Conceptual Responsibility: Structural Constraints on Idea Transmission and the Ethics of Intellectual Communication

Boris Kriger[1,2]

[1]Information Physics Institute, Gosport, Hampshire, United Kingdom,
boris.kriger@informationphysicsinstitute.net
[2]Institute of Integrative and Interdisciplinary Research, Toronto, Canada,
boriskriger@interdisciplinary-institute.org

Abstract

Ideas transmitted through social networks systematically degrade. This paper establishes this claim through converging evidence from three independent domains: experimental psychology (Bartlett's serial reproduction paradigm and subsequent replications), cultural evolution theory (iterated learning models), and structural analysis drawing on information-theoretic concepts as illuminating analogies. We argue that cognitive and communicative constraints, while not identical to Shannon channel limitations, share formal properties that make significant fidelity loss expectable in mass transmission contexts. This structural expectability—not mathematical certainty—transforms the ethics of intellectual work. We ground the resulting normative framework in a responsibility-for-risk model: when actors can foresee that their actions create significant risks of harm, they acquire obligations to mitigate those risks proportionate to foreseeability and severity. The paper articulates seven principles of Conceptual Responsibility, organized into two categories: principles derived from transmission dynamics (Principles 1–5) and principles derived from social epistemology (Principles 6–7). We acknowledge that transmission can also improve ideas through selection, refinement, and creative synthesis, and situate our framework within the broader literature on testimony and communicative responsibility. The framework aims to provide practical guidance for intellectuals navigating the tension between communicative reach and interpretive fidelity.

Keywords: idea transmission, serial reproduction, iterated learning, conceptual responsibility, intellectual ethics, cultural evolution, testimony, risk ethics

1 Introduction: The Problem and the Argument

Philosophical and scientific ideas routinely enter history with meanings their creators never intended. Nietzsche's *Übermensch* becomes a Nazi racial ideal. Darwin's natural selection becomes Social Darwinist brutality. Marx's critique of alienation becomes Stalinist apologetics. These transformations are typically treated as contingent failures—mistakes that better communication or more careful reading might have prevented.

This paper argues that such transformations, while not strictly inevitable in any individual case, reflect structural features of mass communication that make significant distortion highly probable and therefore expectable. We develop this argument through three complementary approaches, drawing on multiple theoretical traditions:

- Empirical evidence from serial reproduction experiments demonstrating systematic degradation patterns

- Formal models from cultural evolution theory explaining why transmission bottlenecks favor simplified forms
- Structural analogies from information theory illuminating the constraints that generate these patterns
- Normative resources from risk ethics and the epistemology of testimony grounding authorial obligations

We then argue that this structural expectability generates genuine ethical obligations for authors, grounded in a responsibility-for-risk framework. The paper concludes with seven principles of Conceptual Responsibility specifying these bounded obligations.

1.1 Clarifying the Central Claim

Our thesis is not that distortion is "mathematically inevitable" in the sense of logical necessity. Rather, we claim that:

- Systematic distortion patterns have been empirically demonstrated across multiple experimental paradigms
- These patterns can be explained by structural features of human cognition and social communication
- Structural analogies from information theory illuminate why these patterns emerge
- The resulting expectability of distortion grounds genuine (though bounded) ethical obligations

This is weaker than claiming isomorphism with Shannon channels but stronger than merely noting that distortion sometimes occurs. The question is whether the structural constraints are sufficiently robust and well-understood to generate normative implications.

1.2 Relation to the Epistemology of Testimony

This paper addresses territory adjacent to the philosophical literature on testimony (Coady, 1992; Lackey, 2008; Fricker, 2007). That literature examines when hearers are justified in accepting testimony and what epistemic responsibilities speakers bear. Our concerns are related but distinct:

The testimony literature primarily concerns individual communicative acts and their epistemic properties. We concern ourselves with the aggregate dynamics of ideas transmitted through chains of speakers—what might be called "iterated testimony." The relevant question is not "Is this particular testimony reliable?" but "What happens to ideas as they pass through many testifiers?"

Additionally, the testimony literature focuses primarily on epistemic norms (truth, justification, knowledge). We emphasize ethical norms (harm, responsibility, mitigation). While these domains overlap—false testimony can cause harm—they are not identical. An idea might be transmitted with full epistemic sincerity at each step yet still degrade into something harmful through cumulative transformation.

