Communications of the Association for Information Systems

Special Issue DOI: 10.17705/1CAIS.05441 ISSN: 1529-3181

Designing Incentives for Attracting Peer Reviewers to Information System Conferences

Arthur Carvalho

Farmer School of Business
Miami University
USA
arthur.carvalho@miamioh.edu
0000-0002-5381-3588

Chad Anderson

Farmer School of Business
Miami University
USA
0000-0002-7163-5578

Liudmila Zavolokina

University of Zurich
Switzerland
University of Lausanne
Switzerland
0000-0002-0762-3505

Abstract:

Information systems (IS) conferences, as venues for the introduction of new knowledge to the IS community, require effective peer review systems to evaluate submitted research for quality, validity, and originality. We argue in this paper that questionable practices and degrading review quality may arise without direct incentives beyond reviewer altruism to engage in the peer review process. In particular, we highlight potential issues with arguably common practices in some IS conferences, such as peer review invitations sent to researchers who have also submitted papers for publication consideration and the increasing number of reviews performed by graduate students. To address these issues, we suggest three solutions: 1) quid pro quo rules; 2) the use of incentive-compatible methods whose scores are linked to relevant rewards; and 3) the use of blockchain-based tokens in tandem with smart contracts and zero-knowledge proofs. We conclude by offering directions the IS community can take to further study the highlighted issues and implement the proposed solutions.

Keywords: Blockchain, Incentive Engineering, Peer Review, Tokenization, Zero-Knowledge Proofs.

This manuscript underwent peer review. It was received 4/03/2023 and was with the authors for four months for one revision. Cathy Urquhart served as Associate Editor.

1 Introduction

Information systems (IS) conferences are vehicles to introduce new knowledge to the IS community. In order to maintain high-quality standards in the knowledge disseminated, IS conferences require effective peer review systems. Peer review is a process by which experts in a field evaluate and critique scholarly papers. The goal is to assess those papers in terms of their quality, validity, originality, and fit within the scope of the underlying publication outlet (e.g., conferences, journals). Such experts are commonly referred to as peers, hence the term "peer review." Although the specifics vary from outlet to outlet, the peer review process typically involves four steps. First, authors submit their work for publication consideration in a specific outlet. Second, the editors of that outlet screen each submission and invite qualified reviewers to evaluate the work's quality and suitability for publication. Third, the reviewers read the manuscript, provide feedback to the authors, and recommend to the editors whether the work should be accepted for publication, rejected, or if revisions are required. Finally, the editor makes a decision and communicates it to the authors.

Peer review is a critical part of the scientific communication process as it helps to ensure that the work published in academic outlets meets the community standards of accuracy, quality, and significance. It also provides a means for authors to receive feedback and improve their work before it is eventually published. While peer review can greatly enhance the quality of published research, it also suffers from a number of limitations that include reviewer bias (Frachtenberg & McConville, 2022; Lee et al., 2012), conflicts of interest (Balietti et al., 2016; Anderson et al., 2007), variable review quality (Gardner et al., 2012), lack of qualified reviewers (Stafford, 2018; Urquhart et al., 2017) lack of transparency (Smith, 2006), and limited diversity (Eve et al., 2021). Addressing these issues requires meaningful effort by research communities, including better training, support, and incentives for reviewers, potentially necessitating the development of new peer review models. Below, we discuss specific challenges for peer review in IS conferences. We then suggest potential ideas for the IS community to contemplate in order to address those issues and improve the peer review process in our conferences.

2 Challenges in IS Conference Reviewing

Little has been published specifically on the peer review process in IS conferences. One notable exception was a report by the 2015 International Conference on Information Systems (ICIS) program chairs (Urquhart et al., 2017). That report provides an account of the challenges of chairing the ICIS conference with recommendations for addressing those challenges. One of the challenges described by the authors was finding enough volunteers to review the conference's 1,198 submissions. Urquhart et al. (2017) noted that the difficulty in recruiting volunteers spanned the track chair, associate editor (AE), and reviewer roles. They theorized that it was at least partially the result of universities placing more value on publications and thus under-incentivizing service work (e.g., conference reviewing). While not specific to conference reviewing, debates published by the Communications of the Association for Information Systems (CAIS) on the quality of reviews (Jennex, 2016) and the difficulty with recruiting quality reviewers (Stafford, 2018) both note this issue of misaligned incentives that reduce the number of scholars willing to contribute their time as peer reviewers.

