# Identity Collapse and the Platforming of Fraud: Structural Trust Degradation in Facebook's Design

Flyxion

## 1 Introduction

The persistence of fraud, the erosion of institutional trust, and the normalization of deception on Facebook are often framed as problems of moderation capacity, malicious users, or unfortunate side effects of unprecedented scale. This framing is inadequate. The central claim of this essay is that these outcomes are neither accidental nor emergent in any strong sense, but rather the predictable and foreseeable consequences of architectural decisions that weaken identity constraints, incentivize imitation, and optimize interaction without regard to legitimacy.

A growing body of scholarship has documented how engagement-optimized platforms extract attention as a resource while externalizing social cost (Wu 2016; Zuboff 2019). Within such systems, behavioral metrics displace normative constraints, and optimization targets replace truth, trust, or accountability as governing principles. Facebook exemplifies this logic at planetary scale. By permitting non-unique identities, encouraging uncredited remixing of personal and commercial media, constraining user refusal, and coupling monetization promises to structurally unattainable thresholds, the platform converts trust into an extractable commodity rather than a conserved social resource.

These dynamics were not discovered retroactively. They were visible to ordinary users almost immediately and have been repeatedly documented by journalists, researchers, and internal whistleblowers (Haugen 2021; Horwitz and Seetharaman 2021). That they persist reflects not uncertainty but prioritization. The system does not merely fail to prevent scams. It creates an environment in which scamming becomes adaptive, automated, and economically rational.

## 2 Identity as a Degraded Namespace

At the foundation of any trust-bearing information system lies a basic requirement: identities must be uniquely bound to histories. This requirement is not ethical but structural. Without persistent, injective mappings between identifiers and entities, attribution becomes probabilistic and accountability collapses. Facebook systematically violates this requirement by allowing individuals, businesses, and organizations to share identical names, profile images, and visual branding without meaningful namespace separation.

Economic theory predicts the consequences of such ambiguity. In environments where quality signals cannot be reliably distinguished, low-quality and deceptive actors drive out legitimate ones,

producing what Akerlof famously described as a market for lemons (Akerlof 1970). On Facebook, this logic manifests not only in commercial exchange but in social identity itself. Pages named after real businesses or public figures routinely appear with no followers, no engagement history, and no provenance, yet are visually indistinguishable from the entities they impersonate.

Over time, this saturates the platform with low-credibility replicas, degrading trust not only in Facebook as an intermediary but in online identity as a concept. Trust ceases to be an accumulated property grounded in continuity and instead becomes a fleeting inference drawn from surface resemblance. As Lessig argued in an earlier era of digital governance, architecture is law (Lessig 1999). When the architecture refuses to enforce identity integrity, impersonation becomes a native operation rather than an edge case.

## 3 Remix Culture as Identity Laundering

Facebook's promotion of unrestricted remixing, resharing, and reposting of images, videos, and music further accelerates identity degradation. While presented as participatory creativity, this policy functions in practice as identity laundering. Content circulates divorced from authorship, context, and provenance, allowing credibility to be transferred without accountability. What appears as cultural openness in fact operates as a mechanism for stripping signals of origin.

This dynamic disproportionately benefits malicious actors. Scammers are not required to establish trust through sustained behavior. They merely appropriate it by proximity, repackaging familiar imagery or viral media to inherit emotional resonance and perceived legitimacy. Engagement-based ranking systems reward this behavior, as remixed content often performs better than original material under metrics optimized for interaction rather than integrity (Campbell 1979; Goodhart 1975).

The broader epistemic consequence is a degradation of contextual integrity. As Nissenbaum has argued, information derives meaning from its normative context of flow (Nissenbaum 2010). When content is systematically detached from its origins, audiences are trained to accept attribution loss as normal, while creators experience the erosion of authorship as a structural condition rather than an aberration. The result is an environment in which legitimacy is endlessly transferable and therefore increasingly meaningless.