We thus position our framework as complementary to the epistemology of testimony, extending its concerns to iterated transmission and foregrounding ethical alongside epistemic considerations.

2 Empirical Foundations: The Evidence for Systematic Distortion

2.1 Bartlett's Serial Reproduction Paradigm

Frederic Bartlett's serial reproduction experiments (Bartlett, 1932) established the foundational paradigm for studying transmission distortion. In the classic design, participants receive a stimulus, reproduce it from memory, and their reproduction becomes the stimulus for the next participant—analogous to the children's game of "telephone."

Bartlett's famous "War of the Ghosts" study demonstrated systematic changes across transmission chains:

- **Shortening**: Material became progressively shorter as details were lost
- **Rationalization**: Unfamiliar elements were modified to fit participants' cultural schemas
- **Conventionalization**: Unusual features were replaced with more familiar forms
- **Assimilation**: Content was distorted toward receivers' prior expectations

2.2 Modern Replications and Extensions

Contemporary research has replicated and extended Bartlett's findings with more rigorous methodology. Roediger et al. (2014) directly compared serial reproduction (transmission between individuals) with repeated reproduction (the same individual recalling multiple times). Key findings:

- Serial reproduction produced significantly greater forgetting of original material than repeated reproduction
- The proportion of accurate original content declined markedly across transmission links
- Schema-consistent intrusions (ideas not in the original but fitting expectations) remained stable or increased

Crucially, these effects occur even when participants are explicitly instructed to be accurate (Gauld & Stephenson, 1967; Bergman & Roediger, 1999). The distortion is not primarily a matter of carelessness but reflects structural features of memory and communication.

2.3 Addressing Methodological Objections

Critics have noted that serial reproduction experiments use artificial conditions: single exposure, no opportunity for clarification, no access to written records. These conditions may exaggerate degradation relative to real-world transmission.

We acknowledge this limitation but note three responses:

First, even under artificial conditions, the experiments reveal underlying cognitive tendencies that operate in all transmission contexts, even if counteracted by error-correction mechanisms in some settings. The artificial conditions isolate the signal; real-world transmission combines this signal with noise and correction in varying proportions.

Second, many real-world transmission contexts do approximate the experimental conditions: brief media mentions, casual conversation, social media sharing, word-of-mouth reputation. These are precisely the contexts through which ideas reach mass audiences.

Third, subsequent research has examined transmission in more naturalistic settings (rumor transmission, urban legend propagation, scientific citation networks) and found similar patterns of systematic distortion, though often attenuated by corrective factors.

3 Iterated Learning: Why Bottlenecks Favor Simplified Forms

The iterated learning paradigm provides formal models explaining the serial reproduction findings. In these models, each generation learns from a finite sample of the previous generation's behavior, creating a "transmission bottleneck."

3.1 The Core Mechanism

Mathematical analysis (Griffiths & Kalish, 2007; Kirby et al., 2014) reveals that transmission bottlenecks systematically bias cultural evolution. The key insight: when learners must reconstruct a system from incomplete data, their prior expectations fill the gaps. Over generations, the transmitted system converges toward these priors.

Xu & Griffiths (2010) formalized this as a Bayesian process. Let θ represent the true state of some cultural convention (e.g., the meaning of an idea), and let x represent the data observed by a learner. Each learner forms a posterior distribution $P(\theta|x)$ combining observed data with prior expectations $P(\theta)$, then generates behavior that becomes data for the next learner. The mathematical result: serial reproduction converges to the prior distribution, regardless of the starting point:

$$\lim_{n\to\infty} P(\theta|x_1 \dots x_n) \to P(\theta) \tag{1}$$

where θ is the transmitted content, x represents observations, and $P(\theta)$ is the prior distribution of receiver expectations. This convergence occurs even when individual transmitters intend perfect fidelity. The bottleneck itself drives simplification toward whatever forms the population expects.

3.2 When Constraints Are Generative

Importantly, iterated learning research also shows that transmission constraints can be generative, not merely destructive. Kirby et al. (2008) demonstrated that artificial languages transmitted through iterated learning become more structured and learnable over generations. Transmission bottlenecks can create systematicity where none existed.

