The current issue and full text archive of this journal is available on Emerald Insight at:
https://www.emerald.com/insight/0263-5577.htm

IMDS
125,11

3024

Toward trustworthy content: the role of challengers, juries and veracity bonds in digital media platforms

Lucas Barbosa
PwC Australia, Sydney, Australia
Sam Kirshner
UNSW Sydney, Sydney, Australia
Rob Kopel
PwC Australia, Brisbane, Australia
Eric Tze Kuan Lim
UNSW Sydney, Sydney, Australia, and
Tom Pagram
PwC Australia, Sydney, Australia

Received 27 November 2024
Revised 1 February 2025
31 March 2025
Accepted 2 May 2025

Abstract

Purpose – Public trust in digital media and online content has declining considerably over the past 50 years. Traditional interventions such as fact-checking, warning labels, content classification systems and artificial intelligence (AI) detection tools have inherent limitations in both scalability and effectiveness. This study explores decentralized governance mechanisms for content trust, including challengers, juries and blockchain-based veracity bonds and evaluates their impact on content quality and trust.

Design/methodology/approach – Three scenario-based experiments were conducted to examine user perceptions of credibility, author commitment, procedural justice and fairness. Study 1 introduced a challenge mechanism to dispute content accuracy. Study 2 implemented a jury mechanism for impartial evaluation of challenges. Study 3 explored the use of financial stakes through veracity and counter-veracity bonds to promote accountability.

Findings – The challenge mechanism significantly enhanced perceptions of credibility by empowering users to scrutinize content. Veracity bonds improved perceived commitment by signaling accountability through financial stakes, while counter-veracity bonds reduced fairness perceptions. We found that jury mechanisms without financial bonds had limited impact on content commitment.

Practical implications – Features such as challengers and veracity bonds help authors demonstrate a commitment to accuracy and quality, enhance accountability by incentivizing truthfulness alongside the inherent goals of attention and virality and encourage community participation in countering misinformation.

Originality/value – This study offers an innovative approach to rebuilding trust in digital media and online content by introducing novel evaluation mechanisms, including veracity bonds and counter-veracity bonds. We extend the Heuristic-Systematic Model of information processing by introducing a new dimension: the Investment Heuristic. Unlike conventional heuristics based on nudges or fact-checking, this approach integrates financial and participatory incentives directly into the platform's design.

Keywords Digital media, Incentives, Content trust, Decentralized governance, Misinformation

Paper type Research paper

© Lucas Barbosa, Sam Kirshner, Rob Kopel, Eric Tze Kuan Lim and Tom Pagram. Published by Emerald Publishing Limited. This article is published under the Creative Commons Attribution (CC BY 4.0) licence. Anyone may reproduce, distribute, translate and create derivative works of this article (for both commercial and non-commercial purposes), subject to full attribution to the original publication and authors. The full terms of this licence may be seen at http://creativecommons.org/licences/by/4.0/legalcode

Blinded for review.

Order of the authorship is arranged alphabetically by last name and every author has contributed equally to the paper.

Industrial Management & Data Systems
Vol. 125 No. 11, 2025
pp. 3024-3048
Emerald Publishing Limited
e-ISSN: 1758-5783
p-ISSN: 0263-5577
DOI 10.1108/IMDS-11-2024-1177

Downloaded from http://ftp.nowpublishers.com/imds/article-pdf/125/11/3024/10941121/imds-11-2024-1177en.pdf by guest on 24 June 2026

Introduction

An increasing body of research highlights a steady erosion of public trust in the media ecosystem (Fletcher *et al*., 2025). The trust decline is often attributed to the spread of false or misleading information through social media and content-sharing platforms, which are known to prioritize user engagement metrics over other metrics such as content credibility (Globig and Sharot, 2024). In a 2021 poll, 95% of Americans viewed misinformation as a widespread and serious issue in contemporary society [1], however, distrust in media was a phenomenon that started prior to the diffusion of social media with a long-running poll [2] suggesting that Americans' trust in mass media has been declining since 1976, with no significant recovery over the past five decades (Figure 1).

Past research identifies a range of potential factors that contribute to trust erosion, including increased media fragmentation (Peng and Yang, 2022), the formation of echo chambers (Donkers and Ziegler, 2021), political polarization (Suiter and Fletcher, 2020), perceived biases (Ardèvol-Abreu *et al*., 2017), and the proliferation of AI-generated and synthetic content (Morosoli *et al*., 2024). Without appropriate intervention, there may continue to be profound consequences for democratic processes and the collective understanding of critical issues such as climate change and public health (Bharti, 2020; Linden *et al*., 2017).

In response, researchers and firms have explored interventions, such as fact-checking initiatives, classification systems designed to label misleading content, tools for detecting AI-generated content, and features to promote the critical analysis of information (e.g. users rating the credibility of news articles; Moravec *et al*., 2022). Yet, the effectiveness of current interventions vary considerably based on individual user behaviors and the socio-political leanings of media platforms (Aslett *et al*., 2023). For example, warning labels applied to fake news stories were found to inadvertently increase trust in unlabeled content for some users – a phenomenon known as the "Implied Truth Effect" (Pennycook *et al*., 2020). In other scenarios, users demonstrated confirmation bias, leading to a rejection of warning labels and fact-checking (Moravec *et al*., 2019).

For interventions to be successful, they must consider the profit motives of commercial content platforms. Most commercial content platforms earn revenue through the advertising that they serve and are incentivized to prioritize attention-maximizing content, such as sensational or polarizing content, over the dissemination of accurate and credible content (Globig and Sharot, 2024). Until incentive structures are aligned to prioritize accuracy and

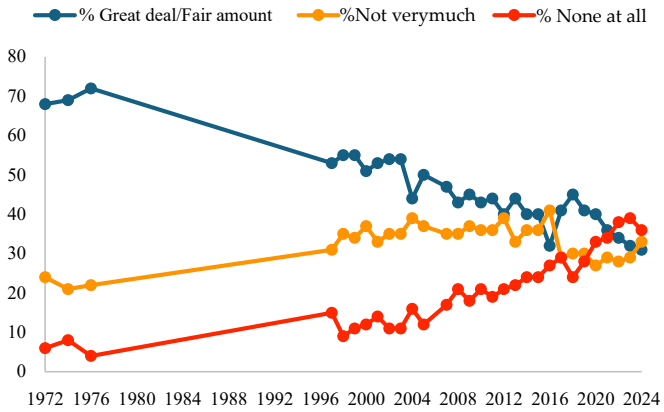

Figure 1. Percentage responses to: "In general, how much trust and confidence do you have in the mass media – such as newspapers, TV and radio – when it comes to reporting the news fully, accurately and fairly – a great deal, a fair amount, not very much or none at all?" (Gallup Poll). Source: Recreated from source at Gallup Poll – https://news.gallup.com/poll/651977/americans-trust-media-remains-trend-low.aspx

Downloaded from http://ftp.nowpublishers.com/imds/article-pdf/125/11/3024/10941121/imds-11-2024-1177en.pdf by guest on 24 June 2026

reliability it will be difficult to implement effective, scalable solutions for restoring trust in the quality and credibility of the information that the platforms serve. Inspired by distributed affordances of Blockchain and novel governance models, digital media and content-sharing platforms are beginning to experiment with novel intervention techniques to better align incentives. For example, Civil leveraged blockchain to promote transparency through token-based governance systems (Al-Saqaf and Edwardsson, 2019), and Steemit incentivizes content creators using cryptocurrency rewards for engagement with platform goals (Liu *et al*., 2022).

Blockchain-based decentralized journalism presents a compelling opportunity to address the distrust issues affecting digital media and content platforms by enforcing contractual transparency and accountability. Blockchain's decentralized protocol ensures that all information is immutable and verifiable to resist tampering or manipulation (Tyma *et al*., 2022) and enables the use of smart contracts to automate and enforce journalistic standards, ensuring that all parties adhere to pre-agreed ethical guidelines and transparency protocols (Spanò *et al*., 2022). Immutable and verifiable smart contracts can incorporate mechanisms to incentivize truth-telling and conciliation by embedding reward systems or penalties based on verifiable outcomes.

In this paper we explore a set of mechanisms for addressing the core issue of incentives alignment issue by encouraging the prioritizing of content credibility and accountability in a scalable manner alongside attention and engagements metrics. Specifically, we propose a system with four mechanisms: a challenge mechanism, a jury-based adjudication system, the novel concepts of veracity (Hoskinson, 2023) and counter-veracity bonds as financial incentives, and blockchain-based smart contracts for distributed governance.

First, the proposed system would allow users to challenge the content posted by authors. The challenge mechanism empowers users to dispute misinformation rather than relying solely on centralized authorities. Research on crowdsourced verification and participatory fact-checking suggests that user-driven scrutiny enhances transparency and strengthens perceptions of fairness. A real-world parallel can be found in Twitter's "Community Notes" (formerly Birdwatch), which allows users to collaboratively identify misleading posts and provide explanatory annotations [3]. The success of features such as Community Notes and Wikipedia's peer-editing system highlights the role of collective engagement in verifying content accuracy. By allowing users to challenge the veracity of content, the platform increases the detection of misinformation while promoting greater user involvement in the credibility assessment process.

Research by Drolsbach *et al*. (2024) highlights the effectiveness of this approach, demonstrating that community-generated annotations increase user trust compared to simple misinformation flags. This increase in trust stems from both the transparency of peer-generated content and the inclusion of fact-checking explanations, which help users better understand why content might be misleading. By incorporating a similar crowdsourced challenge mechanism, we aim to foster critical engagement and accountability, prompting the first research question (RQ1): How would the presence of challengers impact the perception of the media platform?

Second, we introduce a jury-based adjudication system to enhance perceptions of fairness and procedural justice. A key reason for declining trust in digital content and media platforms is the belief that moderation and fact-checking decisions are biased, particularly when controlled by a central authority. Instead of relying on opaque moderation systems, our approach entrusts content disputes to a randomly selected jury of impartial users. The implementation of juries draws on principles from decentralized governance, including the use of dispute resolution models such as Kleros and Minds' jury appeals system. Research in procedural justice suggests that individuals are more likely to accept outcomes when they perceive decision-making processes to be fair, even when the results are unfavorable. By ensuring that disputes are resolved transparently and collectively, the jury mechanism aims to reinforce trust in platform governance.

Downloaded from http://ftp.nowpublishers.com/imds/article-pdf/125/11/3024/10941121/imds-11-2024-1177en.pdf by guest on 24 June 2026

We propose that the outcomes of challenges should be evaluated by juries composed of impartial participants. The jury system would act as a decentralized arbiter so that that the validity of challenges can be assessed fairly. By entrusting the resolution of disputes to a decentralized jury, rather than a central authority, we can promote a sense of procedural justice. A similar system is employed by the social network Minds, which has implemented a Jury System to review appeals. On Minds, a jury consists of 12 randomly selected active users who are not subscribed to the user under review, ensuring impartiality [4]. Jurors can choose to participate, pass, or opt out of the jury pool entirely. Decisions are made based on majority agreement, with at least 75% of jurors required to vote in favor for an appeal to be upheld. Drawing inspiration from such implementations, our proposed jury mechanism seeks to replicate these benefits, leading to our second research question (RQ2): How does having a jury to examine challenges influence the perceived commitment of a media platform?

