its claim belongs to.

It should be noted here that this classification is a working tool of this article, not a citation from any official or established classification elsewhere. Its justification lies not in the authority of a source but in what the rest of this section demonstrates: that, in practice, different criteria hold for these five domains.

Even this distinction, however, is not enough on its own, because it does not answer which criterion each domain is exempt from and which it is not. Without this specification, the distinction collapses into a slogan that justifies any laxity. This article proposes adding a second axis, the type of error, and with two axes in place, instead of a mere list of domains, we get a table whose every cell can carry an independent verdict.

We can thus separate four types of error:

Factual error: a mismatch between a statement in the text and the state of the world, or a misleading simplification of it. Test: can this statement, in principle, be independently verified?

Inferential error: the conclusion does not follow from the premises: an ambiguous middle term, an unwarranted generalisation, a formal fallacy. Test: can one accept the premises and still reject the conclusion?

Attribution error: crediting an opinion to someone who did not hold it, or distorting their position. This differs from factual error, because beyond being wrong, it also has a moral dimension, bearing on whether it harms a third party.

Rhetorical-aesthetic error: here the failure is form's failure to perform its function: things like ambiguity, inconsistent tone, tedium and inefficient structure, all of which drain a piece of its functional value.

This error needs separating, though, from a different shortcoming that is easily confused with it: reconstruction fidelity. On this view, restating someone's position more broadly or with fewer details is not misattribution, so long as it does not invert their position or credit them with something they never said. It is a shortfall in granularity, and it falls under factual or inferential error, not attribution.

As you can see, we have deliberately left internal contradiction out of this list, because contradiction is the violation of one statement by another, and it can occur inside any of these four types: two inconsistent factual claims, two opposing lines of argument, two different attributions to the same thinker, or two clashing tones. We treat it as an independent requirement, one that no domain can waive.

That the criteria for these four types of error differ across domains is not a new claim; it rests on three independent supports:

The first is Kuhn, who, in the postscript he added to The Structure of Scientific Revolutions, proposed the concept of a disciplinary matrix and showed that evaluative criteria are internal to a community and cannot be derived from outside the scientific community (Kuhn, 1970).

The second is Ludwik Fleck, who, in Genesis and Development of a Scientific Fact, shows that the thought style of every thought collective determines what even counts as an error in the first place. Error, on this view, is not merely evaluation; it is perception (Fleck, trans. 1979).

The third is Bourdieu, who adds that every field has its own symbolic capital. Stringency in the scientific field and stringency in the literary field are two different currencies with no exchange rate between them (Bourdieu, 1975).

These views help us see clearly what amounts to a category error in models. A language model belongs to no field, and so it applies the criteria of every field at once, with no hierarchy among them. That is why it faults a public-intellectual piece for lacking quantitative methodology, and a literary text for lacking citations. This is a misattribution of criteria, not the discovery of a flaw.

We have now reached the point where we can formalise these cross-domain differences in the criterion for tolerating error using five degrees of tolerance, as follows:

Zero tolerance: a single instance destroys the whole text's credibility;

High: tolerance is high; the error occurs and may cause harm, but it is not serious or fatal, and it is remediable;

Medium: it is a weakness, not a violation;

Low: the writer has discretion;

Irrelevant: the criterion simply does not apply to that domain.

The table below shows how these five degrees of tolerance map onto the five functional domains. Here, though, we should note that these five degrees measure only a domain's tolerance for a shortfall in precision. Three other things, outright wrongness, misattribution and internal contradiction, belong to a different order. They damage a text's credibility, not its granularity, and unlike a shortfall in precision, damage to credibility is not offset in any domain by purpose or audience. A novel with no claim to represent reality still cannot contradict itself, and a painting whose form is decisive still cannot misrepresent someone's view. For this reason, these three sit outside the table: beneath every cell of it, equally and without exception, they are rejected.