This finding complicates any uniformly pessimistic framing. We therefore distinguish two senses of "distortion":

- **Loss of original content**: Information present in the source but absent in transmitted versions (degradation)
- **Transformation toward receiver expectations**: Content is modified to fit schemas and increase transmissibility

The iterated learning literature shows that the second process can sometimes improve ideas—making them clearer, more systematic, better adapted to communicative needs. However, "improvement" by transmission criteria may constitute "distortion" by fidelity criteria. A more learnable version may be less accurate to the original.

Our normative framework addresses contexts where fidelity to original meaning matters—where distortion constitutes harm rather than refinement. Not all transmission contexts are like this, but many important ones are.

4 Structural Analogies from Information Theory

We now introduce concepts from information theory, with explicit clarification of their status. These are structural analogies that illuminate why the empirical patterns occur, not claims of mathematical identity between Shannon channels and human cognition.

4.1 The Status of the Analogy

Shannon's information theory applies to well-defined channels with measurable parameters: bandwidth in hertz, noise power in watts, information in bits. Human cognition is not such a channel. We cannot measure the "bit rate" of idea transmission or the "channel capacity" of working memory in Shannon's precise sense.

Nevertheless, the analogy is not merely metaphorical. The formal properties that make Shannon channels subject to capacity limits—finite processing resources, interference from other signals, reconstruction from incomplete data—have structural analogues in human cognition:

- **Shannon**: Finite bandwidth limits transmission rate → **Cognition**: Limited attention and working memory limit information uptake
- **Shannon**: Noise corrupts signals → **Cognition**: Interference from prior knowledge and competing inputs corrupts encoding
- **Shannon**: Lossy compression is irreversible → **Cognition**: Information lost in encoding cannot be recovered from memory
- **Shannon**: Generation loss compounds with repeated compression → **Cognition**: Serial reproduction compounds degradation

These structural parallels explain why similar dynamics emerge even though the substrates differ. The analogy is productive because it generates predictions (tested in the empirical literature) and suggests interventions (encoded in the normative principles).

4.2 Channel Capacity as a Structural Concept

We use "channel capacity" to denote the upper bound on complexity that can be accurately transmitted through a given medium. For human cognition, this bound is established empirically rather than derived mathematically:

- **Working memory**: Cowan (2001) estimate suggests a capacity of approximately 4 chunks, significantly lower than Miller's (1956) original "7±2" which conflated capacity with chunking strategies
- **Attention**: Selective and easily depleted; only a fraction of available information is processed
- **Long-term memory**: Reconstructive rather than reproductive; stored information is transformed by retrieval processes

These empirically established limitations function analogously to channel capacity: they impose upper bounds on transmission fidelity that cannot be overcome by effort or intention alone.

4.3 The Compression Analogy

Information theory distinguishes lossless compression (original perfectly recoverable) from lossy compression (some information permanently discarded). Human cognitive processing of complex ideas is structurally lossy:

- Ideas exceeding working memory capacity must be selectively encoded
- Selective encoding discards some components while retaining others

- Discarded components cannot be recovered from the encoded representation
- Each transmission step applies additional lossy compression to an already-compressed signal

This explains the generation loss observed in serial reproduction. It is not that each transmitter makes independent random errors; rather, each applies systematic compression that compounds previous compressions.

4.4 What the Analogy Does Not Claim

To be explicit about the limits of our claim:

- We do **not** claim that Shannon's theorems apply directly to idea transmission
- We do **not** claim that "channel capacity" for cognition can be measured in bits
- We do **not** claim that distortion is logically inevitable in all cases
- We **do** claim that structural features of cognition create predictable, systematic distortion patterns
- We **do** claim that these patterns are sufficiently robust to ground normative conclusions about foreseeable risk

5 A Heuristic Model of Idea Transformation

We now present a heuristic model for analyzing idea transformation. This is an organizing framework, not a formal theory. The notation provides precision for discussion, not mathematical derivation.

5.1 Ideas as Structured Information

Consider an idea as articulated by its author as comprising several components:

- P = Core propositions (the central claims)
- R = Relational structure (logical/evidential connections)
- C = Contextual qualifiers (scope limitations, caveats)
- A = Argumentative scaffolding (supporting reasoning)

This decomposition is analytic, not ontological—real ideas do not come labeled with these components. But the distinction is useful because empirical research shows differential degradation rates: C and R degrade fastest, then A, then P.