Third, financial incentives play a critical role in encouraging responsible content creation beyond procedural fairness. To embed direct accountability, we introduce veracity and counter-veracity bonds to promote accountability for responsible content creation. In such a model, authors would be required to post a veracity bond when they publish content, and challengers would post a counter-veracity bond when challenging an author's content. These are financial stakes in the form of bonds attached to claims of veracity, which are forfeited depending on the outcome of the challenge. If a jury upholds the challenge, the author forfeits their veracity bond, while the challenger's counter-veracity bond is forfeited if the challenge is denied.

Unlike passive fact-checking interventions, which rely on nudging users toward more accurate content, financial bonds create tangible incentives for accuracy. Drawing on signaling theory and behavioral economics, the veracity bond mechanism requires authors to stake financial resources on the accuracy of their posts. These mechanisms introduce "skin in the game" principles that reinforce credibility by ensuring that both content creators and challengers face consequences for misinformation or unsubstantiated claims. This system of bonds incentivizes both parties to act in good faith and could support a higher level of accountability, as both authors and challengers face tangible consequences for their actions.

We are not aware of content platform providers that currently operate using veracity bonds, however there is discussion amongst industry practitioners within decentralized communities (e.g. Charles Hoskinson [5], founder of Cardano) in its ability to curb disinformation. Thus, our third research question (RQ3) investigates: How do veracity and counter-veracity bonds influence the perceived commitment of a media platform?

The fourth mechanism in this system is decentralized governance, replacing a central authority with a transparent protocol to oversee content disputes and to enforce content policies. By distributing control among users, it is possible to encourage greater participation in decision-making processes and to positively influence users' perceptions of the platform's commitment to transparency and fairness. This leads to our final research question (RQ4): How does decentralized governance affect the perceived commitment of a media platform?

The selection of these mechanisms was informed by their ability to realign incentives toward accuracy and procedural fairness. While algorithmic fact-checking and automated moderation could provide alternative solutions, these approaches introduce risks related to algorithmic bias and lack of transparency. Instead, challenges, juries, and veracity bonds offer a more participatory and decentralized approach to trust-building. Collectively, these mechanisms foster a media environment where users actively engage in content verification, where disputes are resolved through transparent and inclusive processes, and where financial incentives deter misinformation. By integrating these components, platform can have a multi-layered framework that strengthens credibility and accountability in a manner that is resilient to both cognitive biases and structural weaknesses in traditional media governance.

Our research tackles structural misalignments of content creation and dissemination on media platforms and has the following contributions:

Downloaded from http://ftp.nowpublishers.com/imds/article-pdf/125/11/3024/10941121/imds-11-2024-1177en.pdf by guest on 24 June 2026

(1) First, we provide a proof of concept of innovative platform mechanisms (i.e. challenges, juries, and veracity bonds), showing their potential for enhancing media credibility and accountability.

(2) Second, we introduce the novel concepts of veracity and counter-veracity bonds to the academic literature, offering a new lens for aligning incentives in digital media.

(3) Third, to theoretically address our research questions, we propose a framework for understanding trust-building, highlighting the role of procedural fairness and decentralized governance.

(4) Finally, to develop our hypotheses, we extend the Heuristic-Systematic Model (HSM) by proposing the "Investment Heuristic," a psychological cue that highlights the role of financial stakes in fostering perceived commitment to truth and credibility of the platform.

Taken together, our study offers an innovative approach for addressing issues of media trust and misinformation in the digital age and provides many avenues for future research.

Theoretical background and hypothesis development

Theoretical background

Signaling Theory. Signaling theory is used to understand how individuals or entities communicate information when one party has more information than the other. It is relevant across fields like economics, biology, and anthropology and helps explain behaviors such as advertising, courtship, and social interactions (Bird and Smith, 2005). Signals can be either costly or cost-free, each with implications for credibility. Costly signals require significant investment or risk, making them more reliable, whereas cost-free signals, also known as "cheap talk," [6] may still convey valuable information in certain contexts (Bergstrom and Lachmann, 1998; Silva and Sigmund, 2024). Costly signaling ensures that only those who genuinely possess the qualities being signaled can afford to send the signal, thus enhancing credibility. Signaling theory applies to advertising, where companies invest in expensive campaigns to signal product quality, and social behaviors, where individuals display costly traits to signal competence or commitment (Noble, 1999; Johnston, 1999). These signals help receivers make informed decisions, reducing uncertainty under asymmetric information.

Signaling theory serves as an overarching lens for this study because it directly addresses how individuals interpret and respond to costly signals in environments characterized by information asymmetry (Bird and Smith, 2005; Bergstrom and Lachmann, 1998). Media platforms, especially decentralized ones, operate in contexts where users must evaluate the credibility of content creators and processes without direct interaction. The theory's emphasis on "skin in the game" provides a robust explanation for how mechanisms like veracity bonds function as credible signals of commitment to truthfulness. Moreover, signaling theory's applicability across domains (e.g. advertising, biology, and organizational behavior) highlights its relevance to media trust, where financial and procedural signals are central to user perceptions (Noble, 1999; Johnston, 1999).

While signaling theory provides the foundation for understanding how visible cues influence credibility, psychological mechanisms underpin how users perceive and process these signals. Specifically, we focus on the Heuristic-Systematic Model (HSM), which explains how individuals use heuristic cues, such as financial stakes, can lead to credibility judgments without engaging in cognitively demanding systematic evaluation (Chaiken, 1980). Additionally, Cognitive Load Theory suggests that clear, effortful signals reduce information processing demands, fostering greater user trust (Sweller, 1988). By integrating these theories, we utilize signaling theory as the overarching abstract guiding framework as the scaffold for integrating the psychological drivers at the more operational level behind user perceptions of trust and fairness in content platforms.

Downloaded from http://ftp.nowpublishers.com/imds/article-pdf/125/11/3024/10941121/imds-11-2024-1177en.pdf by guest on 24 June 2026

Heuristic-Systematic Model. HSM explains how individuals process information to form judgments, operating under two distinct modes: systematic and heuristic processing (Davis and Tuttle, 2013; Chaiken, 1980). Systematic processing involves deliberate and effortful analysis of information, requiring cognitive engagement to evaluate its relevance and credibility (Chauhan and Gupta, 2024). In contrast, heuristic processing relies on mental shortcuts or simple decision rules, such as "experts can be trusted" or "long messages are valid," allowing for quicker judgments with less cognitive effort (Chaiken and Maheswaran, 1994).

HSM's dual-process framework has been widely applied across domains such as disaster response, e-commerce, and online communities, where users must frequently make rapid judgments in information-rich environments (Davis and Tuttle, 2013). For instance, in disaster scenarios, time-sensitive decision-making requires individuals to rely on heuristic cues like source credibility or message length due to overwhelming volumes of information (Sun and Xie, 2024). Similarly, in e-commerce, users often depend on heuristics such as product ratings and reviews to evaluate credibility, bypassing more systematic scrutiny (Chauhan and Gupta, 2024). These examples highlight HSM's relevance in contexts characterized by environmental and cognitive constraints, where efficient information processing is essential.

In today's information-saturated environment, users are increasingly reliant on heuristic processing. Social media platforms expose individuals to overwhelming volumes of content, leading to cognitive overload (Sun and Xie, 2024). Under such conditions, heuristic cues enable users to quickly process and evaluate information without expending significant mental resources (Sundar, 2008; Sun and Xie, 2024). This preference for heuristic reasoning is amplified in time-sensitive or cognitively constrained contexts, where efficiency in decision-making becomes paramount (Davis and Tuttle, 2013). These shortcuts often influence how users assess the credibility and trust of online content.

Sun and Xie (2024), in their meta-analysis, identify five key types of heuristics that influence user judgments in online environments: endorsement, self-confirmation, authority, affective, and fact-checking heuristics. Each plays a distinct role in shaping how users process information and form judgments, often based on simple, easily interpreted cues.

(1) Endorsement Heuristic: Information endorsed by others, such as through likes or shares, is perceived as credible, leveraging the bandwagon effect (Sundar, 2008). However, this heuristic can inadvertently amplify misinformation when popularity metrics signal false credibility (Sun and Xie, 2024).

(2) Self-Confirmation Heuristic: Users are more likely to trust and share content consistent with their pre-existing beliefs, even if it is explicitly labeled as false (Moravec *et al.*, 2019). Motivated reasoning further reinforces this behavior, as individuals prioritize congruent information to reduce cognitive effort (Kunda, 1990).

(3) Authority Heuristic: Credibility judgments often rely on the perceived expertise or reputation of the information source (Metzger and Flanagin, 2013). However, Sun and Xie (2024) note that source credibility alone may not always predict behavior, particularly when content quality conflicts with authority signals.

(4) Affective Heuristic: Emotional responses, such as anxiety, play a significant role in influencing judgment and sharing behavior. Anxiety, in particular, has been shown to increase misinformation sharing as users seek to alleviate discomfort by disseminating content (Zajonc, 1980).

(5) Fact-Checking Heuristic: The presence or absence of fact-checking cues, such as "verified" or "disputed" labels, significantly impacts credibility judgments. These cues act as cognitive shortcuts, allowing users to quickly assess content reliability (Pennycook *et al.*, 2020).

Downloaded from http://ftp.nowpublishers.com/imds/article-pdf/125/11/3024/10941121/imds-11-2024-1177en.pdf by guest on 24 June 2026

HSM is particularly well-suited for studying user perceptions in decentralized media platforms due to its dual-process framework. The model accounts for how users engage with signals such as challenges, juror decisions, and veracity bonds (Hoskinson, 2023) primarily through heuristic strategies. By offering features that provide clear and accessible cues, the platform reduces the cognitive burden on users evaluating credibility and encourages efficient decision-making. This approach ensures that the platform remains intuitive and accessible to a broad range of users, regardless of their motivation or capacity to process information in detail.

Procedural Justice Theory. Procedural justice plays a crucial role in influencing the trustworthiness of information in decision-making processes. It refers to the fairness of the processes that lead to outcomes, and its perception can significantly impact how individuals trust the information provided by authorities. When decision-making processes are perceived as fair, individuals are more likely to trust the information and decisions made by authorities, which in turn affects their acceptance and satisfaction with the outcomes (Tyler, 2000). This relationship is mediated by factors such as the perceived ability, benevolence, and integrity of the authority figures involved in the decision-making process.

Procedural justice enhances trust in authorities by ensuring that decision-making processes are transparent, participatory, and respectful. When individuals perceive that they have been treated fairly, they are more likely to trust the motives and integrity of the decision-makers, which enhances the perceived trustworthiness of the information provided (Bos *et al*., 1998). People are more willing to accept decisions when they believe the procedures used to reach those decisions are fair, even if the outcomes are unfavorable. This acceptance is crucial, as fair procedures can mitigate negative reactions and foster trust in the decision-making process (Bianchi *et al*., 2015). Trust in decision-making authorities also dictates how procedural justice influences fairness perceptions. When trust in authorities is low, fair procedures become even more critical in shaping positive perceptions (Bianchi *et al*., 2015).

