

# When the Proxy Breaks

<https://gerolds.github.io/textbook/textbook/posts/writers-room-for-expertise/>

# Contents

1. The proxy collapse pattern
2. Historical evidence
  - 2.1 Photography and the death of “faithful depiction”
  - 2.2 Synthesizers, Auto-Tune, and the Luddites
3. The economic unlock
  - 3.1 The writers’ room model
4. The trap, and what replaces the old proxy
5. Closing questions
6. Appendix: Sources and Extended Reading
  - 6.1 On Proxy Collapse and Measurement
  - 6.2 On Photography and Artistic Legitimacy
  - 6.3 On Music Technology and Authenticity Discourse
  - 6.4 On the Luddites and Skilled Labor
  - 6.5 On the Economics of Niche Content
  - 6.6 On Epistemic Signals and Trust
  - 6.7 On Writers’ Rooms and Collaborative Creation
  - 6.8 On AI and Writing (Contemporary Discourse)
  - 6.9 Extended Discussions

# When the Proxy Breaks

**When a craft is hard, its surface markers become a credibility shortcut.** Polish implies effort; effort implies care; care implies truth or taste. Audiences learn to treat “looks professional” as “probably worth my attention” because the cost of producing that surface deters most fakers. The proxy works until a technology collapses the cost. Then the correlation between surface and merit breaks, and audiences experience the break as betrayal.

The backlash against LLM-assisted writing follows a repeating pattern: technology cheapens a craft surface, audiences lose a trusted shortcut, and the disruption gets framed as moral crisis rather than measurement problem. The resolution is predictable: value migrates from the newly-cheap surface to the decisions that remain expensive, and communities stabilize around signals that reward outcomes over tool purity.

Below we show that pattern through historical cases, then illustrate the consequences with a silver lining: **the economics of expert publishing change.**

---

## 1. The proxy collapse pattern

A proxy is a stand-in measurement. You cannot directly observe an author’s care, so you observe their polish instead. You cannot verify every claim, so you use fluency as evidence that someone did the work. Proxies let readers allocate attention without inspecting every artifact from first principles.

Proxies fail when their cost structure changes. If polish becomes cheap, polish stops filtering for care. The surface that once signaled “someone invested here” now signals only “someone had access to the tool.” The audience’s heuristic misfires. They feel deceived, even when no deception occurred.

**This is a measurement problem, not a moral one.** The author using the new tool is not cheating; they are producing the same surface at lower cost. The audience is not irrational for feeling betrayed; their shortcut genuinely stopped working. But framing

the disruption as cheating obscures what needs to happen: the audience must learn a new test.

The moral framing persists because it protects status hierarchies. If you built a career on a scarce craft, you have an incentive to keep the old proxy socially enforced after it is empirically broken. “Authentic” becomes a gatekeeping term rather than a quality descriptor. The fight shifts from “does this work?” to “did you earn it the hard way?” Once that happens, the debate is about preserving moats, not evaluating outcomes.

**Claim 1:** When a proxy becomes cheap, the first reaction is moral outrage that disguises an evaluation upgrade the audience must eventually make.

---

## 2. Historical evidence

The pattern has played out across media for two centuries.

### 2.1. Photography and the death of “faithful depiction”

Before cameras, faithful depiction required years of training. Realism was expensive, so realism became a proxy for artistic seriousness. Photography mechanized that surface overnight.

The backlash was immediate and moralized. Baudelaire’s 1859 essay “The Modern Public and Photography” argued that mechanical reproduction would corrupt art by letting the masses bypass cultivated judgment. Critics insisted photography could not be art because a machine produced the image. This is proxy defense: the argument is not “these images are bad” but “these images did not cost enough to count.”

The argument collapsed because it confused difficulty of production with value of result. Once photography became common, mere realism stopped differentiating. Artistic value migrated to choices still requiring judgment: framing, timing, subject selection, relationship to reality. Photography did not kill painting; it relocated artistry from brushwork to decisions.

LLM tools force the same relocation: from typing to claim-selection.