5.2 Differential Degradation Rates

The differential degradation pattern has clear cognitive explanations:

- **Contextual qualifiers** (C): Often encoded as "hedges" peripheral to main content; low salience predicts early loss
- **Relational structure** (R): Requires maintaining multiple elements simultaneously in working memory; high cognitive load predicts degradation

- **Argumentative scaffolding** (A): Often lengthy and complex; exceeds typical transmission bandwidth
- **Core propositions** (P): Typically most salient and memorable; highest probability of survival

5.3 The Attractor State: A Qualified Claim

We hypothesize that ideas entering mass transmission tend toward an attractor state comprising core propositions stripped of context. However, this claim requires qualification:

The attractor dynamic applies primarily to ideas entering mass public discourse through low-bandwidth channels (media mentions, casual conversation, social media). In specialized expert communities with high-bandwidth channels (academic discourse, professional training), complex ideas can maintain complexity indefinitely. Quantum mechanics remains complex among physicists even while being caricatured in popular culture.

The framework thus predicts that the same idea will have different transmission fates in different contexts. Mass transmission drives toward slogans; specialized transmission can maintain complexity. The ethical implications primarily concern ideas that will (foreseeably) enter mass transmission.

6 When Transmission Improves Ideas

Any adequate account must address cases where transmission improves rather than degrades ideas. We identify three mechanisms:

6.1 Selection

Cultural transmission is selective: not all ideas spread equally. This selection can favor better ideas—clearer formulations, more useful concepts, more accurate claims.

However, selection operates on transmissibility, not truth. Ideas that spread are those that are easy to understand, emotionally resonant, and compatible with existing beliefs—properties that correlate imperfectly with accuracy or value.

6.2 Refinement

Commentary traditions, scholarly debate, and pedagogical reformulation can clarify and improve ideas over time. Aristotle's thought was refined by centuries of commentary; Darwin's theory was improved by the Modern Synthesis.

These refinement processes typically require institutional support that resists the degradation dynamics we describe.

6.3 Creative Synthesis

Transmission can generate novel combinations that are better than any source. The recombination of fragments from different traditions can produce creative syntheses neither tradition could have achieved alone.

We acknowledge this possibility while noting that it is orthogonal to our concerns. Creative synthesis produces new ideas; our framework concerns fidelity to existing ideas.

7 Normative Grounding: From Expectability to Obligation

We now address the is-ought transition: why does the structural expectability of distortion generate authorial obligations rather than excuses?

7.1 The Responsibility-for-Risk Framework

We ground our normative claims in a responsibility-for-risk model drawn from applied ethics (Jonas, 1984; Shrader-Frechette, 1991). The core principle:

> *When an agent can foresee that their action creates a significant risk of harm, and can take reasonable steps to mitigate that risk, they acquire an obligation to do so, proportionate to the foreseeability and severity of potential harm.*

This principle operates across many domains: product liability, environmental regulation, professional ethics, informed consent. It does not require certainty of harm—risk is sufficient.

7.2 Application to Idea Transmission

Applied to intellectual work:

- Authors release ideas into transmission environments with known degradation properties
- Degraded ideas can cause harm (ideological weaponization, policy mistakes, damaged reputations)
- Authors can take reasonable steps to mitigate degradation risk (the seven principles)
- Therefore, authors acquire obligations proportionate to foreseeability and potential severity

7.3 The Foreseeability Standard

A key objection asks: at what probability threshold does foreseeability generate obligation? We propose a practical standard:

> *A distortion is foreseeable for purposes of authorial responsibility when a thoughtful author, aware of the transmission environment and the features of their text, would recognize it as a significant risk—not merely conceivable, but probable enough to warrant consideration in framing decisions.*

Factors relevant to this judgment include:

- **Historical precedent**: Has similar material been distorted in similar ways before?
- **Textual features**: Does the text contain extractable fragments, provocative formulations, or ambiguous claims?
- **Transmission context**: Will the idea reach mass audiences through low-bandwidth channels?
- **Ideological environment**: Are there motivated actors likely to appropriate the idea for harmful purposes?

7.4 Why Expectability Generates Obligation Rather Than Excuse

One might argue: "If distortion is structural, authors cannot be blamed for it." We reject this inference:

First, structural constraints are not absolute barriers but factors that can be partially addressed through design choices. An author cannot prevent all compression, but they can influence what survives compression.