Domain	Factual	Inferential	Attribution	Rhetorical
Scientific-empirical	Zero tolerance	Zero tolerance	High	Low
Philosophical-analytical	Medium	Zero tolerance	Zero tolerance	Medium
Public intellectual	High	High	High	Zero tolerance
Literary-fictional	Irrelevant	Irrelevant	Medium	Zero tolerance
Visual-artistic	Irrelevant	Irrelevant	Low	Decisive

As we can see from the table:

In the scientific-empirical row, both factual and inferential errors carry zero tolerance, because the entire value of a scientific statement lies in others being able to rely on it. Data that turns out wrong

even once removes the whole paper from the citation cycle. In this row, attribution is high, not zero tolerance, because in an empirical text the literature review serves an auxiliary role, and a broader, less detailed but honest restatement of someone else's view, though a shortcoming, does not invalidate the experimental result.

Rhetoric is low, and the evidence bears this out: papers with very poor prose are routinely published and cited all the same.

In the philosophical-analytical row, inference carries zero tolerance, because in analytic philosophy the argument is the product itself, not a path towards it; an inferential error here is the equivalent of fabricating data in empirical science. Attribution also carries zero tolerance, because a philosophical text usually proceeds by reconstructing others' positions, and if the reconstruction is wrong, any critique built on it collapses into a straw man and the whole argumentative move loses its validity.

Factual accuracy is medium, because in a thought experiment the empirical truth of the premises is beside the point. Rhetoric is medium, because in a conceptual text verbal ambiguity quickly turns into conceptual ambiguity.

In the public-intellectual row, no cell is unimportant, and this point is often overlooked: this domain is not more lenient; it has different criteria. Rhetoric here carries zero tolerance, because the reader has no obligation to keep reading; an unintelligible text has failed completely in this domain, even if every individual statement in it is true. Factual accuracy and inference are high, not zero tolerance, not because a false statement is permitted here, which it never is, but because imprecise simplification is an intrinsic part of this genre, and separating legitimate simplification from wrongness is a matter of degree, not a binary.

In the literary-fictional row, factual accuracy and inference are irrelevant, and these cells carry the greatest risk of misunderstanding. This does not mean a novel may contain anything false. It means a novel, as such, makes no claim to represent reality or to formal argument that could be measured against the outside world or against logical validity. This exemption does not extend to internal contradiction, however. As stated earlier, a mismatch between two parts of a story remains, unconditionally, an instance of error, regardless of whether the novel claims to represent reality. Attribution is medium, because representing real people in fiction carries a moral limit that never fully disappears.

In the visual-artistic row, the rhetoric column is labelled decisive rather than zero tolerance, because in this domain form is not merely an evaluative criterion; it is the work itself. The content/form distinction, which held across the previous four rows, collapses here, and this collapse is exactly why applying textual criteria to a work of art is a category error.

Beyond these, three further conclusions emerge from the table, none visible in any single row or column:

One: no column carries zero tolerance across every row, meaning no universal criterion exists for the degree of required precision. This claim is limited to precision, not to wrongness or contradiction.

Two: no row is lenient across every column. Every domain carries at least one zero-tolerance or decisive cell. So distinguishing domains does not licence laxity; it relocates where stringency applies, rather than reducing it.

Three: the attribution column shows the least spread, ranging from low to zero tolerance, with no irrelevant cell at all. This shows that attribution error comes closest to a cross-domain criterion, probably because it is the only error type that, beyond its epistemic dimension, carries a moral one and harms a third party. If we were to keep only one criterion out of all of them, it would be to represent others' views correctly.

We now have a complete tool for judging any given objection: what type of error it is, which functional domain it applies to, and whether it exceeds that domain's permitted threshold. This tool alone, however, does not explain why the demand for unlimited precision, despite all these distinctions, keeps recurring, insistently. That question forms the theoretical core of this article, and we turn to it in the next section.