## 4 Forced Exposure and the Impossibility of Refusal

Ethical systems require the possibility of refusal. On Facebook, this possibility is deliberately constrained. Users cannot reliably disable entire content classes such as reels, cannot permanently reduce advertising density, and cannot effectively block categories of interaction that repeatedly violate their preferences. Blocking operates locally and transiently, while exposure operates globally and persistently.

This asymmetry ensures that engagement is not chosen but extracted. Attention becomes compulsory rather than voluntary, and users are subjected to algorithmically injected content regardless of expressed intent. As Wu has shown, attention markets historically rely on scarcity and coercion

rather than consent (Wu 2016). Facebook extends this logic into the digital domain by making refusal costly, incomplete, or illusory.

From the perspective of governance, this is not a usability flaw but a revenue requirement. Guaranteed impression volume underwrites advertising markets and monetization schemes. The inability to opt out is therefore not accidental. It is a prerequisite for the platform's economic model, which treats user attention as a resource to be allocated rather than an agency to be respected (Zuboff 2019).

## 5 Monetization as Coercive Hope

Facebook's monetization architecture exemplifies a particularly insidious form of extraction that operates through aspirational signaling rather than direct coercion. Pages and creators are encouraged to pursue monetization through dashboards, notifications, and prompts that imply attainability and progress. Yet the actual requirements for eligibility remain prohibitively high, opaque, and frequently retroactive. Creators often discover only after sustained labor that they must maintain tens of thousands of followers, produce specific content formats such as reels, and meet engagement thresholds that reset on short time horizons.

This structure functions as coerced hope. Labor is extracted in advance, visibility is rationed algorithmically, and failure is individualized rather than attributed to systemic throttling. When reach collapses, creators are offered paid promotion as a remedy, effectively charging users to recover visibility that was artificially withheld. This dynamic exemplifies Goodhart's Law: when engagement metrics become targets, they cease to function as meaningful measures of value (Goodhart 1975). What is optimized is not quality or trust, but metric performance divorced from substance.

The broader political economy of this arrangement resembles what Varoufakis has described as technofeudal dependency, in which access to audiences is controlled by platform lords who extract rent from aspirational labor (Varoufakis 2023). Cory Doctorow has similarly argued that platforms enclose the means of computation and then sell partial relief from the constraints they themselves impose (Doctorow 2023). In this context, monetization does not reward excellence. It monetizes desperation.

## 6 Advertising Saturation and Institutional Erosion

Advertising on Facebook is not merely frequent; it is structurally invasive. Sponsored content appears at regular intervals regardless of user behavior, and blocking individual advertisers does not reduce ad density. It merely triggers replacement. The rate remains constant while the quality often declines. Many advertisements are themselves scams, low-quality dropshipping schemes, or impersonations of legitimate businesses.

This saturation produces what may be described as epistemic pollution. When deceptive content becomes ambient background noise, users lose the ability to treat any signal as reliable. Trust collapses not because deception is rare, but because it is ubiquitous. As Akerlof's analysis predicts, when low-quality signals dominate, high-quality actors withdraw or are ignored (Akerlof 1970). The platform

benefits from this collapse insofar as skepticism toward organic content increases dependence on paid amplification and platform-mediated verification.

Network theory suggests that such effects scale nonlinearly. As platforms grow, local trust failures propagate through dense connectivity, amplifying institutional erosion across the network (Barabási 2016). What begins as annoyance becomes cynicism, and what begins as cynicism becomes disengagement from institutions altogether. The harm is therefore not confined to individual users but diffuses outward into the broader social fabric.

## 7 Scam Recycling and Platform Incentives

When scams are reported and removed from Facebook, the response is intentionally incomplete. Accounts are deleted, but IP addresses, behavioral signatures, and reuse patterns are not meaningfully constrained. This allows scammers to reappear almost immediately under new names, often with improved tactics refined through prior failures. Each iteration generates advertising revenue, engagement metrics, or transaction fees before eventual removal.