Second, the reasoning proves too much. Many harms involve structural factors beyond individual control; we do not therefore excuse actors who knowingly contribute to these harms when they could take mitigating action.

Third, the relevant comparison is not between actual outcomes and perfect transmission, but between outcomes with and without responsible authorial choices.

8 The Seven Principles of Conceptual Responsibility

We organize the principles into two categories reflecting their theoretical sources:

- **Principles 1–5**: Derived from transmission dynamics (cognitive constraints, compression, fragmentation)
- **Principles 6–7**: Derived from social epistemology (authority effects, distributed responsibility)

This organization acknowledges that the framework draws on multiple theoretical traditions rather than claiming derivation from a single unified apparatus.

8.1 Part A: Principles from Transmission Dynamics

Principle 1: *Explicit Boundary Articulation*

Authors should make scope limitations and conditions of applicability maximally explicit and salient.

Empirical basis: Contextual qualifiers (C) degrade fastest in transmission. Low-salience hedges are lost early. Boundary information survives longer when encoded prominently.

Normative grounding: If loss of boundaries foreseeably enables harmful misapplication, and prominence increases survival probability, authors have reason to make boundaries prominent.

Practical guidance:

- State scope in the opening paragraph, not just in caveats
- Repeat key limitations at each major argumentative turn
- Make boundary statements as memorable as core claims

Principle 2: *Anticipatory Misreading Analysis*

Authors should identify foreseeable distortions and address them proactively in the text.

Empirical basis: Transmission converges toward receiver priors (Xu & Griffiths, 2010). Authors can model likely priors and anticipate the distortions they will induce.

Normative grounding: The foreseeability standard: if a thoughtful author would recognize a distortion as a significant risk, they have reason to address it.

Practical guidance:

- Explicitly state what the argument does not claim
- Anticipate the most likely simplified versions and explain why they are inadequate
- Address known ideological appropriation risks directly

Principle 3: *Ethical Application Guidance*

Authors should explicitly articulate normative constraints on how their ideas should be applied.

Empirical basis: Compressed ideas become instrumental tools. Normative constraints are lost with other contextual information unless given high salience.

Normative grounding: If harmful applications are foreseeable and normative guidance could reduce them, authors have reason to provide such guidance.

Practical guidance:

- Distinguish legitimate from illegitimate applications
- Articulate the values that should govern implementation
- Make ethical constraints as prominent as practical implications

Principle 4: *Active Discursive Engagement*

Living authors should monitor significant interpretations and actively correct prominent distortions.

Empirical basis: Transmission degradation can be partially counteracted by error-correction. Author intervention introduces corrective signal into the transmission system.

Normative grounding: If ongoing engagement can reduce foreseeable harm at reasonable cost, authors have reason to engage.

Scope limitation: This obligation applies only to living authors, scales with prominence and potential harm, and is bounded by reasonable capacity.

Principle 5: *Structural Fragmentation Resistance*

Authors should design texts to resist extraction of dangerous fragments.

Empirical basis: Fragmentable texts generate fragments that can be recombined into harmful constructs. Systematic argumentation resists fragmentation better than aphoristic style.

Normative grounding: If fragmentability foreseeably increases harm risk, authors have reason to reduce fragmentability, balanced against communicative needs.

The trade-off: Fragmentation resistance exists in tension with memorability. The goal is graceful degradation: ensuring that the fragments that do emerge are less harmful than alternatives.

8.2 Part B: Principles from Social Epistemology

Principle 6: *Epistemic Humility*

Authors should present ideas with appropriate uncertainty markers and openness to revision.

Empirical basis: Authority status reduces critical engagement (Cialdini, 2001; Milgram, 1974), accelerating dogmatization. This is a social-epistemic phenomenon, not a transmission-bandwidth constraint.

Normative grounding: If authority effects foreseeably increase harm from distorted dogma, authors have reason to resist their own authority effects. This draws on norms of epistemic responsibility in testimony (Fricker, 2007).

Practical guidance:

- Acknowledge limitations of evidence and possibility of error
- Present strongest counterarguments
- Explicitly invite criticism

Principle 7: *Downstream Responsibility Gradient*

Authorial responsibility diminishes as ideas pass through more intermediaries.