---

6. The Demand for Precision as a Technique of Ignorance Production

So far, the argument has been that unlimited stringency is inefficient from the standpoint of cognitive economy: it costs, and its returns diminish. This argument, though correct, is weaker than it needs to be, because it only makes the demand for more precision look costly, not problematic.

Inefficiency is a practical flaw. If a writer is willing to pay its price, no answer seems left to give. The claim of this section is stronger: the demand for ever more precision is, beyond mere inefficiency, a mechanism for producing ignorance. This claim is not metaphorical or analogical. The agnotology literature has already identified, named and documented this technique, though in an entirely different context: public policy and industrial strategy. What this section does, then, is show the structural identity between that research finding and what today's writer, holding a tool that has made critique costless, now inflicts on themselves.

Within the agnotology tradition, the watershed documenting how an ever-growing demand for precision can serve as a mechanism for producing ignorance is the body of research on the tobacco industry's strategy in the second half of the twentieth century. The central finding of that research, formulated most clearly in Merchants of Doubt (Oreskes & Conway, 2010), is that the industry's strategy was not to deny scientific findings. Outright denial is a vulnerable position; it can be refuted, and every refutation discredits the denier. The effective strategy was something else: an apparent acceptance that the issue mattered, paired with a constant demand for more research, more precise data and greater certainty before any action. Unlike denial, this position is irrefutable, because no body of evidence is ever so complete that no gap can be found in it. The most famous document of this strategy is an internal tobacco-industry memo that explicitly describes doubt as the company's manufactured product, a product deployed against the body of settled facts in the public mind.

Freudenburg, Gramling and Davidson raised this strategy from a historical case to a conceptual level, naming it Scientific Certainty Argumentation Methods, a formulation useful for our discussion (Freudenburg, Gramling, & Davidson, 2008). On this view, the industry's strategy raises the burden of proof so high that no action becomes permissible, and it does this not by violating scientific standards but by applying those same standards without any limit. David Michaels, writing from his own position within regulatory policy, has documented this same mechanism in the chemical and pharmaceutical industries (Michaels, 2008).

This technique has three overt features, all three important for our discussion:

First, this method is parasitic. It sets no new or false standard; every one of its demands, controlling for confounding variables, adequate sample size, replicability, is a legitimate and correct demand of scientific methodology. Its paralysing effect comes not from the wrongness of these demands but from applying them without boundary and without end.

Second, the technique is immune to critique. Whoever demands more precision stands in a position of defending science, not opposing it, so any objection to them looks like opposition to scientific rigour itself. This rhetorical asymmetry is part of what makes the technique effective.

Third, the technique aims at delay, not refutation. The goal here is not to prove the opposing claim wrong; it is the unlimited postponement of the moment of decision, and the result, in practice, is the same result as successful denial, without bearing its costs.

With this laid out, we can now state our claim: the structure of the three features above maps, without remainder, onto the situation of a writer who exposes their text to the unending critique of language models. To show this claim holds, we can check the correspondence feature by feature:

Being parasitic: the objections a model returns are often not, in themselves, wrong. 'This claim needs more empirical evidence', 'this concept is not sufficiently differentiated', 'the research literature in this area is not fully covered': taken as isolated statements, all three are true of any text, and this very universality makes them epistemically worthless. An objection that applies to every text conveys no information about this particular text. In Bateson's terms, it is not a difference that makes a difference.

Immunity to critique: a writer who rejects the list of objections is placed in a difficult position; they appear to be dodging critique. This asymmetry, it seems, is even sharper here than in the policy case, because the other party is not a suspect stakeholder but a tool with no interest in delay whatsoever, and it therefore appears neutral.

Aiming at delay: the practical outcome is the same in both cases: the text is not published, the decision is not made, the intervention does not happen. So the difference lies only in scale, not in structure.

So far, we have reached an analogy, but we should also note a difference that can make the article's claim more serious still.