This dynamic creates a profound learning asymmetry. Attackers learn globally across attempts, while the platform forgets locally at each deletion. As Elad's work on sparse and redundant representations suggests, adversarial actors adapt rapidly in underconstrained environments, exploiting gaps left by brittle classifiers (Elad 2010). Removal without memory functions as negative learning. It erases information that could have constrained future abuse.

Internal documents and whistleblower testimony indicate that these patterns are well understood within the company (Haugen 2021; Horwitz and Seetharaman 2021). Their persistence therefore cannot be attributed to ignorance. The platform's refusal to impose durable consequences is not neutral. It actively subsidizes fraud by lowering the cost of iteration and externalizing harm onto users. Over time, scammers become more automated, more persuasive, and more desperate, escalating both sophistication and volume. Fraud, under these conditions, is not an anomaly. It is an equilibrium.

## 8 Metric Gamification and the Simulation of Legitimacy

In the absence of reliable identity and historical continuity, Facebook substitutes metric performance for legitimacy. Likes, followers, views, and engagement ratios are treated as proxies for trust despite being trivially manipulable. This substitution is predictable. When direct signals of quality are unavailable, users infer credibility from visible counters, regardless of their provenance (Campbell 1979).

The result is a gamified legitimacy economy in which appearances dominate outcomes. Empty pages, engagement farms, and deceptive growth tactics flourish, while honest actors are penalized for refusing manipulation. Legitimacy is no longer conserved through behavior. It is manufactured, traded, and exhausted. As Simon warned in his analysis of information-rich environments, cognitive overload forces reliance on simplified heuristics that systems can easily exploit (Simon 1971).

This simulation of legitimacy completes the scam pipeline. Quantitative activity is converted into qualitative trust, which is then monetized. The platform's metrics do not merely measure behavior. They shape it, selecting for actors who optimize surface signals rather than substantive contribution. Trust degradation, in this sense, is not a side effect of metricization. It is its logical conclusion.

## 9 Formal Systems Interpretation

From the perspective of formal systems theory, Facebook may be understood as an identity-processing automaton that operates without injective mappings. In any well-formed symbolic system, references must map uniquely and persistently to histories in order for inference to be sound. When multiple entities share indistinguishable names, images, and surface features, the system abandons injectivity and instead permits many-to-one and one-to-many mappings between symbol and referent. This violates the basic conditions required for reliable reasoning.

In such a system, truth becomes undecidable at the interface level. Users are not provided with sufficient information to distinguish authentic states from counterfeit ones because the representational grammar does not encode disambiguation. The platform therefore behaves like a non-deterministic machine whose outputs cannot be verified without external context that the system itself suppresses. Fraud does not exploit a loophole in this architecture; it exploits the architecture itself.

The removal of historical continuity further transforms the platform into a memoryless process. Accounts may be deleted, renamed, or repurposed without preserving a persistent event record accessible to other participants. In formal terms, state transitions are permitted without conservation laws. As Shannon's theory of communication makes clear, information that is not preserved cannot be recovered, and noise accumulates as a result (Shannon 1948). By discarding identity history in favor of throughput, the platform erases the very invariants required to constrain adversarial behavior.

This loss of invariants produces emergent pathologies at scale. As Anderson famously observed, qualitative changes arise when quantitative scale crosses critical thresholds (Anderson 1972). What might appear as tolerable ambiguity in small systems becomes catastrophic in global ones. Bad actors are not merely tolerated but rendered structurally indistinguishable from good ones, because the system refuses to encode the information necessary to tell them apart.

## 10 Thermodynamic Framing of Trust and Entropy

Trust within a social system behaves analogously to a low-entropy resource. It accumulates slowly through consistent behavior, verified identity, and historical continuity, and it dissipates rapidly under conditions of ambiguity and deception. Facebook operates as a high-entropy engine that converts trust gradients into engagement flows. Identity ambiguity, metric gamification, and forced exposure function as entropy injectors that accelerate the decay of reliable signals.