Theoretical basis: This principle reflects the distributed nature of testimonial chains (Lackey, 2008). Each transmitter makes choices for which they bear proximate responsibility; the original author's responsibility is mediated by these intervening choices.

Normative grounding: Responsibility-for-risk is proportionate to proximity and control. Authors have greatest control over initial framing; this control diminishes as intermediaries intervene.

Formal representation: We suggest that author responsibility decays with transmission distance. A plausible functional form is exponential:

$$R(n) \approx R_0 \times \lambda^n, \quad \text{where } 0 < \lambda < 1 \tag{2}$$

Here $R(n)$ represents author responsibility after n transmission steps, R_0 is initial responsibility, and λ is a decay factor. We do not claim this specific functional form is uniquely justified—it represents a qualitative claim that responsibility decays, with exponential form as one plausible specification. Alternative forms (linear, hyperbolic) would express different intuitions about decay rate. Empirical work on responsibility attribution across causal chains could inform this choice, though such work remains methodologically challenging.

9 Objections and Replies

9.1 The Metaphor Objection

Objection: The information-theoretic analogy is merely metaphorical and cannot support strong conclusions.

Reply: We agree that the analogy is not a mathematical identity. However, productive analogies in science need not be identities. The question is whether the analogy generates correct predictions and useful interventions. The structural framing predicts differential degradation rates, convergence toward simplified forms, and generation loss in serial transmission—all confirmed by independent empirical research. This is sufficient for a productive analogy, even without isomorphism.

9.2 The Is-Ought Objection

Objection: The paper moves illegitimately from descriptive claims to normative conclusions.

Reply: The normative conclusions follow not from the descriptive claims alone but from those claims combined with the responsibility-for-risk framework, which is an independently defensible normative principle. The structure is: (1) Transmission creates foreseeable risk of harmful distortion (empirical); (2) Agents who create foreseeable risks have obligations to mitigate those risks (normative principle); (3) Therefore, authors have obligations to mitigate distortion risk (conclusion).

9.3 The Pessimism Objection

Objection: The framework is uniformly pessimistic, ignoring cases where transmission improves ideas.

Reply: Section 6 explicitly addresses positive transmission dynamics. We acknowledge selection, refinement, and creative synthesis as mechanisms of improvement. The framework applies to contexts where fidelity matters and degradation constitutes harm—not to all transmission contexts universally.

9.4 The Paralysis Objection

Objection: If all communication is subject to distortion risk, the framework generates paralysis.

Reply: The framework generates bounded obligations, not prohibitions. Authors need not guarantee perfect transmission; they need take reasonable mitigating steps proportionate to foreseeability and potential harm.

9.5 The Digital Media Objection

Objection: Digital media increase bandwidth, reducing rather than exacerbating degradation.

Reply: Digital media have complex effects. They increase available bandwidth but decrease attention per item; they enable access to primary sources but also accelerate spread of simplified versions. The net effect on transmission fidelity is empirically contested and likely context-dependent. We withdraw any strong claim about digital media necessarily worsening distortion; the framework applies to whatever transmission dynamics actually obtain in a given context, which must be assessed empirically.

10 Methodological Notes on Testing the Framework

The framework generates empirically testable predictions. We sketch potential operationalizations while flagging open methodological questions.

10.1 Measuring Transmission Fidelity

Transmission fidelity can be operationalized through:

- **Propositional accuracy**: Proportion of original claims preserved across transmission steps (requires coding scheme for identifying propositions)
- **Semantic similarity**: Cosine similarity of text embeddings between original and transmitted versions (open questions: which embedding model? trained on what corpus? validated against human judgments?)
- **Expert coding**: Trained raters assess degree of distortion on validated scales (requires development and validation of coding instruments)

Each operationalization involves methodological choices that affect results. We flag these as open questions for future research rather than solved problems.

10.2 Testing Differential Degradation

The prediction that contextual qualifiers degrade faster than core claims can be tested by:

- Coding original texts for component types (P, R, C, A) using a validated coding scheme
- Tracking survival rates of each component type across transmission
- Comparing decay curves

Challenges include reliable coding of component types and sufficient sample sizes for statistical power.