In the classic pattern, an external, interested agent deliberately deploys the technique. The goal is defined, the interest is defined, and resistance is therefore possible: one can identify the agent, expose their motive and reread their claims in light of their interest. Indeed, the entire exposé literature of agnotology is built on this very possibility.

In the situation this article discusses, however, no interested agent exists. The language model gains nothing from delaying publication. The writer clearly loses out. No one deploys this technique, and yet the behaviour keeps recurring and keeps reaching the same result.

This phenomenon can, in the precise sense of the word, be called an instance of emergent ignorance: a pattern with every feature of an organised technique of ignorance production, arising from the structural interaction of two neutral mechanisms, with no central producer. Those two mechanisms are, on one side, the fact that supplying critique has become costless, and on the other, the writer's internalised norm of self-criticism, which is, in itself, healthy and useful, not a sign of weakness.

The subtle point here is that neither of these two, on its own, is harmful. An abundant supply of critique, absent a norm of accepting critique, is simply ignored, and a norm of self-criticism, under

conditions of limited critique supply, is exactly what makes research possible. It is the combination of the two that produces paralysis, and this is precisely what marks it as emergent: a property that appears at the level of the system while being absent from either component beforehand.

This is why we might call this phenomenon self-imposed ignorance through stringency, a phenomenon whose distinguishing feature is that victim and agent are one and the same.

This diagnosis, on its own, offers no path to action, because knowing which mechanism traps us does not tell us what to do about the next objection a model raises. For that, we need a practical criterion, and the next section takes up this task.

---

7. A Practical Criterion

So far, we have shown that the demand for unlimited precision is unwarranted. But this, too, is a negative conclusion, and on its own it is not merely useless but dangerous. A writer who takes away only this conclusion has acquired a tool for dismissing any critique whatsoever.

In Section 3, drawing on a philosophical tradition, we examined a positive criterion. Here we seek a practical criterion, one that can be applied at the moment of facing a specific objection and tell us whether to accept it or reject it.

Our criterion needs three properties: it must be usable, not merely correct; it must cut both ways, able to reject as well as accept; and it must apply to its own user as well.

Before anything else, we must set out three minimal conditions: three categories of objection that must always be accepted and corrected, ahead of the four questions to come and regardless of the text's functional domain:

One: attribution error. Misquoting someone else's view is impermissible in any field, as we saw reading across the table in Section 5, because attribution error is the only error type that carries a moral dimension beyond its epistemic one: it harms a third party who is absent from the conversation and cannot defend themselves. Weighing costs and benefits makes no sense here, because someone else pays the cost.

Two: internal contradiction. Two parts of a text that violate each other defeat the text in the very thing it claims to do, regardless of domain. Even a literary text, exempt from the criterion of factual accuracy, is not exempt from internal coherence.

Three: a false factual statement, wherever the text claims to represent reality. The final qualifier here matters, because fiction is exempt from this criterion only because it makes no such claim. Any text that presents a statement as fact is answerable for its correctness, even if it is a literary or artistic text.

These three are minimal conditions and admit no exception. What follows applies only above this minimal floor.

Once past these conditions, the following four questions are asked in order:

First question: what type is this objection?

The objection must be placed in one of Section 5's four categories: factual, inferential, attributional or rhetorical. This question is more decisive than it looks, because most objections phrased as factual are in fact rhetorical. 'This claim is not documented' may mean the claim is wrong, or it

may mean the reader is not convinced. The two demand entirely different responses: the first calls for correcting the statement, the second for restructuring the argument.

Second question: which field does this objection come from? Is the standard behind the objection the standard of the field this text belongs to?

This question directly answers the phenomenon analysed in Section 5. Faulting a conceptual paper for lacking methodology, faulting a literary text for lacking citations, and faulting an empirical paper for dry prose are all of this kind: not wrong, but misplaced.