An additional challenge in recruiting conference reviewers is the nature of conference timelines, which require that all submissions be assigned to reviewers and the reviews completed within a very short window of time. As a result, AEs are asked to secure commitments from reviewers well ahead of the submission deadline to ensure the review process stays on track. Urquhart et al. (2017) note how this results in fit issues as pre-committed reviewers often lack the expertise to effectively review assigned papers. We have experienced that issue firsthand as conference AEs. For example, in the ICIS Information Systems in Healthcare track, the one connecting thread for submissions is the healthcare context. Consequently, track submissions can span various methodologies, technologies, and theories, making it likely that a pre-committed set of reviewers will not have the expertise to effectively review their assigned papers. The result might be lower average review quality than if editors waited to invite reviewers based on their fit with each paper.

Conference review quality is also impacted by the fact that the burden of reviewing tends to fall disproportionally on junior faculty and doctoral students (Chua et al., 2018). We have personally observed an increasing number of students serving as reviewers for IS conferences. On the one hand, there is benefit to the students and the IS community as new reviewers are trained and become more experienced. However, on the other hand, review quality tends to deteriorate when the reviewer pool is predominantly composed of inexperienced reviewers. Some conferences have responded to this issue by requiring that the set of reviewers for any paper must not be exclusively doctoral students.

Given the above, the two challenges we focus on regarding conference peer reviewing in IS are a shortage of qualified reviewers and inconsistent review quality. While there is a natural solution to the former issue (i.e., attract more qualified reviewers), there is no clear-cut solution to the latter, as quality reviews result from several factors. For example, reviewers might lack expertise, have time constraints, be unwilling to be critical of other researchers, or lack training. This paper focuses on a contributing quality factor that we have experienced in IS conferences related to conflicts of interest. While publication in any quality outlet is competitive, prestigious academic conferences are particularly so, where even high-quality submissions may be rejected due to venue space and time constraints for presentations. Within that setting, we have observed the practice of conference track editors/chairs inviting researchers who submit their work to a track and also review papers submitted to the same track. Such practices can create a conflict of interest for reviewers to intentionally review a paper negatively to raise their own chances of acceptance. The experiments by Balietti et al. (2016) have indeed corroborated that hypothesis, where they found that "*even if outcomes of peer review are generally valid, competition increases editorial type II errors and encourages self-interested referees to behave strategically.*" Here, Type II error means that more high-quality submissions are rejected.

In what follows, we suggest potential ideas for the IS community to contemplate in order to improve the peer-review process in IS conferences. The ultimate goal behind the offered suggestions is to boost the number of qualified reviewers while also addressing other issues, such as misaligned incentives due to conflicts of interest.

3 Incentivizing Reviewers in IS Conferences

Given the importance of peer review, guidelines by senior scholars have been proposed to help reviewers (Rai, 2016) and authors (Pang & Thatcher, 2023) navigate the process. While such guidelines are a step in the right direction, issues with the peer review process described above continue to persist. Consequently, we propose three potential changes to the peer review process in IS conferences that can increase the number of peer reviewers while addressing the hypothesized misaligned incentives due to conflicts of interest: 1) quid pro quo rules; 2) the use of incentive-compatible scoring methods that reward truth-telling; and 3) the use of blockchain-based tokens coupled with smart contracts and zero-knowledge proofs.

3.1 Quid Pro Quo Rules

One of the most straightforward approaches to increasing the number of peer reviewers is to impose quid pro quo rules (Jennex, 2016). In the context of IS conferences, a quid pro quo rule would require authors who submit a paper for consideration to also review papers for the conference. For example, suppose a manuscript co-authored by two researchers is submitted to a conference, and two peers will review this submission. In that case, the two co-authors must give back to the conference by collectively reviewing two papers. Quid pro quo rules are part of what Avital (2018) called the "operational regulation" of peer review (i.e., the bureaucratic mechanisms that govern the peer-review process). We note that some conferences have already implemented quid pro quo rules. For example, in its 2023 edition, the prestigious ACM CHI Conference on Human Factors in Computing Systems required authors to acknowledge at the submission time that they would review papers for the conference if an invitation happened.