10.3 Testing the Principles

Each principle generates specific predictions:

- **Principle 1**: Texts with high-salience boundaries will show slower boundary degradation
- **Principle 2**: Texts that address foreseeable misreadings will show lower rates of those misreadings
- **Principle 5**: Systematic argumentative texts will show lower variance in transmitted meanings than aphoristic texts

These predictions are testable through controlled transmission experiments comparing matched texts varying on the relevant dimensions.

11 Conclusion

Ideas transmitted through social networks systematically degrade. This claim is supported by converging evidence from experimental psychology, cultural evolution theory, and structural analysis drawing on information-theoretic concepts. While the theoretical resources we employ are diverse—including empirical findings, formal models, structural analogies, and normative frameworks from risk ethics and the epistemology of testimony—they converge on a common picture: transmission constraints make significant distortion expectable in mass communication contexts.

The structural expectability of distortion transforms intellectual ethics. Authors cannot prevent their ideas from being compressed, simplified, and fragmented. But they can influence what survives these processes. The seven principles of Conceptual Responsibility—organized into principles derived from transmission dynamics (Boundary Articulation, Misreading Analysis, Ethical Guidance, Discursive Engagement, Fragmentation Resistance) and principles derived from social epistemology (Epistemic Humility, Responsibility Gradient)—provide guidance for navigating the tension between communicative reach and interpretive fidelity.

We have tried to be explicit about what our framework claims and what it does not. It does not claim mathematical inevitability of distortion in all cases. It does not claim that transmission is purely destructive. It does not claim that authors bear unlimited responsibility for all downstream uses. What it does claim is that foreseeable risks generate bounded obligations, that structural features of transmission create such risks, and that responsible intellectual work requires attending to them.

In a world where ideas shape institutions and lives, this framework offers a path toward more responsible intellectual practice—not by guaranteeing perfect communication, which is impossible, but by designing for graceful degradation in an imperfect communicative world.

References

Bartlett, F. C. (1932). *Remembering: A Study in Experimental and Social Psychology*. Cambridge University Press.

Bergman, E. T., & Roediger, H. L. (1999). Can Bartlett's repeated reproduction experiments be replicated? *Memory & Cognition*, 27(6), 937–947.

Cialdini, R. B. (2001). *Influence: Science and Practice* (4th ed.). Allyn & Bacon.

Coady, C. A. J. (1992). *Testimony: A Philosophical Study*. Oxford University Press.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, 24(1), 87–114.

Cover, T. M., & Thomas, J. A. (2006). *Elements of Information Theory* (2nd ed.). Wiley.

Fricker, M. (2007). *Epistemic Injustice: Power and the Ethics of Knowing*. Oxford University Press.

Gauld, A., & Stephenson, G. M. (1967). Some experiments relating to Bartlett's theory of remembering. *British Journal of Psychology*, 58(1-2), 39–49.

Griffiths, T. L., & Kalish, M. L. (2007). Language evolution by iterated learning with Bayesian agents. *Cognitive Science*, 31(3), 441–480.

Jonas, H. (1984). *The Imperative of Responsibility: In Search of an Ethics for the Technological Age*. University of Chicago Press.

Kirby, S., Cornish, H., & Smith, K. (2008). Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *PNAS*, 105(31), 10681–10686.

Kirby, S., Griffiths, T., & Smith, K. (2014). Iterated learning and the evolution of language. *Current Opinion in Neurobiology*, 28, 108–114.

Kriger, B. (2016). The ethical responsibility of philosophers: Ideas as agents of change. *The Common Sense World*.

Lackey, J. (2008). *Learning from Words: Testimony as a Source of Knowledge*. Oxford University Press.

Milgram, S. (1974). *Obedience to Authority: An Experimental View*. Harper & Row.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 63(2), 81–97.

Roediger, H. L., Meade, M. L., Gallo, D. A., & Olson, K. R. (2014). Bartlett revisited: Direct comparison of repeated reproduction and serial reproduction techniques. *Journal of Applied Research in Memory and Cognition*, 3(4), 266–271.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27(3), 379–423.

Shrader-Frechette, K. S. (1991). *Risk and Rationality: Philosophical Foundations for Populist Reforms*. University of California Press.

Xu, J., & Griffiths, T. L. (2010). A rational analysis of the effects of memory biases on serial reproduction. *Cognitive Psychology*, 60(2), 107–126.