In this framing, scams act as localized entropy spikes that the platform neither contains nor dampens. Instead of isolating these regions and preventing recurrence, the system allows entropy to diffuse across the network, raising the baseline noise level. As entropy increases, users become less able to

discriminate signal from noise, and the marginal cost of deception decreases. This favors actors who externalize harm and penalizes those who invest in coherence.

The platform does not merely tolerate entropy production. It profits from it. Each instance of confusion generates additional engagement, reporting interactions, ad impressions, or monetization attempts. The system thus violates any analogue of a second law that would require entropy to be bounded or compensated. Instead, disorder becomes a revenue source. As studies of critical transitions suggest, systems that allow unchecked entropy accumulation eventually undergo irreversible phase shifts (Scheffer et al. 2009). In the present case, the terminal state is semantic saturation, in which all content appears equally unreliable and trust collapses as a meaningful variable.

## 11 Indictment of Engagement-Optimized Governance

Engagement-optimized governance replaces normative constraints with behavioral metrics. Decisions are evaluated not by their effect on truth, safety, or trust, but by their contribution to measurable interaction. Under this regime, harm is acceptable provided it is engaging, and deception is tolerated provided it sustains activity. Governance becomes a function of throughput rather than stewardship.

Facebook exemplifies this logic by structuring remediation mechanisms to be minimally disruptive to engagement flows. Reporting interfaces constrain expression to predefined categories that suppress novel or systemic harms. Blocking tools operate at the individual level and are deliberately isolated, preventing the emergence of collective defense. Scam removal is reactive and transient, calibrated to placate complaints without altering underlying incentives. As Tufekci has argued, such algorithmic governance systems exhibit forms of agency that escape traditional accountability frameworks (Tufekci 2015).

This mode of governance is not neutral. It constitutes a choice to privilege short-term behavioral extraction over long-term system viability. By refusing to encode trust-preserving constraints into the platform's architecture, Facebook effectively governs by entropy maximization. The resulting environment selects for manipulation, automation, and escalation, while rendering ethical participation increasingly costly and irrational.

The ethical failure here is not that scams occur, but that the system learns from them. Each successful deception informs future optimization, while each victim subsidizes the refinement of extraction strategies. Fraud, in this environment, is not an anomaly but a rational strategy. Trust degradation is not a side effect but a feature.

## 12 Toward Constraint-Restoring Architectures

If the failures described thus far were merely emergent or accidental, remediation might plausibly require novel technologies or unprecedented forms of governance. This is not the case. The core problem is not a lack of sophistication but a refusal to enforce constraints that are already well understood in other domains of computing and institutional design. Any meaningful reform must therefore begin by restoring structural constraints rather than by intensifying content moderation.

One such constraint concerns expressive reporting. Current reporting mechanisms force users to select from narrow taxonomies that suppress context and novelty, rendering systemic harms illegible by design. A system that cannot represent a problem cannot learn from it. Allowing full narrative reports, preserved as first-class objects linked to outcomes, would not only improve detection but also create an auditable record of institutional response. The absence of such mechanisms reflects prioritization choices rather than technical impossibility.

Equally important is the restoration of collective defense. Blocking, reporting, and detection currently operate at the level of isolated individuals, forcing each user to rediscover the same threats independently. This asymmetry benefits attackers, who learn globally, while victims learn locally. Network science suggests that defensive information propagates more effectively when shared across connected nodes (Barabási 2016). Enabling shared block lists and scam signatures would reverse this asymmetry, allowing trust-preserving signals to scale.

Most fundamentally, platforms must adopt event-historical architectures in which identities are inseparable from their action histories. Accounts should function as append-only ledgers rather than disposable shells. Name changes, removals, and enforcement actions must remain legible as part of a public record. Without historical persistence, enforcement cannot generalize, and without generalization, deterrence fails. These principles are commonplace in safety-critical systems, where error logs are preserved precisely to prevent recurrence (Simon 1971).