This is the most dangerous question on the list, because the easiest way to dodge a legitimate criticism is to declare that the critic has misunderstood the field. To guard against this risk, I add the following condition:

Rejecting an objection on the ground that it is misplaced by field is permissible only when the writer can name the text's field explicitly and in advance, exactly as this article's introduction did. If a writer cannot say which field their text belongs to, they have no right to call any criterion out of field.

Third question: is this objection bounded? Does it specify its own condition of satisfaction? Can one picture what change would count as resolving it?

For example, 'the citation on page 12 is wrong' is bounded, but 'more evidence is needed' is unbounded, unless it says what evidence and how much. An unbounded objection, in the terms of Section 6, reproduces the same paralysing structure and must either be bounded or set aside.

From this article's perspective, the practical way to bound an objection is to ask the critic, or the model: what specific change would you count as resolving this? If no specific answer comes, the objection was never really an objection at all.

Fourth question: does fixing this objection change the outcome?

This is Bateson's criterion. If, after fixing the objection, the text's central claim remains what it was, the objection is of the immunising kind, not the corrective kind. A text made more immune to a hypothetical critic, while saying nothing new, has not improved. It has only grown heavier.

This is the point to introduce the concept of defensive writing, by analogy with defensive medicine: an action or intervention whose goal is not to improve the outcome but to reduce the agent's own exposure to being called to account. In medicine, this defensive approach produces unnecessary tests and wasted cost. In writing, it produces hedging clauses, weakening qualifiers and ritual citations, none of which make the text more correct, all of which make it slower and more obscure. What the two concepts share is that defensive action is not merely useless; it damages the final quality.

These four questions, of course, are no new invention. Each is a workable form of a judgement already argued for independently, as follows:

The first question comes from the Aristotelian distinction among different natures of subject matter, formalised in Section 5 into four categories of error.

The second question derives from Bourdieu's field theory and Kuhn's disciplinary matrix, and answers directly to the model's fieldlessness. The third question comes from the analysis of scientific certainty argumentation methods, which identifies unboundedness as the defining mark of a paralysing demand. The fourth question is simply Bateson's definition of information, and, at the level of decision, Simon's criterion of satisficing.

What these four questions actually add is order. Error type is asked before field, because the three floor categories, attribution, internal contradiction and false factual statement, are exempt from field-based judgement, and if the order were reversed, a writer could dismiss an attribution error by claiming it is out of field. Boundedness comes before effect on outcome, because an unbounded objection cannot be evaluated for its effectiveness at all.

On this basis, a text is ready for publication when it is not wrong, does not contradict itself, has not misrepresented anyone's view, is not needlessly obscure, and carries no decisive shortcoming by the standards of its own field, not when no hypothetical critic could raise any objection to it at all, because that goal is never achieved, and a text that waits for it will never be published. This is the same conclusion as the Agrippan trilemma in Section 1, this time not as a theoretical problem but as a description of the writer's desk.

Finally, one specific practical consequence of this criterion for working with models is worth stating. A common mistake is to ask a model to criticise harshly. Because of the mechanism explained in Section 2, this request always produces an unbounded list that leaves the text's signal indistinguishable from the prompt's noise. The right request has a different structure: state the text's field, specify the error category wanted, and ask for a boundary.

For example: this text is a philosophical essay for a general reader; list only attribution errors and internal contradictions, and for each one, say what specific change would resolve it.

The difference between these two requests is the difference between an endless list and a finite, actionable one, and this, at the practical level, is exactly what this entire article does at the theoretical level.

---

8. Anticipated Questions, Objections, and Open Horizons

It is obvious that this article's theoretical argument, and its approach and outlook, can be criticised and objected to. In this section we discuss some likely objections, and where we lack a clear answer, we say so explicitly.

One. When no global body sets a standard for precision and stringency in writing, why should anyone accept this article's standard?

