EXTENDED ABSTRACT

Trust in Digital Spaces

An Application of Coleman's Concept of Trust to Online Interactions. *By Anna Lena Fehlhaber*

Trust is a foundational category of social order and cooperative action, yet the conditions under which it emerges are undergoing fundamental transformation in the digital age. Platforms, algorithms, and data-driven infrastructures are reshaping the dynamics through which individuals assess trustworthiness, enter into trust relationships, and expose themselves to the risk of trust violation. While Coleman's theory of trust has been extensively examined in analogous contexts, its application to digital communication environments remains largely unexplored: existing studies address intermediation processes only in isolated instances (cf. Ritzer-Angerer 2018). Against this backdrop, the central research question is: To what extent can Coleman's theories of trust be transferred to digital trust relationships, and what adaptations are normatively necessary from a media ethics perspective?

Coleman (1990) conceptualizes trust as a rational decision under uncertainty, anchored in methodological individualism. The trustor transfers a resource or control to the trustee in the expectation that the latter will not betray that trust. Since the trustee's response only becomes observable after the trustor has acted, an irresolvable temporal asymmetry arises, rendering trust a calculated wager. Coleman formalizes this: a rationally acting individual extends trust when the expected gain exceeds the expected loss, expressed as $p \cdot G > (1-p) \cdot L$, where p denotes the probability of expectation fulfillment, G the potential gain, and L the potential loss in the event of a trust violation. Central to the model are identifiable actors, functioning reputation mechanisms, and the possibility of sanctioning trust violations through intermediaries such as contracts or insurance. Granovetter (1985) criticizes the fact that Coleman's dyadic focus underestimates the embeddedness of trust in broader networks. Nevertheless, the model provides a flexible analytical framework for digital contexts, as it does not depend on physical co-presence but rests on the abstract logic of risk calculation under uncertainty.

Digital spaces, ranging from pseudonymous social media feeds and anonymous imageboards to darknet marketplaces, fragment or obscure the identities of interaction partners. Cole-

Anna Lena Fehlhaber serves as technology strategy lead for AI and cyber security and is a visiting researcher at the Hasso Plattner Institute and Hannover Medical School, focusing on security and trust in human-technology relationships.

man implicitly presupposes actors whose identities are sufficiently known to permit potential sanctioning or reputational harm (1990, pp. 175f.). In many digital settings, this precondition is substantially weakened: nonverbal cues are absent, reputational information is easily manipulated, and the informational basis for rational risk assessment is severely constrained. Empirical research on crypto markets has shown, however, that complete anonymity does not necessarily prevent the emergence of trust; instead, informal social and technical substitute mechanisms develop. This underscores the necessity of extending Coleman's model to encompass informal and dynamic trust-building processes.

A key distinction arises between platform-mediated contexts, such as eBay or Airbnb with their institutionalized rating and guarantee systems, and unregulated peer-to-peer relationships in messenger groups or decentralized darknet marketplaces. In platform-mediated environments, the platform functions as a digital intermediary in the Colemanesque sense: it structures communication, enforces participation rules, and enables sanctions such as account suspensions or refund claims, fully consistent with Coleman's (1990, p. 182) own discussion of institutional intermediaries. In unregulated settings, by contrast, trust must emerge without formal safeguards and relies on informal reputational signals and technical traceability. Coleman's core assumption of rational risk calculation remains operative in both cases, yet becomes considerably more demanding in the absence of institutional anchor points.

Digital communication operates at a speed and scale that far exceeds the stable, locally embedded social environments Coleman primarily analyzed (1990, p. 104). Real-time social proof mechanisms — ratings, likes, and comments — continuously update reputational signals, such that a single virally distributed trust violation can trigger a cascade of distrust. While Coleman acknowledges feedback effects, his analysis focuses on medium- to long-term dynamics in stable social systems. The temporal compression and global reach of digital networks generate highly volatile trust relationships, in which trust increasingly rests on snapshots rather than on established interaction histories. Technical possibilities such as the mass purchase of fabricated reviews or the artificial accumulation of reputation through intermediaries fundamentally undermine the informational preconditions of rational trust calculation (cf. He et al. 2023).

The most consequential departure from Coleman's framework arises from the growing role of machine learning-based algorithms as de facto trust intermediaries. These systems curate content, prioritize interactions, and produce rankings and reputation scores, yet lack the social foundation of personal relationships, clear norms, and collective negotiation processes that characterize Coleman's intermediaries. Their decision-making logic is frequently opaque, potentially biased, and barely contestable by users (cf. Rudin 2019). Explainable AI (XAI) approaches address this opacity through methods such as feature importance or counterfactual explanations, yet they create merely cognitive interpretability — they substitute neither personal accountability nor the collective norm formation that Coleman presupposes as the basis of rational calculation.

Critics of Coleman's rationalist approach emphasize its limited capacity to capture cultural and emotional dimensions of trust. Digital interaction spaces make these deficits particularly apparent: affective dynamics such as online firestorms and virally disseminated social media campaigns can destabilize established trust relationships far more rapidly than rational recalculation would suggest. Cultural norms and habitualized practices shape the interpretation of trust signals in ways that vary considerably across platform contexts — a finding that illustrates that digital trust cannot be reduced to individual calculation but is structured by sociocultural conditions.

The analysis demonstrates that Coleman's model of rational trust decisions under uncertainty remains fundamentally valid in digital spaces, yet requires substantial extension along technical, cultural, and affective dimensions. Platforms equipped with rating and guarantee systems function as Colemanesque intermediaries; however, the opacity of algorithmic processes, the manipulability of reputation scores, and the volatility of affective dynamics considerably complicate and distort the rational trust calculus. For communication studies and media ethics, this entails a twofold platform responsibility: first, to design identity and reputation architectures that effectively constrain manipulation; and second, to address affective escalation dynamics, such as firestorms, through transparent moderation practices, rather than amplifying them algorithmically. Future-oriented trust research will need to draw more extensively on interdisciplinary approaches spanning communication studies, psychology, and computer science in order to adequately capture the interplay of algorithmic infrastructure, platform governance, and human agency in digital publics.

References

Coleman, James S. 1990. Foundations of Social Theory. Cambridge, MA.

Granovetter, Mark. 1985. "Economic Action and Social Structure: The Problem of Embeddedness." American Journal of Sociology 91, no. 3: 481–510.

He, Sherry, Brett Hollenbeck, and Davide Proserpio. 2023. "Leveraging Social Media to Buy Fake Reviews." Communications of the ACM 66, no. 10: 98–105. https://doi.org/10.1145/361542.

Ritzer-Angerer, Petra. 2018. "Digital Transformation and Trust (Dis)intermediation – Coleman's Models Thirty Years On." SSRN Electronic Journal. https://doi.org/10.2139/ssrn.3271566.

Full Article (German):
Anna Lena Fehlhaber: Vertrauen in digitalen Räumen. Eine Anwendung von Colemans Vertrauenskonzept auf Online-Interaktionen. In: Communicatio Socialis (2/2026) 59: 282-293. DOI: 10.5771/0010-3497-2026-2-282