Hypothesis development

This study integrates signaling theory with psychological theories of decision-making to propose a comprehensive framework. Signaling theory explains the role of costly signals, such as veracity bonds, in enhancing perceived credibility and commitment (Bird and Smith, 2005). Simultaneously, procedural justice theory highlights the importance of fairness perceptions (Tyler, 2000; Bos *et al*., 1998), while psychological mechanisms like HSM and cognitive biases elucidate how users process these signals. For example, veracity bonds not only serve as costly signals but also align with cognitive heuristics, enabling users to evaluate trustworthiness efficiently (Pennycook and Rand, 2019). This multi-theoretical perspective strengthens the basis for the following hypotheses, which address credibility, commitment, and procedural fairness.

The challenge feature operates as a form of costly signaling, whereby users invest time and effort to contest content they find questionable, showcasing their commitment to accuracy (Bird and Smith, 2005). This mechanism not only aligns with Signaling Theory, which posits that signals involving cost or risk enhance credibility (Aumann and Hart, 2003), but also leverages the Fact-Checking Heuristic, a cognitive shortcut that enables users to evaluate content credibility based on verification cues. Fact-checking heuristics are triggered when users encounter signals indicating that information has been scrutinized, such as visual flags, labels, or annotations highlighting disputed or verified content (Vu and Chen, 2023). Similarly, the presence of a challenge feature provides an implicit cue that contested content has undergone review by others, signaling to users that the platform promotes transparency and accountability. These cues reduce cognitive effort by allowing users to rely on easily interpreted signs of verification, rather than engaging in extensive evaluation themselves (Chaiken and Maheswaran, 1994; Vu and Chen, 2023).

Empirical research shows the effectiveness of fact-checking cues in enhancing trust and reducing misinformation believability (Moravec *et al*., 2020; Pennycook *et al*., 2020).

Downloaded from http://ftp.nowpublishers.com/imds/article-pdf/125/11/3024/10941121/imds-11-2024-1177en.pdf by guest on 24 June 2026

By incorporating a participatory challenge mechanism, the platform provides a community-driven approach to fact-checking, reinforcing perceptions of credibility and fostering trust.

H1. The availability of a challenge feature on the platform will increase perceived platform credibility.

Signaling Theory posits that when content creators subscribe and are playing within a standard or consistent set of rules geared toward ensuring accuracy (e.g. through responsible reporting practices or openly accepting challenges) they signal a strong commitment to credibility (Bird and Smith, 2005). This is particularly effective when creators demonstrate "skin in the game", that is, when they bear some risk of loss if they fail to meet credibility standards, which can be represented through mechanisms like veracity bonds (Bergstrom and Lachmann, 1998). Just as companies use costly advertising to signal product quality, creators who take tangible actions to back their claims build trust among users, who may be reassured by the transparency and risk-sharing (Bianchi *et al*., 2015). Studies indicate that when information sources display genuine, costly investments, audiences are more inclined to trust them, thereby boosting platform credibility.

H2. Higher perceived commitment of producing truthful content by creators will positively influence perceived platform credibility.

Procedural Justice Theory suggests that fair, inclusive processes can strongly influence users' trust in decision-making structures (Tyler, 2000). Decentralized governance embodies procedural fairness by distributing control across users, fostering a more balanced authority that minimizes biases typically associated with centralized governance (Bos *et al*., 1998). By removing traditional principal-agent relationships, decentralized governance also reduces power imbalances, reinforcing users' perceptions of transparency and impartiality. This aligns with Signaling Theory, as the transparent nature of a decentralized system signals integrity and organizational commitment to fairness (Bird and Smith, 2005). Thus, the governance structure of the platform itself becomes a credible mechanism, assuring users that decisions are made collaboratively and equitably.

H3. Decentralized governance structures will positively influence perceived platform credibility.

Procedural Justice Theory emphasizes that trust in decision-making stems largely from perceptions of fairness and transparency within the process (Tyler, 2000). When users believe that the processes guiding content and challenges are fair, they are more likely to view the platform itself as trustworthy and credible (Frazier *et al*., 2010). Fair processes foster a sense of respect and impartiality, leading users to trust both the information on the platform and the intentions of its creators. This relationship between procedural justice and trust has been demonstrated across multiple studies, where transparent and equitable procedures have been shown to significantly enhance perceptions of trustworthiness, especially in environments with asymmetrical information (Bianchi *et al*., 2015).

H4. Greater procedural justice perceptions will lead to higher perceived platform credibility.

The jury's role in evaluating content challenges aligns with the Endorsement Heuristic, where users interpret the collective judgment of a group as a signal of credibility and fairness (Sun and Xie, 2024). The jury system provides a clear heuristic cue, since the content is endorsed or validated by a diverse group of impartial jurors, which is perceived as more credible and trustworthy (Sundar, 2008; Metzger and Flanagin, 2013). This collective evaluation reduces the cognitive effort required by users to assess the fairness of content challenges, as they can rely on the aggregated judgment of the jury as a proxy for integrity and accuracy.

The wisdom of the crowd effect further strengthens this perception, as decisions made by a representative body of jurors are seen as less prone to bias and more aligned with community

Downloaded from http://ftp.nowpublishers.com/imds/article-pdf/125/11/3024/10941121/imds-11-2024-1177en.pdf by guest on 24 June 2026

standards (Pennycook and Rand, 2019). By embedding this decentralized, community-driven mechanism, the platform reinforces its commitment to truthfulness and procedural fairness, motivating creators to uphold high standards under scrutiny from a collective body. The jury system thus serves as both a functional and symbolic representation of the platform's dedication to fostering trust and accountability. Thus, we posit:

H5. The presence of a jury evaluating challenges will positively impact the perceived commitment of producing truthful content by creators.

While mechanisms like challenges and juries align neatly with existing heuristic types such as the Fact-Checking Heuristic and Endorsement Heuristic, respectively, veracity bonds do not fit within the heuristic taxonomy identified by Sun and Xie (2024). This is unsurprising, as the taxonomy emerged from a meta-analysis of established systems, whereas veracity bonds represent a novel concept. Veracity bonds go beyond traditional mechanisms by embedding financial stakes as a signal of accountability, creating a distinct pathway for fostering credibility.

Drawing on Signaling Theory, veracity bonds function as a form of costly signaling, where creators demonstrate "skin in the game" by risking personal assets and reputation to vouch for the truthfulness of their content (Bird and Smith, 2005). These financial or reputational risks provide a tangible signal to users, showcasing the creator's commitment to accuracy. This signaling mechanism can be conceptualized as the Investment Heuristic: users interpret visible, high-stakes investments as a cognitive shortcut for assessing credibility. Much like performance bonds in corporate contexts, where financial stakes signal commitment to deliver on promises (Farrell, 1995), veracity bonds communicate a strong dedication to truthfulness and accountability.

The Investment Heuristic simplifies the judgment process by allowing users to infer credibility based on the presence of financial or reputational stakes, reducing the cognitive effort required for systematic evaluation. When users see that creators have something to lose, they are more likely to trust the integrity of the content being shared (Silva and Sigmund, 2024). By embedding this novel heuristic into its design, the platform not only ensures that creators are held accountable but also reinforces user trust in its governance framework. Thus, the requirement of veracity bonds enhances perceptions of a creator's commitment to producing truthful content, leading to the following hypothesis:

H6. The requirement of posting veracity bonds from content creators will positively impact the perceived commitment of producing truthful content by creators.

Procedural Justice Theory suggests that perceptions of fairness are maximized when both parties, in this case, creators and challengers, are equally accountable (Tyler, 2000). By requiring both veracity and counter-veracity bonds, the platform ensures that creators and challengers alike have "skin in the game," promoting a balanced power dynamic and fostering perceptions of fairness (Bos *et al*., 1998). This system of dual bonds signals to users that the platform maintains an impartial and just structure, where all participants are equally subject to risk, reducing frivolous challenges and reinforcing integrity (Bird and Smith, 2005).

H7. The combined presence of requiring veracity bonds for content creators and counter-veracity bonds for challengers will reinforce procedural justice perceptions of the platform.

Methodology

Overview of studies

To empirically investigate how various features of a media platform influence perceptions of credibility, commitment, and fairness, three independent studies were conducted using scenario-based experiment with online questionnaires, where participants, recruited from the

Downloaded from http://ftp.nowpublishers.com/imds/article-pdf/125/11/3024/10941121/imds-11-2024-1177en.pdf by guest on 24 June 2026

UK via Prolific, interacted with a hypothetical media platform called MediaSphere. The platform mimicked existing social media environments, such as Twitter/X or Facebook, allowing real-time interaction and participation. Each study introduced new platform mechanisms, building on the previous study, to evaluate how these features impacted user perceptions. The complete materials for all studies are provided in the Online Appendices.

The data collection process and experimental design focused on evaluating high-level design principles of a proposed media platform, such as challenges, juries, and veracity bonds, rather than specific operational or technical implementations. Participants were presented with detailed descriptions of these principles and asked to conceptualize how they might function in their ideal states. This approach intentionally avoided relying on an operational experimental platform or visual interfaces, as these could constrain participants' interpretations or introduce biases tied to specific designs. Instead, participants were encouraged to engage meaningfully with the abstract concepts through follow-up questions that ensured they understood the descriptions and could articulate how the proposed mechanisms might function in their ideal forms. The goal was to assess the fundamental effectiveness of the design principles themselves, rather than their implementation details. While concrete examples or operational features could have been included, they were intentionally excluded to avoid constraining participants' ability to imagine and engage with the abstract concepts fully. This method ensured that participants could thoughtfully engage with the abstract principles under investigation while minimizing biases introduced by specific technical designs.

Procedures and materials

Study 1. The first study ($N = 85$) examined the effect of a challenge feature on perceived credibility. In the study, participants first read about a platform called MediaSphere (Appendix F). They were informed that MediaSphere is a social media platform designed for sharing news content from both mainstream and independent sources, with features similar to those found on platforms like Twitter and Facebook. The description emphasized MediaSphere's focus on fostering real-time interactions and discussions, further contextualizing the platform. Participants [7] were then introduced to MediaSphere's governance structure, which varied depending on the assigned treatment: the platform was described as either operating under a centralized model, where decisions were controlled by a single authority, or a decentralized model, where decision-making power was distributed among users.

Following an attention check on the platform governance structure, participants were then randomized to either the control group, where no challenge feature was available, or the treatment group, where users could challenge the accuracy of posts by submitting evidence or corrections. The control group participants then answered the survey questions, while participants in the treatment group were provided the following description: "MediaSphere allows users to challenge posts by presenting supporting evidence or corrections. If the challenge is successful, the post will either be tagged with a modification notice or removed from the platform entirely. These changes are visible to all users on the platform."

After reading about the platform, participants answered questions designed to measure credibility, focusing on the reputation, accuracy, and quality of the content. These measures, which were collected using a 7-point Likert scale, were adapted from Flanagin and Metzger (2007) and McKnight and Kacmar (2007). We also measured perceptions of author's commitment. Specifically, how the presence of the jury mechanism influenced perceptions of content creators' responsibility and accountability. These measures were adapted from Meyer and Allen (1991) and Skinner (1996). At the end of the study, participants provided demographic information, including gender, age, country of residence, education level, political affiliation, frequency of news consumption, and social media use. Gender and political affiliation were provided as multiple-choice options, while education level and news consumption frequency were rated on ordinal scales.