The answer is that the question is misframed from the start. We have not put our own standard in place of a global body's standard; for writing, no such body ever existed to issue one. What this article calls correctness and precision are working concepts, defined to advance this very discussion, not a citation of a standard set somewhere else. The validity of this definition should not be measured by searching for the authority that issued it, but by whether it works over the course of the article: whether it lets us see something that would otherwise go unseen.

Two. This article's argument hands any weak writer a tool for dismissing any legitimate criticism. In other words, if a demand for more precision is treated as suspect from the outset, no criterion is left for correcting a text, and the result is exactly the relativism this article claims to oppose.

The answer is that what constitutes a technique of ignorance production requires three conditions at once:

- Unboundedness: the demand states no threshold whose reaching would satisfy it;
- Fieldlessness: the demand draws on the criteria of a field the text does not belong to;

●No effect on outcome: resolving the objection leaves the text's central claim unchanged and merely makes it more immune to a hypothetical critic.

A criticism that has none of these three, pointing to a wrong citation, an internal contradiction, an inference that does not follow from its premises, is a criticism that must always be accepted, and drawing this distinction between the two classes of critique, legitimate and illegitimate, is exactly the work whose absence created the current situation. A writer who treats every objection with equal seriousness is acting just as much without a criterion as a writer who dismisses them all.

Three. Where has the claim that harsh criticism and undeserved praise are two forms of a single mechanism actually been tested?

In answer, this claim is inferred from existing research on model sycophancy, and part of that research, feedback sycophancy tests, comes close to testing this claim directly. More recent research (Cheng et al., 2026) confirms this broader dependency, that a model's output is a function of user expectation and satisfaction more than of content accuracy, at a larger scale, with more than two thousand participants. But the more specific form of this article's claim, varying the prompt itself between 'criticise harshly' and 'approve' on a single fixed text, and testing whether the two outputs trace back to a shared source, has not yet been tested directly. Until exactly this test is run, the claim is a hypothesis with close empirical support, not a direct finding.

Four. What this article calls ignorance production might really just be a waste of attention, or perfectionism.

We should distinguish between ignorance production in the precise agnotology sense and mere delay or wasted time. To establish the stronger claim, one would need to show that this process actually weakens justified belief or systematically distorts decision-making, not merely that it delays publication. This article has not produced such evidence; it has shown only structural isomorphism.

Five. Doesn't the claim that no interested agent exists conflict with the economics of AI platforms?

One might object that, under pay-per-use business models, where every extra message generates direct revenue, the model provider profits from long cycles of revision. If so, this weakens Section 6's claim that agentlessness is what distinguishes this phenomenon from the classic technique of doubt production, the very thing that elevated it from mere structural resemblance to a genuine instance of the same category, since a real interested agent, however inconspicuous, would then exist.

This objection holds more force for so-called pay-per-use models than for unlimited subscription models, where a long revision cycle is a net computational cost to the provider, not a profit. This article has not adjudicated, and does not adjudicate, between the two readings, emergent and agentless ignorance, or a reproduction of the classic technique with an agent who is simply less visible under certain business models.

Conclusion

This article has shown that the question in its title, how much precision is really enough, was wrongly framed from the start. Precision itself was never in dispute. The right question is how much stringency is enough.

Stopping revision, too, has never meant finding an objective point. It has always been a conventional decision, one that external circumstances used to make on the writer's behalf. Today, though, with those constraints gone, a decision that was always implicit must now be made explicitly. In the absence of that explicitness, an unlimited supply of critique and the norm of self-criticism together form a pattern that functions as a technique of ignorance production, without any agent deploying it.

So what has changed is not the nature of precision; it is the responsibility for setting the threshold, which has passed from external circumstances to the writer. Writing is no longer scarce. What has become scarce is the capacity to say this sentence: this is enough, and I can say why.

From this view, stopping revision and the pursuit of optimisation is not an epistemic failure. It is something every writer has always done. The only difference today is that it must be done consciously, knowing for what purpose, in what place, and what degree of stringency is enough.