## 13 Identity as Namespace

In any coherent information system, identity must function as a namespace rather than as a mutable label. A namespace enforces uniqueness, persistence, and referential stability. It ensures that symbols map to histories rather than to appearances, and that actions can be attributed without ambiguity. When identity is treated merely as display content, stripped of structural constraints, the system forfeits its ability to support trust, accountability, or meaning.

Facebook's architecture explicitly rejects identity as namespace. Names are non-unique, visual representations are easily duplicated, and accounts may be renamed or repurposed without preserving a publicly legible history. Identity is treated as content rather than as structure and is therefore optimized for engagement in the same way as posts, images, or videos. Once identity itself becomes subject to metric pressure, it is inevitably gamed. Scammers and impersonators are not deviating from the system's logic. They are executing it efficiently.

The collapse of identity as namespace produces cascading failures. Attribution becomes unreliable, reporting loses specificity, and enforcement actions fail to generalize because they are applied to disposable shells rather than to persistent entities. When an account is removed without preserving its event history, the system deletes information that could have constrained future abuse. This is analogous to deleting error logs in a safety-critical system while leaving the fault unaddressed.

Attempts to layer verification badges, reputation scores, or machine-learning classifiers atop a broken namespace cannot restore coherence. Such measures add complexity without addressing the underlying category error. As Floridi has argued, informational environments require structural in-

tegrity at the level of representation, not merely post hoc correction (Floridi 2014). Without identity as namespace, higher-order governance remains impossible.

## 14 Axiomatic Foundations

The arguments advanced in this work rest on a set of structural axioms concerning identity, trust, and system behavior. These axioms are not normative prescriptions but descriptive constraints required for any system that purports to support coherent social interaction at scale.

The axiom of identity persistence holds that an identity must remain bound to its action history. Any operation that allows an identity to discard or obscure its past without preserving public traceability constitutes an information loss that degrades accountability. The axiom of namespace uniqueness requires that identity symbols map injectively to entities within the system's domain. When this condition is violated, attribution becomes probabilistic and impersonation becomes structurally enabled.

The axiom of historical legibility requires that identity histories remain visible to other participants. Enforcement actions that erase rather than record history destroy the evidence required for trust calibration. The axiom of constraint precedence states that structural constraints must be enforced prior to optimization. Systems that optimize engagement before establishing identity integrity and accountability will inevitably amplify exploitative behaviors. Finally, the axiom of entropy boundedness requires that trust entropy be actively constrained. When deception and ambiguity recur without durable suppression, disorder accumulates faster than corrective mechanisms can dissipate it.

Together, these axioms define the minimum conditions under which trust can exist as a conserved quantity rather than a fleeting illusion. Their violation does not produce isolated harms but induces systemic pathologies that scale with participation. The observed failures of large social platforms are therefore not anomalies or moderation lapses but predictable consequences of axioms that were never enforced.

## 15 Scam Amplification as a Structural Theorem

The axioms articulated above admit a set of unavoidable consequences that may be stated in theorem-like form. These consequences are not empirical generalizations contingent on particular actors or moments in time, but logical entailments of the system's structural properties. Scam amplification, in this sense, is not an unfortunate byproduct of scale but a predictable outcome of engagement-optimized architectures that violate identity and constraint axioms.

When identity persistence and namespace uniqueness are absent, impersonation-based fraud emerges as a dominant strategy. The cost of impersonation approaches zero while the potential payoff remains bounded below by basic human trust heuristics. Under such conditions, rational adversaries converge on deceptive tactics because no countervailing structural penalty exists within the system. As Akerlof's analysis would predict, low-cost deception crowds out high-cost honesty (Akerlof 1970).

Local enforcement without historical continuity further guarantees scam persistence. Each enforcement action removes only a surface instantiation while erasing information that could have constrained future abuse. Attackers learn globally across attempts, while the platform forgets locally at each deletion. This asymmetry ensures that successive scam iterations improve over time. Elad's work on adversarial adaptation in underconstrained representational spaces provides a useful analogy: when detection is brittle and memoryless, attackers exploit sparsity and iterate faster than defenses can adapt (Elad 2010).