Downloaded from http://ftp.nowpublishers.com/imds/article-pdf/125/11/3024/10941121/imds-11-2024-1177en.pdf by guest on 24 June 2026

Study 2. Study 2 ($N = 98$) built upon the findings of Study 1 by introducing a jury mechanism to adjudicate challenges. Study 2 had the same procedure and text as the treatment group of Study 1. Thus, all participants were exposed to the challenge feature. However, they were randomly assigned to either a no-jury condition (control group), where challenges were simply flagged but not formally reviewed, (this is the same as the treatment group of Study 1) or a jury condition, where challenges were evaluated by a randomly selected group of platform users. These users voted on the validity of the challenge, and their decisions, along with the reasoning behind them, were made visible to all participants on the platform.

Participants were given the following explanation in the jury condition: "When a challenge is made, a randomly selected jury of platform users will review the evidence provided by the challenger and vote on the validity of the claim. The jury's decision, along with their reasoning, will be visible to all users." In addition to measuring credibility and commitment, Study 2 included measures of procedural justice, which assessed participants' perceptions of fairness in the process and outcomes related to challenges. These measures were adapted from Colquitt (2001), Leventhal (1976), and Leventhal (1980). The measures for credibility, commitment, and procedural justice are provided in Table A.1.

Study 3. Study 3 ($N = 192$) introduced financial mechanisms in the form of veracity bonds and counter-veracity bonds. All participants were exposed to both the challenge feature and the jury mechanism, as described in Study 2. In Study 3, the participants were randomly assigned to one of four conditions based on the presence of financial bonds: (1) no bonds, (2) veracity bond only, (3) counter-veracity bond only, or (4) both veracity bond and counter-veracity bond. Thus, the no bond condition is the same as the jury treatment of Study 2.

In the veracity bond conditions, content creators were required to post a financial bond when publishing their stories. Participants in this condition were told: "When creating a post, content creators must attach a financial bond backing the accuracy of their content. If the post is successfully challenged, the bond is forfeited to the challenger." In the counter-veracity bond conditions, challengers were required to post a bond when submitting a dispute. Participants in this condition were provided the following explanation: "When challenging a post, users must attach a financial bond to their claim. If the challenge is successful, the bond is returned to the challenger. If unsuccessful, the bond is forfeited to the content creator." To see the connection between studies, Figure 2 lists the study treatments and Figure 3 provides a graphical overview of the relevant platform mechanisms being tested and the key measures being evaluated in each study.

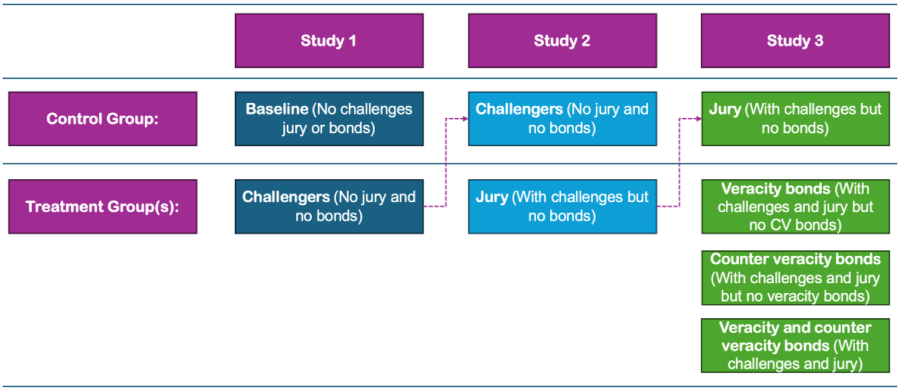

Figure 2. Overview of the treatments groups and their relationships across studies. Source: Authors' own creation

Downloaded from http://ftp.nowpublishers.com/imds/article-pdf/125/11/3024/10941121/imds-11-2024-1177en.pdf by guest on 24 June 2026

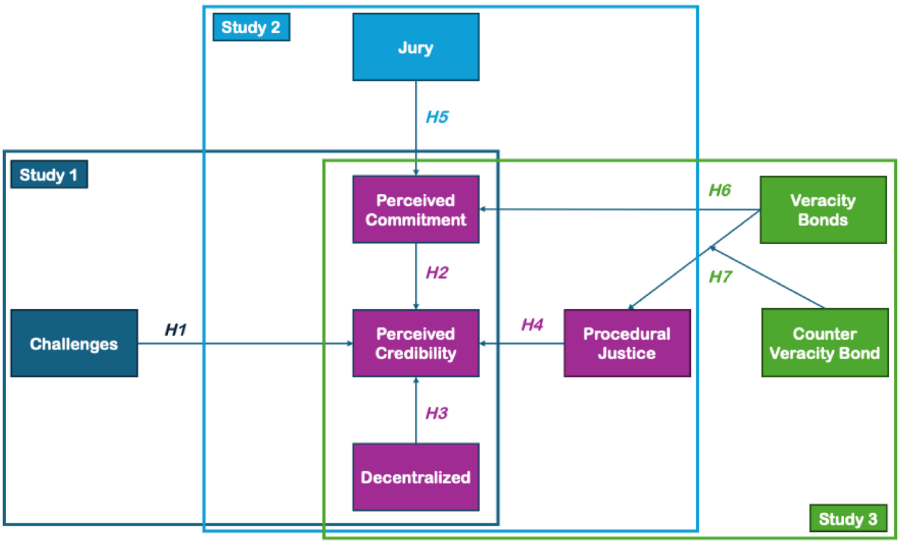

Figure 3. Overview of studies and research model. Source: Authors' own creation

Treatment salience validation. Our studies do not include manipulation checks due to their potential to disrupt the experimental process. As Hauser *et al.* (2018) caution, manipulation checks can introduce biases, particularly when participants in control groups are asked about interventions they have not experienced. For example, in Study 1, asking control group participants about the presence of a challenge mechanism could confuse them, lead to disengagement, or prompt demand characteristics that interfere with subsequent responses.

To ensure that the interventions were sufficiently salient without introducing these risks, we conducted a separate "treatment salience validation" experiment. This study ($N = 178$) replicated the verbatim text and conditions of the main experiments, randomizing participants across six treatments (Control, Challenger, Jury, Veracity Bond, Counter-Veracity Bond, and Both Bonds). As the texts were identical to the main experiments, participants were still required to pass the same attention checks before being exposed to the treatment information. After reading the treatment text, they answered four salience questions, indicating whether each mechanism (e.g. challenger, jury, veracity bond, counter-veracity bond) was present in the scenario. Responses were on a three-point scale: No (-1), Unsure (0), and Yes ($+1$). Including an "Unsure" option allowed participants to indicate genuine uncertainty about the presence of a mechanism, reducing the likelihood of forced or inaccurate responses. This approach minimized noise in the data by ensuring that participants who were unsure did not default to "Yes" or "No," which could have skewed the interpretation of salience. The full statistical results and analysis for this validation are provided in Appendix B.

Statistical comparisons using *t*-tests confirmed that each treatment group demonstrated significantly higher salience for their respective features compared to control and other unrelated groups ($p < 0.0001$). Specifically: (1) The Control group differed significantly from the Challenger group for the challenge mechanism question. (2) The Control and Challenger groups differed significantly from the Jury and Bond groups for the jury mechanism question. (3) The Veracity Bond and Both Bonds groups were significantly different from the Control, Challenger, Jury, and Counter-Veracity Bond groups for the veracity bond question. (4) The Counter-Veracity Bond and Both Bonds groups were significantly different from all other groups for the counter-veracity bond question.

Downloaded from http://ftp.nowpublishers.com/imds/article-pdf/125/11/3024/10941121/imds-11-2024-1177en.pdf by guest on 24 June 2026

Results

Study 1

To examine the impact of the Challenge mechanism and Decentralized Governance on user perceptions of platform credibility, we conducted a structural equation modeling (SEM) analysis using latent constructs for Credibility and Content Commitment.

Model Fit and Measurement Validity. Model fit was evaluated using multiple indices to assess the adequacy of the model. The Comparative Fit Index (CFI) was 0.954, and the Tucker-Lewis Index (TLI) was 0.934, which approach or exceed the conventional cut-off value (e.g. 0.95 for CFI), indicating an acceptable fit to the data. The standardized parameter estimates for factor loadings are summarized in Table C.1. The factor loadings for indicators of Content Commitment, Credibility, Challenge, and Decentralize were all significant, supporting the construct validity of these latent variables. For instance, the loadings for Credibility indicators ranged from 0.833 ($p < 0.001$) to 1.345 ($p < 0.001$), indicating that these items effectively represent the credibility construct. The loadings for Content Commitment indicators were also strong, with estimates ranging from 1.000 to 2.389 (all $p < 0.001$), supporting this construct's representation. Variance estimates for each latent variable are provided in Table C.2. Credibility exhibited a variance estimate of 0.223 (SE = 0.060, $p < 0.001$), while Content Commitment had a variance of 0.255 (SE = 0.123, $p = 0.039$). Challenge and Decentralize showed variance estimates of 0.248 (SE = 0.038, $p < 0.001$) and 0.235 (SE = 0.036, $p < 0.001$), respectively. These variance estimates in Table 1. Confirm the distinct contributions of each construct within the model, supporting the overall measurement structure.

Hypothesis Testing. Regression analysis was conducted to examine the effect of the Challenge mechanism, the Decentralized Governance and Content Commitment on the perceived credibility of the platform. The Challenge mechanism had a significant positive effect on perceived credibility, with an estimate of 0.539 (SE = 0.132, $z = 4.073$, $p < 0.001$), suggesting that the presence of a challenge feature enhances trust in platform content. This finding aligns with the hypothesis H1 that allowing users to contest information increases the perceived reliability of the platform. The estimate of 1.008 (SE = 0.274, $z = 3.685$, $p < 0.001$) provides strong support for H2 that greater Content Commitment increases the perceived Credibility of the platform. Finally, Decentralized Governance exhibited a marginally positive effect on credibility, with an estimate of 0.226 (SE = 0.129, $z = 1.751$, $p = 0.08$) providing mixed support for H3.

Study 2

To build on the findings from Study 1, Study 2 examined the impact of the Jury mechanism and Decentralized Governance on perceptions of platform credibility and content commitment. Participants were randomly assigned to conditions with and without a jury to adjudicate challenges, while both groups had access to the platform's challenge feature.