## 2.2. Synthesizers, Auto-Tune, and the Luddites

Music technology shows the same pattern. When synthesizers entered popular music, bands signaled “no synthesizers were used on this album” on packaging, not to describe their sound but to preserve the authenticity proxy. Auto-Tune attracted the complaint “anyone can sing now,” which is not an aesthetic objection but fear of proxy pollution. Once these tools stabilized, artistry shifted to production choices. “Synthetic” stopped being a defect and became a palette.

The Luddites were not irrational technophobes; they were skilled workers responding to machines that threatened wages and bargaining power. The rhetoric around “inferior machine-made goods” served the same function as “soulless AI writing”: it preserved the old proxy (handcraft = quality) against an empirical challenge. Quality systems eventually emerged not by banning machines but by formalizing outcome standards independent of production method.

**Claim 2:** The biggest backlashes happen when technology destroys a shortcut for detecting effort, not when it destroys the underlying capability.

In each case, the old proxy was replaced by a new signal of skill once the tool normalized.

---

## 3. The economic unlock

Every proxy collapse shares a second-order effect: it lets new producers enter markets that were previously gated by the cost of the surface, not the cost of the substance. Photography let people without painting skills record images. Synthesizers let composers without orchestras produce full arrangements. The printing press let scholars without scriptoria distribute ideas. The gate was never “do you have something worth showing” but “can you afford the surface that proves you are serious.” When the surface gets cheap, the question finally reduces to whether you have something worth showing.

For writing, the gated resource was prose assembly: turning expertise into readable text. That gate kept advanced, niche content non-economic.

As a consequence, the internet skews towards beginners because beginner content amortizes. A tutorial reaches millions; an advanced essay reaches hundreds. The same hour of writing labor yields vastly different returns, so creators rationally target breadth. Advanced knowledge stays locked in conversations, internal memos, and papers nobody reads. Scarcity of translation labor, not scarcity of insight, explains why searching your own field returns explanations written for outsiders.

When translation cost drops, the calculus shifts. A piece reaching 500 readers was never worth 20 hours of labor; at 2 hours the math changes. Series and deep dives appear because follow-up writing is no longer a second job. Experts reach audiences directly instead of waiting for journalists to translate their work (often badly).

**Claim 3:** LLMs remove a moat that kept expert synthesis non-economic. The bottleneck was never “experts have nothing to say”; it was “the labor of saying it isn’t worth the audience size.”

### 3.3. The writers’ room model

The name comes from television: a writers’ room is a collaborative structure where multiple contributors shape a single output. One person might pitch the story, another breaks it into scenes, another punches up dialogue. The showrunner holds the vision; the room provides labor and iteration. No one pretends the showrunner typed every word, and no one thinks that diminishes the result.

Applied to LLM-assisted writing, the model works like this: the human brings the claims, constraints, and accountability. The model brings drafting labor, restructuring, and variant generation. The human decides what matters, what is true, and what to publish. The model handles the assembly work that turns raw thinking into readable prose.

A practical workflow: First, capture. Dump raw notes, bullet reasoning, constraints, examples, counterexamples. Externalize the mental model without requiring polish. Second, compile. The model turns capture into coherent narrative, organizing structure and generating draft prose. This is where translation labor would be paid; the model pays it instead. Third, verify. The expert prunes, tightens, and signs. Cut what is wrong. Sharpen what is vague. Confirm the final version says what you mean.

This separation clarifies what delegation means. Delegation is not abdication. Insight lives in the selection of what matters and the boundary of what is defensible; typing is not the value. If the piece is wrong, the author is responsible. The model does not get blamed, and “the AI wrote it” is not a defense.

---

#### 4. The trap, and what replaces the old proxy

Commoditized assembly also commoditizes the appearance of competence. Shipping plausible nonsense gets easier. And here the analogy meets its limit: **a camera cannot fabricate a false claim; an LLM can.** Photography mechanized a surface without generating propositional content. LLMs generate statements that can be true or false, supported or unsupported. This is not merely a cheaper surface; it is a machine that produces the *appearance* of reasoning. The failure mode is not just “low-effort work gets published” but “fluent falsehoods get published.”