Engagement-coupled distribution then amplifies these behaviors superlinearly with platform size. Content ranking systems optimized for interaction preferentially elevate emotionally salient, urgent, or manipulative signals, precisely the features exploited by scams. As networks grow denser, such signals propagate more rapidly, while moderation capacity scales at best linearly (Barabási 2016). The stable equilibrium of this dynamic is not containment but saturation.

## 16 Regulatory Interpretation and the Standard of Obvious Harm

From a regulatory perspective, the failures described in this work do not fall within the category of unforeseen side effects or emergent complexity. They satisfy the standard of obvious harm. The mechanisms by which identity ambiguity enables impersonation, by which engagement optimization amplifies deception, and by which disposable accounts facilitate recidivism are neither subtle nor novel. They are evident to any competent observer interacting with the system for even a short period of time.

This places the platform in a condition of constructive knowledge. The harms are persistent, widely reported, repeatedly documented, and directly observable through routine use. Whistleblower testimony and internal research disclosures confirm that these effects were known within the organization long before they became matters of public controversy (Haugen 2021; Horwitz and Seetharaman 2021). Under ordinary standards applied to infrastructure, finance, or consumer safety, such knowledge would trigger an obligation to redesign the system. Continued operation without correction would be classified not as error but as negligence.

Claims that these problems are too complex to solve or require further study do not withstand scrutiny. The underlying principles are elementary. Identity that is not unique cannot support attribution. Enforcement that erases history cannot deter repeat abuse. Metrics that substitute for legitimacy will be gamed. These are not esoteric insights. They are foundational concepts taught in introductory computer science, systems engineering, and administrative law (Simon 1971; Lessig 1999). Their violation at planetary scale does not render them uncertain. It renders their consequences more severe.

## 17 Second-Order and Adaptive Harms at Population Scale

Beyond direct fraud and deception, large-scale platform governance produces second-order harms that operate through adaptive substitution and attentional displacement. When particular categories

of content are banned or heavily suppressed, including genuinely harmful material, the result is not elimination but mutation. Actors rapidly generate near-isomorphic substitutes that evade detection by differing just enough from what algorithms expect. This phenomenon is well understood in security contexts: rigid classifiers incentivize camouflage.

At scale, this produces an arms race in which increasingly distorted, sensationalized, or obfuscated content proliferates while enforcement mechanisms fall perpetually behind. The net effect is not moral improvement but increased sophistication of deception. When entire domains of legitimate informational content are restricted, such as news content in certain jurisdictions, engagement does not decrease. It is redirected toward lower-quality substitutes, including tabloid fragments, outrage clips, and low-information visual media that perform better under engagement metrics (Wu 2016; Bridle 2018).

The cumulative effect of these substitutions is attentional degradation. Attention is diverted away from activities that build long-term individual and societal capacity, such as learning languages, mathematics, engineering, medicine, or skilled trades, and toward content optimized for rapid consumption and emotional reaction. Over long time horizons and across billions of users, this constitutes a systematic reallocation of cognitive resources. As Floridi has argued, polluted informational environments degrade not only knowledge but agency itself (Floridi 2014).

## 18 Conclusion: Foreseeability, Responsibility, and the Possibility of Repair

The failures analyzed in this work are not comparable to historical infrastructure disasters whose harms were unknown or disputed at the time of deployment. Industrial pollutants and early public health hazards were often introduced under conditions of genuine uncertainty. By contrast, the trust degradation, fraud amplification, and attentional erosion produced by large-scale social platforms were evident almost immediately and have remained visible throughout their operation.

The core mechanisms are simple and widely understood. Non-unique identity enables impersonation. Disposable accounts enable recidivism. Engagement-optimized ranking amplifies deception. Suppressing structured information without altering incentives produces lower-quality substitutes. These dynamics are observed by ordinary users through routine interaction, documented by researchers and journalists, and acknowledged implicitly through reactive design changes. Foreseeable harm imposes a higher standard of responsibility.