Model Fit and Measurement Validity. Model fit was evaluated using a variety of indices, all of which indicate an acceptable fit for the data. The CFI was 0.933, and the TLI was 0.914, both close to or exceeding the recommended threshold, indicating an adequate model fit. Factor

Table 1. Regression coefficients for Study 1

Outcome	Predictor	B	SE	*z*	*p*	95% CI Lower	Upper
Credibility	Content Commitment	1.008	0.274	3.685	<0.001	0.472	1.544
	Challenge	0.539	0.132	4.073	<0.001	0.280	0.798
	Decentralize	0.226	0.129	1.751	0.080	−0.027	0.478

Source(s): Authors' own creation

Downloaded from http://ftp.nowpublishers.com/imds/article-pdf/125/11/3024/10941121/imds-11-2024-1177en.pdf by guest on 24 June 2026

loadings for each latent variable (Table D.1) confirm that the indicators for Content Commitment, Credibility, and Procedural Justice are strongly related to their respective constructs. For example, the factor loadings for Credibility indicators ranged from 0.901 to 1.122 (all $p < 0.001$), supporting the construct's validity. Similarly, Content Commitment indicators ranged from 1.000 to 1.725, and Procedural Justice indicators ranged from 0.650 to 1.274, all highly significant ($p < 0.001$), confirming that these items effectively represent their intended constructs. Variance estimates for each factor (Table D.2), support the reliability of these constructs. Credibility had a variance estimate of 0.250 (SE = 0.060, $p < 0.001$), and Content Commitment had a variance of 0.508 (SE = 0.155, $p = 0.001$). Procedural Justice and the Jury variable both showed significant variance estimates as well, suggesting distinct and meaningful contributions to the model.

Hypothesis Testing. The regression results, presented in Table 2, provide insight into the hypotheses tested in Study 2. The presence of a jury mechanism did not significantly affect Content Commitment (estimate = 0.070, SE = 0.148, $z = 0.475$, $p = 0.635$), suggesting that adding a jury alone may not significantly enhance users' perceptions of authors' commitment to content quality. This finding does not support H5, which hypothesized that the Jury mechanism would positively influence Perceived Commitment. However, Content Commitment itself was a significant predictor of Credibility, with an estimate of 0.411 (SE = 0.118, $z = 3.472$, $p < 0.001$), supporting H2. Procedural Justice also had a significant effect on Credibility (estimate = 0.531, SE = 0.110, $z = 4.837$, $p < 0.001$), providing strong support for H4. Decentralized Governance showed a marginal effect on Credibility (estimate = 0.225, SE = 0.125, $z = 1.799$, $p = 0.072$), suggesting a potential trend where decentralized decision-making structures may contribute modestly to platform trustworthiness. As in Study 1, the result here again offers partial support for H3, indicating that decentralized governance may have a positive but limited effect on perceived credibility.

Study 3

Study 3 investigated the effects of veracity and counter-veracity bonds on perceptions of platform credibility, commitment, and procedural justice. Building on the features tested in previous studies, Study 3 introduced these financial mechanisms to examine their potential to enhance fairness perceptions and commitment. Participants were randomly assigned to four conditions: no bonds, veracity bond only, counter-veracity bond only, or both bonds.

Model Fit and Measurement Validity. The model demonstrated excellent fit, with a CFI of 0.977 and a TLI of 0.971, indicating a close fit to the data and supporting the adequacy of the model structure. Factor loadings for each latent variable (Table E.1) confirmed strong relationships between indicators and their respective constructs. For instance, Credibility indicators had loadings ranging from 0.744 to 1.017, all highly significant ($p < 0.001$), supporting the validity of this construct. Similarly, Procedural Justice indicators ranged from 0.672 to 1.066, demonstrating strong alignment with the intended construct. Variance estimates, presented in Table E.2, indicate the reliability of each construct, with significant variance across all factors. For example, Credibility had a variance estimate of 0.225

Table 2. Regression coefficients for Study 2

Outcome	Predictor	B	SE	*z*	*p*	95% CI Lower	95% CI Upper
Content commitment	Jury	0.070	0.148	0.475	0.635	−0.220	0.360
Credibility	Content commitment	0.411	0.118	3.472	<0.001	0.179	0.643
	Procedural justice	0.531	0.110	4.837	<0.001	0.316	0.746
	Decentralized	0.225	0.125	1.799	0.072	−0.020	0.470

Source(s): Authors' own creation

Downloaded from http://ftp.nowpublishers.com/imds/article-pdf/125/11/3024/10941121/imds-11-2024-1177en.pdf by guest on 24 June 2026

(SE = 0.041, $p < 0.001$), and Procedural Justice showed substantial variance at 1.385 (SE = 0.161, $p < 0.001$), confirming that these constructs were meaningfully captured in the model.

Hypothesis Testing. In Study 3, regression analysis as shown in Table 3 was conducted to examine the effects of veracity bonds, counter-veracity bonds, and their interaction on perceptions of Content Commitment, Procedural Justice and Content Credibility. We examined whether veracity bonds would positively impact Content Commitment. The results indicated a modest significant effect of veracity bonds on Content Commitment (estimate = 0.186, SE = 0.091, $z = 2.042$, $p = 0.041$), supporting H6. This suggests that the presence of veracity bonds enhances users' perceptions of content creators' commitment to quality, indicating that financial penalty may indeed signal increased accountability.

Requiring content creators to post a financial bond did not substantially impact participants' perceptions of fairness in platform processes (estimate = 0.007, SE = 0.228, $z = 0.031$, $p = 0.975$). On the other hand, counter-veracity bonds, which require challengers to post a bond when disputing content, demonstrated a significant negative effect on procedural justice (estimate = −0.652, SE = 0.212, $z = -3.073$, $p = 0.002$). This finding indicates that participants may view financial requirements for challengers as a barrier, potentially detracting from their sense of fairness in the challenge process. Additionally, the interaction effect between veracity and counter-veracity bonds, as proposed in H7, was marginally significant (estimate = 0.539, SE = 0.294, $z = 1.837$, $p = 0.066$), suggesting a possible combined impact on procedural justice. This suggests that the simultaneous presence of both bonds may lead to a more balanced perception of fairness, as compared to either bond alone. This partial support for H7 implies that using both veracity and counter-veracity bonds may create a mutually reinforcing effect, potentially enhancing procedural justice in a way that each bond alone does not achieve.

Revisiting other hypotheses tested in Study 3, the results provide continued support for H2 and H4. Content Commitment was again a significant predictor of Credibility (estimate = 0.510, SE = 0.113, $z = 4.501$, $p < 0.001$), aligning with H2 and reinforcing the notion that users view committed content creators as more credible. Procedural Justice also had a strong positive effect on Credibility (estimate = 0.593, SE = 0.054, $z = 10.947$, $p < 0.001$), supporting H4 and highlighting that fair and transparent processes enhance users' trust in the platform. In contrast, Decentralized Governance, hypothesized in H3 to positively influence Credibility, did not show a significant effect in this study (estimate = −0.071, SE = 0.088, $z = -0.810$, $p = 0.418$), suggesting that decentralization may have limited impact on credibility perceptions.

General discussion

Across all three studies, Content Commitment and Procedural Justice emerged as central elements in cultivating credibility and trust on the platform, which is the central objective of

Table 3. Regression coefficients for Study 3

Outcome	Predictor	B	SE	*z*	*p*	95% CI Lower	95% CI Upper
Content commitment	Veracity Bond	0.186	0.091	2.042	0.041	0.007	0.365
Credibility	Content Commit	0.510	0.113	4.501	<0.001	0.288	0.732
	Procedural Justice	0.593	0.054	10.947	<0.001	0.487	0.699
	Decentralized	−0.071	0.088	−0.810	0.418	−0.244	0.101
Procedural justice	Veracity Bond	0.007	0.228	0.031	0.975	−0.440	0.454
	Counter (CV) Bond	−0.652	0.212	−3.073	0.002	−1.068	−0.236
	Veracity × CV Bond	0.539	0.294	1.837	0.066	−0.036	1.115

Source(s): Authors' own creation

Downloaded from http://ftp.nowpublishers.com/imds/article-pdf/125/11/3024/10941121/imds-11-2024-1177en.pdf by guest on 24 June 2026

this research. Participants consistently valued a strong commitment to content quality from creators and fair processes within the platform, aligning with prior research that underscores the importance of perceived transparency and fairness in building trust online (Flanagin and Metzger, 2007; McKnight and Kacmar, 2007). When users see that content creators prioritize accuracy and the platform upholds fair standards, they are more inclined to perceive the platform itself as credible and reliable. These findings provide a foundation for understanding how commitment and procedural factors serve as trust-building pillars.

In Study 1, the Challenge mechanism provided users with an option to challenge content, which significantly improved credibility perceptions. This feature likely enhanced users' sense of agency and transparency, as it visibly connected them to the platform's quality-control processes. Unlike passive fact-checking labels, which often fail due to cognitive biases and entrenched beliefs, the Challenge mechanism gave users an active role in verifying information, which likely contributed to enhanced trust.

Study 2 extended this examination with the addition of a Jury mechanism, where a randomly selected group of users evaluated challenges. Interestingly, the jury's presence did not significantly affect Content Commitment, possibly reflecting limitations similar to those seen in real-world fact-checking and accuracy nudges (Szaszi *et al.*, 2022). Research has shown that fact-checking labels, while useful, often struggle to counteract psychological resistance and pre-existing biases (Lin *et al.*, 2016). Participants may have viewed the jury's decisions as too removed from creators' intentions, hence failing to establish a strong connection between the jury process and commitment to quality.

Study 3 introduced veracity and counter-veracity bonds as mechanisms to increase accountability. Veracity bonds, where content creators posted a financial bond to guarantee accuracy, significantly enhanced perceptions of Content Commitment, likely due to the concept of "skin in the game" (Taleb, 2018). This approach seems to resonate with users more than fact-checking or nudges (Szaszi *et al.*, 2022) because it creates tangible, visible accountability. On the other hand, counter-veracity bonds, where challengers posted a financial bond, had a negative impact on Procedural Justice. This suggests that requiring challengers to invest financially may introduce perceived barriers, reducing users' sense of fair access to participation. The combined use of veracity and counter-veracity bonds, however, showed potential in balancing fairness perceptions, indicating that a shared accountability mechanism could foster a sense of equity.

The role of Decentralized Governance across studies was modest, with marginal or non-significant effects on perceived credibility. While decentralization may be valuable in fostering transparency, the findings suggest that it does not contribute as strongly to credibility as other mechanisms. Decentralized structures alone may lack the direct, visible accountability signals that users associate with trustworthiness. This limited effect of decentralization indicates that governance structures may be less central to credibility perceptions than direct accountability and engagement features. Future studies could explore alternative governance models that combine decentralization with more transparent decision-making processes to enhance credibility perceptions.

Theoretical implications

Our study is the first to jointly examine the impact of challenges and juries on perceptions of content credibility across media platforms. By integrating Signaling Theory, the HSM, and Procedural Justice Theory, we theorize the significance of fairness perceptions and the commitment of content creators to truthful reporting as key mechanisms underlying the efficacy of these features. This theoretical integration provides a novel framework for understanding trust-building in decentralized platforms. Beyond testing the effectiveness of challenges and juries, we contribute to the academic literature by introducing the concepts of veracity bonds and counter-veracity bonds. By applying Signaling Theory to digital media contexts, we also extend the HSM through the introduction of a new psychological

Downloaded from http://ftp.nowpublishers.com/imds/article-pdf/125/11/3024/10941121/imds-11-2024-1177en.pdf by guest on 24 June 2026

mechanism: the Investment Heuristic. This mechanism bridges signaling theory and HSM, advancing mid-level theory and offering a more comprehensive perspective on trust-building processes (Chaiken, 1980; Tyler, 2000). Together, our findings demonstrate that costly signals, such as veracity bonds, act as cues that shape users' perceptions of commitment and fairness (Bird and Smith, 2005; Pennycook and Rand, 2019). We believe that there is tremendous value in future research exploring the applicability of veracity bonds and the Investment Heuristic in other domains, such as decentralized finance and governance.