References

Albert, Hans. Treatise on Critical Reason. Translated by Mary Varney Rorty. Princeton University Press, 1985.

Aristotle. Nicomachean Ethics. Translated by W. D. Ross, revised by Lesley Brown. Oxford University Press, 2009.

Bateson, Gregory. Steps to an Ecology of Mind. Chandler Publishing, 1972.

Borges, Jorge Luis. “On Exactitude in Science.” In Collected Fictions, translated by Andrew Hurley. Penguin, 1998.

Bourdieu, Pierre. “The Specificity of the Scientific Field and the Social Conditions of the Progress of Reason.” Social Science Information 14, no. 6 (1975): 19–47.

Cheng, Myra, Cinoo Lee, Pranav Khadpe, Sunny Yu, Dyllan Han, and Dan Jurafsky. “Sycophantic AI Decreases Prosocial Intentions and Promotes Dependence.” Science 391, no. 6792 (2026): eaec8352.

Cole, Stephen, Jonathan R. Cole, and Gary A. Simon. “Chance and Consensus in Peer Review.” Science 214, no. 4523 (1981): 881–886.

Collins, Harry. Changing Order: Replication and Induction in Scientific Practice. University of Chicago Press, 1985.

Duhem, Pierre. The Aim and Structure of Physical Theory. Translated by Philip P. Wiener. Princeton University Press, 1954.

Fleck, Ludwik. Genesis and Development of a Scientific Fact. Translated by Fred Bradley and Thaddeus J. Trenn. University of Chicago Press, 1979.

Freudenburg, William R., Robert Gramling, and Debra J. Davidson. “Scientific Certainty Argumentation Methods (SCAMs): Science and the Politics of Doubt.” Sociological Inquiry 78, no. 1 (2008): 2–38.

Goffman, Erving. The Presentation of Self in Everyday Life. Doubleday, 1959.

James, William. The Will to Believe and Other Essays in Popular Philosophy. Longmans, Green, and Co., 1896.

Kuhn, Thomas S. The Structure of Scientific Revolutions, 2nd ed., with Postscript. University of Chicago Press, 1970.

Michaels, David. Doubt Is Their Product: How Industry’s Assault on Science Threatens Your Health. Oxford University Press, 2008.

Oreskes, Naomi, and Erik M. Conway. Merchants of Doubt: How a Handful of Scientists Obscured the Truth on Issues from Tobacco Smoke to Global Warming. Bloomsbury Press, 2010.

Peirce, Charles Sanders. “Note on the Theory of the Economy of Research.” In United States Coast Survey Report, 1879.

Peters, Douglas P., and Stephen J. Ceci. “Peer-Review Practices of Psychological Journals: The Fate of Published Articles, Submitted Again.” Behavioral and Brain Sciences 5, no. 2 (1982): 187–195.

Popper, Karl. The Logic of Scientific Discovery. Hutchinson, 1959. (Logik der Forschung, 1934.)

Quine, W. V. O. “Two Dogmas of Empiricism.” The Philosophical Review 60, no. 1 (1951): 20–43.

Rescher, Nicholas. Cognitive Economy: The Economic Dimension of the Theory of Knowledge. University of Pittsburgh Press, 1989.

Sextus Empiricus. Outlines of Scepticism. Translated by Julia Annas and Jonathan Barnes. Cambridge University Press, 2000.

Sharma, Mrinank, et al. “Towards Understanding Sycophancy in Language Models.” arXiv preprint, 2023.

Simon, Herbert A. “A Behavioral Model of Rational Choice.” The Quarterly Journal of Economics 69, no. 1 (1955): 99–118.

Simon, Herbert A. “Designing Organizations for an Information-Rich World.” In Computers, Communications, and the Public Interest, edited by Martin Greenberger, 37–72. Johns Hopkins University Press, 1971.