Equally important, the existence of harm does not imply the absence of solutions. Persistent identity histories, expressive reporting mechanisms, shared defensive tools, and constraint-first architectures are well within current technical capability. Many exist in other domains of computing and governance. The platforms in question possess unparalleled volumes of data on abuse patterns, scam recurrence, and attention dynamics. If these tools are insufficient to detect structural harm, they are being misapplied. If they are sufficient but unused, the failure is institutional rather than technical.

At planetary scale, infrastructure is destiny. When failures are foreseeable and repairs achievable,

responsibility is no longer abstract. It is immediate.

## A Addendum: Institutional Rebranding and Legitimacy Transfer

The identity dynamics described throughout this work are not confined to user behavior. They are practiced at the institutional level by platform operators themselves. The corporate rebranding of Facebook to Meta illustrates the same non-injective identity mapping observed in scam behavior, whereby a persistent underlying entity adopts a new public identifier to shed accumulated reputational debt while preserving operational continuity.

This strategy is reinforced through association with high-prestige scientific and humanitarian initiatives, such as the Chan Zuckerberg Biohub, a nonprofit research organization founded by Mark Zuckerberg and Priscilla Chan to pursue ambitious goals in AI-powered biology. The value of such work is not in question. What is structurally relevant is the transfer of legitimacy across domains. Positive identity in one arena is allowed to offset negative identity in another despite the absence of causal separation. This is identity remix at institutional scale.

That these techniques are employed openly underscores the foreseeability of the harms discussed in this paper. The organization demonstrates a sophisticated understanding of how naming, affiliation, and legitimacy function as transferable assets. The failure to enforce analogous constraints within its own platforms therefore cannot plausibly be attributed to ignorance. It reflects selective application of identity integrity, enforced where advantageous and relaxed where profitable.

## References


[1] G. A. Akerlof. The market for "lemons": Quality uncertainty and the market mechanism. *Quarterly Journal of Economics*, 84(3):488–500, 1970.

[2] P. W. Anderson. More is different. *Science*, 177(4047):393–396, 1972.

[3] A.-L. Barabási. *Network Science*. Cambridge University Press, 2016.

[4] J. Bridle. *New Dark Age: Technology and the End of the Future*. Verso, 2018.

[5] D. T. Campbell. Assessing the impact of planned social change. *Evaluation and Program Planning*, 2(1):67–90, 1979.

[6] C. Doctorow. *The Internet Con: How to Seize the Means of Computation*. Verso, 2023.

[7] M. Elad. *Sparse and Redundant Representations*. Springer, 2010.

[8] L. Floridi. *The Fourth Revolution*. Oxford University Press, 2014.

[9] C. A. E. Goodhart. Problems of monetary management. Reserve Bank of Australia, 1975.

[10] F. Haugen. Testimony before the U.S. Senate Subcommittee on Consumer Protection. 2021.

[11] J. Horwitz and D. Seetharaman. The Facebook files. *The Wall Street Journal*, 2021.

[12] J. Lanier. *Ten Arguments for Deleting Your Social Media Accounts Right Now.* Henry Holt, 2018.

[13] L. Lessig. *Code and Other Laws of Cyberspace.* Basic Books, 1999.

[14] M. Scheffer et al. Early-warning signals for critical transitions. *Nature*, 461:53–59, 2009.

[15] C. E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27:379–423, 1948.

[16] H. A. Simon. Designing organizations for an information-rich world. Johns Hopkins University Press, 1971.

[17] Z. Tufekci. Algorithmic harms beyond Facebook and Google. *Colorado Technology Law Journal*, 2015.

[18] Y. Varoufakis. *Technofeudalism.* Melville House, 2023.

[19] T. Wu. *The Attention Merchants.* Knopf, 2016.

[20] S. Zuboff. *The Age of Surveillance Capitalism.* PublicAffairs, 2019.