The consistent importance of Content Commitment and Procedural Justice across all three studies reinforces theories emphasizing these factors as pillars of credibility in user-driven platforms. When users observe that content creators are committed to quality and that platform processes are fair, their trust in the platform is significantly strengthened. This aligns with theories in social psychology and media studies that emphasize the importance of perceived transparency and equity in fostering credibility. The findings suggest that commitment and procedural justice may function as core constructs that underlie trust in digital environments, particularly those reliant on user-generated content.

The mixed effectiveness of different mechanisms provides valuable insights into accountability frameworks in online media. The Challenge mechanism and veracity bonds emerged as effective in enhancing credibility, suggesting that users value direct accountability signals from both content creators and platform structures. Veracity bonds, in particular, embody the "skin in the game" concept from behavioral economics, which posits that visible financial stakes can signify commitment and encourage perceived accountability. The effectiveness of veracity bonds aligns with research that advocates for commitment mechanisms that extend beyond passive labels or nudges, signaling an alternative approach to credibility-building that could address the limitations of fact-checking and nudging interventions (Taleb, 2018).

The jury's role in evaluating content challenges exemplifies the concept of repeated "cheap talk," where jury members' ongoing discussions and judgments can create a shared expectation of integrity, even without formal binding (Aumann and Hart, 2003). Although decentralized and informal, the jury process signals fairness and community accountability, reinforcing perceptions of honesty and commitment to truthfulness (Bos *et al*., 1998). The jury acts as decentralized evaluator, reducing the potential for bias (e.g. based on wisdom of the crowd effects; Pennycook and Rand, 2019) and showing procedural fairness in a way that increases creators' commitment to accuracy. The platform can benefit from a community-driven mechanism, where creators are motivated to uphold standards due to the scrutiny they face from a collective body (Bergstrom and Lachmann, 1998). Thus, the use of a jury system may help mitigate some of these limitations by providing a structure where repeated interactions and shared interests enhance credibility. Yet, the limited impact of the Jury mechanism aligns with literature showing that passive nudges and fact-checking labels often fall short in reducing misinformation due to cognitive biases and social dynamics (Lin *et al*., 2016). This finding suggests that user-driven credibility mechanisms, while promising, may need to involve direct accountability rather than relying solely on peer evaluation to shift credibility perceptions. It raises questions about the effectiveness of purely evaluative features, such as juries, and points to the need for more proactive or "incentivized" engagement mechanisms.

The differing impacts of veracity and counter-veracity bonds underscore the complex relationship between accountability and procedural fairness. While veracity bonds positively impacted perceptions of commitment, counter-veracity bonds reduced perceptions of procedural justice, suggesting that financial barriers for challengers might discourage engagement and create perceived inequities. These findings expand theories of procedural justice by indicating that accountability measures, to be effective, must balance responsibility and accessibility. Imposing financial stakes on challengers, while intended to deter frivolous challenges, may have been perceived as an unfair hurdle that conflicts with user expectations of open engagement. This nuanced insight suggests that while accountability mechanisms are

Downloaded from http://ftp.nowpublishers.com/imds/article-pdf/125/11/3024/10941121/imds-11-2024-1177en.pdf by guest on 24 June 2026

valuable, they must be designed to ensure equitable access, particularly for user-driven content platforms.

The relatively modest influence of Decentralized Governance on credibility across studies contributes to theories on governance in digital media by suggesting that decentralization alone may not strongly impact credibility perceptions. The findings imply that governance structures, while important, may require additional transparency or direct user accountability features to build trust effectively. This insight aligns with literature on digital governance that emphasizes the role of visible accountability over structural decentralization in fostering user trust. Future research could investigate hybrid governance models that combine decentralization with direct user engagement features, potentially enhancing the trustworthiness of platforms reliant on collective decision-making.

The results stress the need for theoretical frameworks that incorporate both visible commitment signals and procedural justice into digital trust-building strategies. This research suggests that an integrated approach, combining features like challenge options and financial accountability, may be more effective than standalone mechanisms. As such, this study contributes to an emerging theoretical understanding of "integrated accountability," which highlights that user perceptions of credibility may be best supported when multiple, balanced mechanisms are present.

Practical implications

The results from this research offer several practical insights for digital media platforms seeking to enhance user trust, encourage engagement, and foster fair and transparent processes. By examining design and operational mechanisms like challenges, juries, financial bonds, and decentralized governance, this study highlights specific design strategies that can strengthen perceived credibility and procedural justice.

Several media platforms have already experimented with mechanisms that align with the proposed challenge, jury, and veracity bond frameworks, demonstrating their feasibility in real-world applications. As discussed, Twitter/X's Community Notes represents a notable example where users to collaboratively challenge and fact-check posts by adding context to potentially misleading information. Research on this feature has shown that user-generated annotations can increase trust in fact-checking efforts and reduce the spread of misinformation. The success of Community Notes highlights the effectiveness of participatory challenge mechanisms, similar to the one proposed in this study.

In addition to participatory verification, decentralized adjudication systems have been explored by platforms like Minds, a social network that incorporates a jury-based moderation system. On Minds, content moderation decisions are reviewed by randomly selected jurors from the platform's user base, ensuring that decisions are made collectively rather than by a central authority. This approach parallels the jury mechanism proposed in this research, reinforcing the argument that decentralized dispute resolution can enhance perceptions of fairness and procedural justice.

Financial commitment mechanisms such as veracity bonds have yet to be adopted in mainstream media platforms. However, their underlying principles can be seen in blockchain-based platforms like Steemit and Brave Rewards. Steemit, a blockchain-based social media platform, uses financial incentives to reward high-quality content, aligning creator incentives with content credibility. Similarly, Brave allows users to support trustworthy content creators through microtransactions, fostering an ecosystem where financial stakes are linked to content quality. These examples demonstrate how economic incentives can be integrated into media environments to promote accountability, similar to the veracity bond mechanism proposed in this study.

Although no platform fully implements all aspects of our proposed trust mechanisms, our results suggest that participatory verification, decentralized governance, and financial accountability can be effective tools in addressing media trust issues. Future media platforms

Downloaded from http://ftp.nowpublishers.com/imds/article-pdf/125/11/3024/10941121/imds-11-2024-1177en.pdf by guest on 24 June 2026

can build on these existing models by integrating them into a unified system that maximizes transparency, procedural fairness, and incentive alignment. By leveraging these strategies, digital media environments can move closer to a model where users have greater control over content credibility, fostering a more trustworthy and resilient information ecosystem.

Given the strong influence of Content Commitment and Procedural Justice on credibility perceptions, platform designers should emphasize these factors in the user experience. To boost content commitment, platforms can implement visible indicators of creator accountability, such as commitment labels or transparency disclosures. Additionally, clear and consistent guidelines for procedural justice, such as fair dispute processes and transparent content policies, will likely increase user trust. Platforms that visibly prioritize quality and fairness in their processes can foster a more credible environment that enhances user retention and engagement.

The challenge mechanism and veracity bonds emerged as effective tools for boosting credibility perceptions, suggesting that users appreciate direct accountability signals. Real-world platforms already demonstrate aspects of these mechanisms. For example, Twitter (now X) uses Community Notes to enable users to collaboratively verify the accuracy of content, a feature that parallels the challenge mechanism (Moravec *et al*., 2023). Similarly, Brave employs an attention-based token system to reward transparent and ethical engagement, echoing the principles behind veracity bonds (Serada *et al*., 2022). These examples, while not the exact implementations as described our experiment, have nevertheless demonstrated the capabilities of the blockchain technology and its smart contract implementations that allow for decentralized features that could help to enhance accountability and credibility in a social media platform.

By ensuring this feature is accessible and straightforward to use, platforms can promote transparency and encourage users to actively participate in maintaining content quality. Veracity bonds represent another actionable mechanism that platforms could adopt, especially in contexts where content accuracy is paramount. Requiring content creators to post a financial commitment could be particularly valuable in specialized areas, such as news platforms, where accuracy is a core value. Platforms might consider offering a graded bond system, where the amount or type of commitment depends on content reach or impact, to allow creators flexibility while still signaling accountability. Veracity bonds can offer users an assurance that content creators have "skin in the game," helping to counteract psychological resistance to fact-checking and providing a stronger credibility signal than passive labels or warnings.

The mixed findings regarding counter-veracity bonds offer important lessons on balancing accountability with procedural fairness. While veracity bonds for content creators can enhance credibility, requiring challengers to post bonds may deter legitimate participation if perceived as a barrier. Platforms should avoid imposing financial requirements on challengers unless they are carefully designed to be accessible and fair. For example, platforms could consider alternatives such as refundable deposits, scaled bond amounts based on user history, or incentives for accurate challenges to mitigate any negative impact on fairness perceptions. By ensuring that accountability mechanisms do not discourage legitimate engagement, platforms can better balance user participation with quality control.

While decentralized governance has potential benefits, its relatively limited impact on credibility in this study suggests that platforms may not be able to rely on decentralization alone to foster trust. In addition, immutability can introduce its own set of challenges in the perpetually evolving information ecosystem of news and media. As more information is revealed certain attestations become true and others false. Thus, off-chain techniques could be employed on top of immutable records to encode these dynamics and hybrid governance models that combine decentralization with direct user accountability features (i.e. challenges and veracity bonds) may be more effective in building credibility. Platforms could also enhance decentralized governance by providing users with transparent insights into decision-making processes, which may help offset the lack of direct accountability signals inherent in purely decentralized structures.

Downloaded from http://ftp.nowpublishers.com/imds/article-pdf/125/11/3024/10941121/imds-11-2024-1177en.pdf by guest on 24 June 2026

The findings suggest that no single mechanism is sufficient for fostering credibility; instead, an integrated approach that combines multiple, mutually reinforcing features may be most effective. Platforms like Civil, Brave, and Steemit illustrate how various mechanisms, such as decentralized governance, financial incentives, and user-driven verification, can be operationalized to promote accountability and trust (Al-Saqaf and Edwardsson, 2019; Liu *et al*., 2022; Serada *et al*., 2022). These examples provide a roadmap for how the proposed features can be adapted to real-world contexts, enhancing the applicability of this study's findings.

Platforms could create a trust-building ecosystem by combining challenges, veracity bonds, and fairness-promoting policies, providing users with a holistic framework that balances commitment, transparency, and procedural justice. Implementing a combination of these features, rather than relying on one in isolation, could significantly improve user perceptions of platform trustworthiness and contribute to a more resilient and credible online environment.