The proxy-collapse frame still applies, but the evaluation upgrade is steeper. When photography arrived, audiences learned to judge framing and intention rather than brushwork. When LLMs arrive, audiences must learn to judge *claims and evidence* rather than fluency. The skill demanded is epistemic discipline, not aesthetic discernment: a harder ask, and one many readers have been avoiding.

**Claim 4:** When the polished surface becomes abundant, visibility of constraints and accountability define quality.

The old regime throttled output through writing friction. That friction incidentally filtered for persistence, not because friction *caused* quality, but because it correlated with the kind of person

willing to push through it. Now the supply of fluent text rises for both sincere experts and people who merely sound like them. What becomes scarce: **attention and verification.**

When old proxies lose signal, new shortcuts replace them.

*Reputational signals:* track record, peer networks, visible correction behavior. *Structural signals inside the work:* explicit claims, stated scope, separation of observation from inference, concrete falsifiers.

### **Example (before and after):**

*Vague:* “LLMs are transforming how experts share knowledge, making it easier to publish advanced material that wouldn’t have been economically viable before.”

*Constrained:* “LLMs reduce the labor cost of turning expertise into readable prose (claim). This makes essays viable for smaller audiences, hundreds instead of millions (scope). The assumption is that the expert’s bottleneck was assembly, not insight; if the bottleneck is actually verification or originality, the tool helps less. This claim would be weakened if LLM-assisted expert output remained flat despite tool availability.”

The constrained version is harder to write even though it is longer. It requires the author to know what they are claiming, where the boundaries are, and what would change their mind. An LLM can draft prose; it cannot supply those commitments without the author’s input.

Once authors internalize constraint density, “LLM-assisted” stops being a credibility category and becomes a workflow footnote.

Disclosure becomes technical clarification, not confession:

*“Drafting assistance: Claude. All claims mine; errors my responsibility.”*

**Claim 5:** New equilibria form when communities stop policing tools and start rewarding verifiable intention and results.

## **5. Closing questions**

- Which proxy do you use to infer “this author knows what they’re doing”? What would you do if that proxy became free tomorrow?

- What verification step do you consistently skip because the old proxy used to cover for it?
- 

## 6. Appendix: Sources and Extended Reading

### 6.4. On Proxy Collapse and Measurement

- **Goodhart's Law** – “When a measure becomes a target, it ceases to be a good measure.” The proxy collapse pattern is a special case: the measure does not become a target, but becomes cheap to satisfy. Marilyn Strathern's generalization (“When a measure becomes a target, it ceases to be a good measure”) and the original formulation by Charles Goodhart (1975) on monetary policy both illuminate why surface signals degrade.
- **Campbell's Law** – Donald T. Campbell, “Assessing the Impact of Planned Social Change” (1979). The more a quantitative indicator is used for decision-making, the more it corrupts the process it was intended to monitor. Relevant to understanding why polish stopped working as a quality filter.

### 6.5. On Photography and Artistic Legitimacy

- **Charles Baudelaire**, “The Modern Public and Photography” (1859), in *The Mirror of Art*. The primary source for photography-as-corruption arguments. Baudelaire's moral framing is remarkably parallel to contemporary AI discourse.
- **Walter Benjamin**, “The Work of Art in the Age of Mechanical Reproduction” (1935). The canonical analysis of how reproduction technology changes art's function and aura. Benjamin is more nuanced than Baudelaire—he sees political potential in the collapse of scarcity.
- **Susan Sontag**, *On Photography* (1977). Essays on how photography changed not just art but perception itself. Sontag traces the way the medium creates new ways of seeing rather than simply recording old ones.

## 6.6. On Music Technology and Authenticity Discourse

- **Trevor Pinch and Frank Trocco**, *Analog Days: The Invention and Impact of the Moog Synthesizer* (2002). Documents the authenticity debates around synthesizers and the “no synthesizers” album disclaimers of the 1970s.
- **Simon Reynolds**, *Retromania: Pop Culture’s Addiction to Its Own Past* (2011). Includes discussion of how technology disruptions get absorbed into nostalgia cycles and authenticity performance.
- **Dave Tompkins**, *How to Wreck a Nice Beach: The Vocoder from World War II to Hip-Hop* (2010). History of vocal processing technology and the recurring “that’s not real singing” critique.