The limited effectiveness of the Jury mechanism parallels real-world challenges with fact-checking and accuracy nudges, which often have limited success in curbing misinformation. Platforms could therefore consider alternatives to passive fact-checking labels by adopting more proactive measures, such as veracity bonds, which may better address user demands for accountability. Given the persistence of misinformation and the psychological barriers to fact-checking, providing content creators with incentives to ensure accuracy, such as veracity bonds or graded trust scores, may enhance the platform's credibility more effectively than conventional fact-checking tags or nudges.

Limitations and future research

This study has limitations that open avenues for future research. Our experiments utilized a hypothetical platform without implementing live blockchain technology. While this study provides valuable insights into mechanisms that can enhance trust in media, it does not implement blockchain technology in a live experimental setting. Instead, the research focuses on evaluating the conceptual effectiveness of decentralized governance principles, veracity bonds, and challenge mechanisms through scenario-based experiments. This approach allowed us to assess user perceptions without the technical complexities of an operational blockchain platform. Overall, this study is positioned as a proof-of-concept to demonstrate how these mechanisms could function in a decentralized governance model. However, the absence of a fully implemented blockchain system remains a limitation, as it prevents direct evaluation of the practical feasibility and real-world adoption of these mechanisms. Future research may build on these findings by developing operational experimental decentralized systems (Tapscott and Tapscott, 2016; Nakamoto, 2008) to validate the mechanisms' effectiveness under real-world conditions.

Blockchain technology offers unique advantages in trust-building, particularly through its ability to provide immutable, transparent, and decentralized record-keeping. The smart contract functionalities that would enable automatic enforcement of veracity bonds and challenge outcomes were theorized in this study but not tested in a functioning blockchain environment. While our findings support the conceptual promise of blockchain-based trust mechanisms, further research is necessary to validate their effectiveness in practice. Future research should address this gap by conducting experiments with an actual blockchain implementation. Developing a working prototype or deploying smart contract-based veracity bonds within a decentralized media platform would allow researchers to examine how these mechanisms function in real user interactions. Testing such a system in a live setting could provide empirical evidence on transaction costs, user adoption challenges, and the potential risks of manipulation or strategic behavior in blockchain-based governance.

Additionally, while the study used randomized conditions, further investigation is needed to assess the long-term effects of these mechanisms, particularly in active social media

Downloaded from http://ftp.nowpublishers.com/imds/article-pdf/125/11/3024/10941121/imds-11-2024-1177en.pdf by guest on 24 June 2026

environments. Future research could explore hybrid governance models that integrate decentralized structures with clear accountability mechanisms, examining whether these models enhance credibility more effectively. Moreover, investigating the effects of veracity and counter-veracity bonds on a larger, more diverse sample could provide deeper insights into their impact on fairness perceptions across various demographic contexts. Additionally, an operational blockchain system would enable a deeper investigation into how decentralized verification and financial stakes impact long-term trust in media platforms.

Despite these limitations, our study lays the groundwork for future exploration by providing a theoretical and experimental basis for blockchain-integrated media trust mechanisms. The findings highlight the potential of decentralized governance, participatory challenge systems, and financial accountability tools as promising solutions. Future studies that implement these features in a fully functional blockchain environment will be crucial in advancing the practical application of these concepts and validating their real-world impact.

Notes

1. https://apnorc.org/wp-content/uploads/2021/10/misinformation_Formatted_v2-002.pdf
2. https://news.gallup.com/poll/651977/americans-trust-media-remains-trend-low.aspx
3. In January 2025, Meta announced the replacement of its third-party fact-checking programs with a system inspired by X's Community Notes (https://www.wired.com/story/meta-ditches-fact-checkers-in-favor-of-x-style-community-notes/).
4. https://www.minds.com/minds/blog/power-to-the-people-the-minds-jury-system-975486713993859072
5. https://www.binance.com/en/square/post/17875447050538
6. Cheap talk refers to communication that is non-binding and cost-free, meaning it does not directly affect the payoffs of the involved parties (Farrell, 1995). Despite its lack of direct costs, cheap talk can still be informative particularly when there is repeated interaction or aligned interests between the parties. Although traditional signaling theory suggests that cheap talk should be uninformative, research shows that it can influence outcomes, especially in repeated interactions where trust can be built over time (Aumann and Hart, 2003; Manelli, 1996).
7. To ensure participant engagement, attention checks were embedded throughout all three studies. For instance, participants were asked to recall the type of content posted on MediaSphere. If participants provided incorrect answers, they were given a warning, and those who failed two attention checks were removed from the study.

References

Al-Saqaf, W. and Edwardsson, M.P. (2019), "Could blockchain save journalism? An explorative study of blockchain's potential to make journalism a more sustainable business", in *Blockchain and Web 3.0*, Routledge, pp. 97-113, doi: 10.4324/9780429029530-7.

Ardèvol-Abreu, A. and Gil de Zúñiga, H. (2017), "Effects of editorial media bias perception and media trust on the use of traditional, citizen, and social media news", *Journalism and Mass Communication Quarterly*, Vol. 94 No. 3, pp. 703-724, doi: 10.1177/1077699016654684.

Aslett, K., Sanderson, Z., Godel, W., Persily, N., Nagler, J. and Tucker, J.A. (2023), "Online searches to evaluate misinformation can increase its perceived veracity", *Nature*, Vol. 620 No. 7974, pp. 542-546, doi: 10.1038/s41586-023-06883-y.

Aumann, R.J. and Hart, S. (2003), "Long cheap talk", *Econometrica*, Vol. 71 No. 6, pp. 1619-1660, doi: 10.1111/1468-0262.00465.

Bergstrom, C.T. and Lachmann, M. (1998), "Signaling among relatives. III. Talk is cheap", *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 95 No. 9, pp. 5100-5105, doi: 10.1073/PNAS.95.9.5100.

Downloaded from http://ftp.nowpublishers.com/imds/article-pdf/125/11/3024/10941121/imds-11-2024-1177en.pdf by guest on 24 June 2026

Bharti, N. (2020), "Controlling the coronavirus narrative", *Science*, Vol. 369 No. 6503, p. 1189, doi: 10.1126/SCIENCE.ABD3662.

Bianchi, E.C., Brockner, J., Bos van den, K., Seifert, M., Moon, H., Dijke van, M. and Cremer, D.D. (2015), "Trust in decision-making authorities dictates the form of the interactive relationship between outcome fairness and procedural fairness", *Personality and Social Psychology Bulletin*, Vol. 41 No. 1, pp. 19-34, doi: 10.1177/0146167214556237.

Bird, R.B. and Smith, E.A. (2005), "Signaling theory, strategic interaction, and symbolic capital", *Current Anthropology*, Vol. 46 No. 2, pp. 221-248, doi: 10.1086/427115.

Bos, K.van den, Wilke, H.A.M. and Lind, E.A. (1998), "When do we need procedural fairness? The role of trust in authority", *Journal of Personality and Social Psychology*, Vol. 75 No. 6, pp. 1449-1458, doi: 10.1037/0022-3514.75.6.1449.

Chaiken, S. (1980), "Heuristic versus systematic information processing and the use of source versus message cues in persuasion", *Journal of Personality and Social Psychology*, Vol. 39 No. 5, pp. 752-766, doi: 10.1037/0022-3514.39.5.752.

Chaiken, S. and Maheswaran, D. (1994), "Heuristic processing can bias systematic processing: effects of source credibility, argument ambiguity, and task importance on attitude judgment", *Journal of Personality and Social Psychology*, Vol. 66 No. 3, pp. 460-473, doi: 10.1037/0022-3514.66.3.460.

Chauhan, S. and Gupta, P. (2024), "Assessing credibility in eWOM: a meta-analysis using the heuristic-systematic model", *Journal of Enterprise Information Management*, Vol. 37 No. 6, pp. 1839-1857, doi: 10.1108/JEIM-01-2024-0027.

Colquitt, J.A. (2001), "On the dimensionality of organizational justice: a construct validation of a measure", *Journal of Applied Psychology*, Vol. 86 No. 3, pp. 386-400, doi: 10.1037/0021-9010.86.3.386.

Davis, J.M. and Tuttle, B.M. (2013), "A heuristic–systematic model of end-user information processing when encountering IS exceptions", *Information and Management*, Vol. 50 Nos 2-3, pp. 125-133, doi: 10.1016/j.im.2012.09.004.

Donkers, T. and Ziegler, J. (2021), "The dual echo chamber: modeling social media polarization for interventional recommending", *Proceedings of the 15th ACM conference on recommender systems*, pp. 12-22, doi: 10.1145/3460231.3474261.

Drolsbach, C.P., Solovev, K. and Pröllochs, N. (2024), "Community notes increase trust in fact-checking on social media", *PNAS Nexus*, Vol. 3 No. 7, p. 217, doi: 10.1093/pnasnexus/pgae217.

Farrell, J. (1995), "Talk is cheap", *The American Economic Review*, Vol. 85 No. 2, pp. 186-190.

Flanagin, A.J. and Metzger, M.J. (2007), "The role of site features, user attributes, and information verification behaviors on the perceived credibility of web-based information", *New Media and Society*, Vol. 9 No. 2, pp. 319-342, doi: 10.1177/1461444807075015.

Fletcher, R., Andı, S., Badrinathan, S., Eddy, K.A., Kalogeropoulos, A., Mont'Alverne, C., Robertson, C.T., Ross Arguedas, A., Schulz, A., Toff, B. and Nielsen, R.K. (2025), "The link between changing news use and trust: longitudinal analysis of 46 countries", *Journal of Communication*, Vol. 75 No. 1, pp. 1-15, doi: 10.1093/joc/jqae044.

Frazier, M.L., Johnson, P.D., Gavin, M.B., Gooty, J. and Snow, D.B. (2010), "Organizational justice, trustworthiness, and trust: a multifoci examination", *Group and Organization Management*, Vol. 35 No. 1, pp. 39-76, doi: 10.1177/1059601109354801.

Globig, L. and Sharot, T. (2024), "Considering information-sharing motives to reduce misinformation", *Current Opinion in Psychology*, Vol. 50, 101852, doi: 10.1016/j.copsyc.2024.101852.

Hauser, D.J., Ellsworth, P.C. and Gonzalez, R. (2018), "Are manipulation checks necessary?", *Frontiers in Psychology*, Vol. 9, p. 998, doi: 10.3389/fpsyg.2018.00998.

Hoskinson, C. (2023), available at: https://cryptoslate.com/hoskinson-reacts-to-push-back-against-coindesk-acquistion (accessed 16 October 2024).

Johnston, J.S. (1999), "Communication and courtship: cheap talk economics and the law of contract formation", *Virginia Law Review*, Vol. 85 No. 3, p. 385, doi: 10.2307/1073700.

Downloaded from http://ftp.nowpublishers.com/imds/article-pdf/125/11/3024/10941121/imds-11-2024-1177en.pdf by guest on 24 June 2026

Kunda, Z. (1990), "The case for motivated reasoning", *Psychological Bulletin*, Vol. 108 No. 3, pp. 480-498, doi: 10.1037/0033-2909.108.3.480.

Leventhal, G.S. (1976), "Fairness in social relationships", in Thibaut, J.W., Spence, J.T. and Carson, R.C. (Eds), *Contemporary Topics in Social Psychology*, General Learning Press, pp. 211-239.

Leventhal, G.S. (1980), "What should be done with equity theory? New approaches to the study of fairness in social relationships", in Gergen, K.J., Greenberg, M.S. and Willis, R.H. (Eds), *Social Exchange: Advances in Theory and Research*, Springer, pp. 27-55.

Lin, X., Spence, P.R. and Lachlan, K.A. (2016), "Social media and credibility indicators: the effect of influence cues", *Computers in Human Behavior*, Vol. 63, pp. 264-271.

Linden, S.van der, Maibach, E., Cook, J., Leiserowitz, A. and Lewandowsky, S. (2017), "Inoculating against misinformation", *Science*, Vol. 358 No. 6367, pp. 1141-1142, doi: 10.1126/SCIENCE.AAR4533.

Liu, Z., Li, Y., Min, Q. and Chang, M. (2022), "User incentive mechanism in blockchain-based online community: an empirical study of Steemit", *Information and Management*, Vol. 59 No. 7, 103596, doi: 10.1016/j.im.2022.103596.

Manelli, A.M. (1996), "Cheap talk and sequential equilibria in signaling games", *Econometrica*, Vol. 64 No. 4, p. 917, doi: 10.2307/2171850.

McKnight, D.H. and Kacmar, C. (2007), "Factors and effects of information credibility", *Proceedings of the 43rd Hawaii International Conference on System Sciences (HICSS-43)*, pp. 1-10.

Metzger, M.J. and Flanagin, A.J. (2013), "Credibility and trust of information in online environments: the use of cognitive heuristics", *Journal of Pragmatics*, Vol. 59 B, pp. 210-220, doi: 10.1016/j.pragma.2013.07.012.

Meyer, J.P. and Allen, N.J. (1991), "A three-component conceptualization of organizational commitment", *Human Resource Management Review*, Vol. 1 No. 1, pp. 61-89, doi: 10.1016/1053-4822(91)90011-z.

Moravec, P.L., Minas, R.K. and Dennis, A.R. (2019), "Fake news on social media: people believe what they want to believe when it makes no sense at all", *Management Information Systems Quarterly*, Vol. 43 No. 4, pp. 1341-1370, doi: 10.25300/misq/2019/15505.

Moravec, P.L., Kim, A. and Dennis, A.R. (2020), "Appealing to sense and sensibility: system 1 and system 2 interventions for fake news on social media", *Information Systems Research*, Vol. 31 No. 3, pp. 987-1006, doi: 10.1287/ISRE.2020.0927.

Moravec, P.L., Kim, A., Dennis, A.R. and Minas, R.K. (2022), "Do you really know if it's true? How asking users to rate stories affects belief in fake news on social media", *Information Systems Research*, Vol. 33 No. 3, pp. 887-907, doi: 10.1287/isre.2021.1090.

Moravec, P.L., Collis, A. and Wolczynski, N. (2023), "Countering state-controlled media propaganda through labeling: evidence from Facebook", *Information Systems Research*, Vol. 34 No. 2, pp. 1041-1059, doi: 10.1287/isre.2022.0305.

Morosoli, S., Resendez, V., Naudts, L., Helberger, N. and de Vreese, C. (2024), "'I resist'. A study of individual attitudes towards generative AI in journalism and acts of resistance, risk perceptions, trust and credibility", *Digital Journalism*, pp. 1-20, doi: 10.1080/21670811.2024.2435579.

Nakamoto, S. (2008), "Bitcoin: a peer-to-peer electronic cash system".

Noble, J. (1999), "Cooperation, conflict and the evolution of communication", *Adaptive behavior*, Vol. 7 No. 3-4, pp. 349-369, doi: 10.1177/105971239900700308.

Peng, Y. and Yang, T. (2022), "Anatomy of audience duplication networks: how individual characteristics differentially contribute to fragmentation in news consumption and trust", *New Media and Society*, Vol. 24 No. 10, pp. 2270-2290, doi: 10.1177/1461444821991559.

Pennycook, G. and Rand, D.G. (2019), "Fighting misinformation on social media using crowdsourced judgments of news source quality", *Proceedings of the National Academy of Sciences*, Vol. 116 No. 7, pp. 2521-2526, doi: 10.1073/pnas.1806781116.

Downloaded from http://ftp.nowpublishers.com/imds/article-pdf/125/11/3024/10941121/imds-11-2024-1177en.pdf by guest on 24 June 2026

Pennycook, G., Bear, A., Collins, E.T. and Rand, D.G. (2020), "The implied truth effect: attaching warnings to a subset of fake news headlines increases perceived accuracy of headlines without warnings", *Management Science*, Vol. 66 No. 11, pp. 4944-4957, doi: 10.1287/MNSC.2019.3478.

Serada, A., Grym, J. and Sihvonen, T. (2022), "The economy of attention on blockchain in the brave browser", in *Futures of Journalism: Technology-Stimulated Evolution in the audience-news Media Relationship*, Springer International Publishing, Cham, pp. 49-62, doi: 10.1007/978-3-030-95073-6_4.

Silva, H.D. and Sigmund, K. (2024), "Dynamics of signaling games", *Siam Review*, Vol. 66 No. 2, pp. 368-387, doi: 10.1137/23m156402x.

Skinner, B.F. (1996), "Are theories of learning necessary?", *Psychological Review*, Vol. 57 No. 4, pp. 193-216, doi: 10.1037/h0054367.

Spanò, R., Massaro, M., Ferri, L., Dumay, J.L. and Schmitz, J. (2022), "Blockchain in accounting, accountability and assurance: an overview", *Accounting, Auditing and Accountability*, Vol. 35 No. 9, pp. 2473-2494, doi: 10.1108/aaaj-06-2022-5850.

Suiter, J. and Fletcher, R. (2020), "Polarization and partisanship: key drivers of distrust in media old and new?", *European Journal of Communication*, Vol. 35 No. 5, pp. 484-501, doi: 10.1177/0267323120903685.

Sundar, S.S. (2008), "The MAIN model: a heuristic approach to understanding technology effects on credibility", in Miriam, J.M. and Andrew, J.F. (Eds), *Digital Media, Youth, and Credibility*, The John D. and Catherine T. MacArthur Foundation Series on Digital Media and Learning, The MIT Press, Cambridge, MA, pp. 73-100.

Sun, Y. and Xie, J. (2024), "Who shares misinformation on social media? A meta-analysis of individual traits related to misinformation sharing", *Computers in Human Behavior*, Vol. 158, 108271, doi: 10.1016/j.chb.2024.108271.

Sweller, J. (1988), "Cognitive load during problem solving: effects on learning", *Cognitive Science*, Vol. 12 No. 2, pp. 257-285, doi: 10.1207/s15516709cog1202_4.

Szaszi, B., Higney, A., Charlton, A.B., Gelman, A., Ziano, I., Aczel, B., Goldstein, D.G., Yeager, D.S. and Tipton, E. (2022), "No reason to expect large and consistent effects of nudge interventions", *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 119 No. 31, e2200732119, doi: 10.1073/pnas.2200732119.

Taleb, N.N. (2018), *Skin in the Game: Hidden Asymmetries in Daily Life*, Random House.

Tapscott, D. and Tapscott, A. (2016), *Blockchain Revolution: How the Technology Behind Bitcoin is Changing Money, Business, and the World*, Penguin.

Tyler, T.R. (2000), "Social justice: outcome and procedure", *International Journal of Psychology*, Vol. 35 No. 2, pp. 117-125, doi: 10.1080/002075900399411.

Tyma, B., Dhillon, R., Sivabalan, P. and Wieder, B. (2022), "Understanding accountability in blockchain systems", *Accounting, Auditing and Accountability*, Vol. 35 No. 9, pp. 2553-2574, doi: 10.1108/aaaj-07-2020-4713.

Vu, H.T. and Chen, Y. (2023), "What influences audience susceptibility to fake health news: an experimental study using a dual model of information processing in credibility assessment", *Health Communication*, Vol. 38 No. 5, pp. 601-613, doi: 10.1080/10410236.2023.2206177.

Zajonc, R.B. (1980), "Feeling and thinking: preferences need no inferences", *American Psychologist*, Vol. 35 No. 2, pp. 151-175, doi: 10.1037/0003-066X.35.2.151.

Further reading

Buterin, V. (2013), *Ethereum: A Next-generation Smart Contract and Decentralized Application Platform*, Ethereum Foundation.

Catalini, C. and Gans, J.S. (2020), "Some simple economics of the blockchain", *Communications of the ACM*, Vol. 63 No. 7, pp. 79-87, doi: 10.1145/3359552.

Downloaded from http://ftp.nowpublishers.com/imds/article-pdf/125/11/3024/10941121/imds-11-2024-1177en.pdf by guest on 24 June 2026

Główczewski, M. and Burdziej, S. (2022), "(In)justice in academia: procedural fairness, students' academic identification, and perceived legitimacy of university authorities", *International Journal of Higher Education*, Vol. 86 No. 1, pp. 163-184, doi: 10.1007/s10734-022-00907-8.

Guess, A.M., Barberá, P., Munzert, S. and Yang, J. (2021), "The consequences of online partisan media", *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 118 No. 14, e2013464118, doi: 10.1073/PNAS.2013464118.

Gupta, V., Han, B.R., Kim, S.-H. and Paek, H. (2020), "Maximizing intervention effectiveness", *Management Science*, Vol. 66 No. 11, pp. 5246-5257, doi: 10.1287/MNSC.2019.3537.

Iyengar, S. and Massey, D.S. (2019), "Scientific communication in a post-truth society", *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 116 No. 16, pp. 7656-7661, doi: 10.1073/PNAS.1805868115.

Kuruppu, S., Dissanayake, D. and Villiers, C.d. (2022), "How can NGO accountability practices be improved with technologies such as blockchain and triple-entry accounting?", *Accounting, Auditing and Accountability*, Vol. 35 No. 9, pp. 2689-2713, doi: 10.1108/aaaj-10-2020-4972.

Lazer, D., Baum, M., Benkler, Y., Berinsky, A., Greenhill, K., Menczer, F., Metzger, M., Nyhan, B., Pennycook, G., Rothschild, D., Schudson, M., Sloman, S.A., Sunstein, C.R., Thorson, E.A., Watts, D.J. and Zittrain, J.L. (2018), "The science of fake news", *Science*, Vol. 359 No. 6380, pp. 1094-1096, doi: 10.1126/science.aao2998.

Nikidehaghani, M., Andrew, J. and Cortese, C.L. (2022), "Algorithmic accountability: robodebt and the making of welfare cheats", *Accounting, Auditing and Accountability*, Vol. 35 No. 5, pp. 1125-1149, doi: 10.1108/aaaj-02-2022-5666.

Supplementary material
The supplementary material for this article can be found online.

Corresponding author
Eric Tze Kuan Lim can be contacted at: e.t.lim@unsw.edu.au

For instructions on how to order reprints of this article, please visit our website:
www.emeraldgrouppublishing.com/licensing/reprints.htm
Or contact us for further details: **permissions@emeraldinsight.com**

Downloaded from http://ftp.nowpublishers.com/imds/article-pdf/125/11/3024/10941121/imds-11-2024-1177en.pdf by guest on 24 June 2026