## 6.7. On the Luddites and Skilled Labor

- **E.P. Thompson**, *The Making of the English Working Class* (1963). The revisionist history that recovered Luddism from “irrational technophobe” caricature. Thompson shows skilled workers responding strategically to machines that threatened wages and autonomy.
- **Eric Hobsbawm**, “The Machine Breakers” (1952), *Past & Present*. Shorter treatment of machine-breaking as rational collective bargaining rather than anti-technology panic.

## 6.8. On the Economics of Niche Content

- **Chris Anderson**, *The Long Tail* (2006). The optimistic case for niche content economics online. Useful as a comparison: Anderson predicted discovery would solve distribution, but attention scarcity reasserted itself.
- **Yochai Benkler**, *The Wealth of Networks* (2006). On peer production and how reduced coordination costs change what gets made. The “expert publishing” unlock is a specific instance of Benkler’s general thesis.
- **Clay Shirky**, *Here Comes Everybody* (2008) and *Cognitive Surplus* (2010). On what happens when participation costs drop. Shirky’s analysis of amateur production applies, with modifications, to expert production when assembly costs drop.

## 6.9. On Epistemic Signals and Trust

- **Harry Frankfurt**, *On Bullshit* (1986/2005). The distinction between lying (caring about truth and inverting it) and bullshit (not caring about truth at all). Relevant to why fluent falsehoods are a distinct problem from mere error.
- **Philip Tetlock**, *Expert Political Judgment* (2005) and *Superforecasting* (2015, with Dan Gardner). On what actually predicts expert accuracy (hint: not credentials or fluency). The “constraint density” argument draws on Tetlock’s findings about hedged, specific claims outperforming confident broad ones.
- **Julia Galef**, *The Scout Mindset* (2021). Popular treatment of epistemic practices that resist motivated reasoning. The “what would change your mind” prompt is central to Galef’s framework.

## 6.10. On Writers’ Rooms and Collaborative Creation

- **Brett Martin**, *Difficult Men* (2013). History of the prestige TV era, including how writers’ rooms actually function and how authorship is distributed.
- **Pamela Douglas**, *Writing the TV Drama Series* (4th ed., 2018). Practitioner’s guide to room structure, including how showrunners maintain voice across multiple writers.

## 6.11. On AI and Writing (Contemporary Discourse)

- **Ted Chiang**, “ChatGPT Is a Blurry JPEG of the Web” (2023), *The New Yorker*. Influential framing of LLMs as lossy compression. Useful for understanding what the model “knows” and what it approximates.
- **Emily Bender et al.**, “On the Dangers of Stochastic Parrots” (2021), *FAccT*. The paper that introduced “stochastic parrots” as a frame for large language models. Important for understanding the “fluent falsehoods” problem.
- **Ethan Mollick**, *Co-Intelligence: Living and Working with AI* (2024). Practical treatment of LLM integration into knowledge work, including writing workflows.

## 6.12. Extended Discussions

The proxy-collapse pattern connects to several broader conversations:

**Signaling theory in economics** – Michael Spence’s work on job-market signaling (Nobel Prize, 2001) explains why costly signals work: they separate types when cheap signals cannot. Proxy collapse is what happens when a previously costly signal becomes cheap.

**Mimetic theory** – René Girard’s work on imitation and rivalry helps explain why the moral framing persists: when everyone can produce the surface, the surface stops conferring distinction, and those who built status on it fight to preserve the old hierarchy.

**Craft vs. art debates** – The proxy collapse pattern recurs wherever a previously difficult technique becomes automated. See also: CNC machining and woodworking, digital typography and lettering, CGI and practical effects. Each debate follows the “is it cheating?” → “does it work?” → “what new skills matter?” arc.

---

*Drafting assistance: Claude. All claims mine; errors my responsibility.*