We acknowledge that quid pro quo rules can create conflicts of interest if not implemented appropriately. Specifically, as previously suggested, space limitations for accepted papers can incentivize authors serving as reviewers to recommend rejecting competing papers to boost their own papers' chances of acceptance. When constraints on the number of accepted papers are defined at the track level, a potential solution is for track editors/chairs to invite authors from other tracks to review papers with some elements within their expertise. This could be facilitated by requiring authors to list all their areas of expertise when they submit a paper. Taking this paper as an example, it could be reviewed by scholars with expertise in blockchain, incentive mechanisms, or familiarity with the peer review process. At the same time, we acknowledge that cooperation between tracks can be hard to achieve, especially in large flagship conferences (e.g., ICIS). Having authors review papers outside the track they submit to would remove the conflict of interest to reject papers solely to increase their chances of acceptance. However, it does not directly incentivize truthful/honest reporting, which is a factor in review quality. We discuss in the following subsection a complementary approach that directly incentivizes truth-telling.

3.2 Truth-Telling Scores

While quid pro quo rules may directly address issues related to the lack of reviewers in IS conferences, they do not directly address the quality and truthfulness of the reviews. In an ideal world, an operational regulation mechanism should be established to assess reviewer quality objectively. Some related ideas have been suggested, such as using fictitious manuscripts to evaluate peer reviewer performance (Baxt et al., 1998). However, we argue that such approaches consume precious resources by taking reviewers from a limited pool to review submissions that are not real. While it can be hard to objectively evaluate the quality of a review, we suggest developing scoring techniques to induce honest reporting. To discuss this idea, we contemplate the peer review process under the lens of decision analysis, in which a decision-maker requests subjective information from experts to make a decision, such as accepting or rejecting a paper. The experts, in turn, observe private signals (i.e., the quality of a manuscript) and, ideally, report that signal truthfully; otherwise, the decision maker may decide based on random input. In other words, experts should not have incentives to misreport their true opinions on a paper. Several truth-inducing methods, also called *incentive-compatible* techniques, have been proposed in the decision analysis and theory literature. For example, proper scoring rules (Carvalho, 2016) are well-established statistical methods that induce risk-neutral experts to report subjective forecasts honestly. The two main issues when applying proper scoring rules in the context of peer review are: 1) reviews are often reported as an answer to a multiple-choice question (e.g., accept, major/minor revisions, reject) instead of forecasts; and 2) proper scoring rules assume that there exists an observed outcome that symbolizes a ground truth. Even if such a ground truth existed in peer review, it would not be observable.

Different techniques have been proposed to address the above limitations. One of them, which we argue may be suitable for peer reviews, is the Bayesian Truth Serum (BTS) method by Prelec (2004). Under that method, reviewers would report not only an answer to a multiple-choice question but also the empirical distribution of answers. Reviewers are then evaluated based on the accuracy of their predictions (i.e., how well they match the empirical frequency) as well as how *surprisingly common* their answers are. For example, an answer endorsed by 50% of the reviewers (say, "rejection") against a predicted frequency of 25% is surprisingly common. Therefore, the reviewers who endorsed that answer should receive a high score. If predictions averaged 75%, an answer endorsed by 50% of the reviewers would be surprisingly uncommon and, consequently, the reviewers who endorsed it should receive a lower score. To better understand the mechanics behind the Bayesian Truth Serum method, consider A as the set of outcomes from the peer review process for a single paper, where $|A| = n$. For example, often $A = \{\text{"}Reject\text{"}, \text{"}Minor\ Revisions\text{"}, \text{"}Major\ Revisions\text{"}, \text{"}Accept\text{"}\}$. The method works on a single decision-making process (e.g., paper submission) at a time, where each reviewer is asked to report two pieces of information. First, the vector $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$, for $x_i \in \{0, 1\}$ and $\sum_{i=1}^{n} x_i = 1$, contains a single non-zero value representing the outcome selected by the reviewer. For example, $\boldsymbol{x} = (0, 0, 0, 1)$ means that a reviewer suggests accepting the paper. Besides endorsing an outcome representing their belief, each reviewer must predict the proportion of reviewers who will endorse each of the n outcomes. Formally, each reviewer also reports a vector $\boldsymbol{y} = (y_1, y_2, \ldots, y_n)$, for $0 \leq y_i \leq 1$ and $\sum_{i=1}^{n} y_i = 1$, encoding the guessed proportion of reviewers endorsing each peer-review outcome. For example, $\boldsymbol{y} = (0, 0, 0.66, 0.34)$ means that a reviewer believes two-thirds of the reviewers will suggest major revisions, whereas about one-third will suggest acceptance.

Given $\boldsymbol{x}$ and $\boldsymbol{y}$, the score received by each reviewer comprises two components, namely an *information score* and a *prediction score*. Say a reviewer endorses the fourth output (e.g., "Accept"), i.e., $x_4 = 1$. That reviewer then receives an information score equal to $\ln \frac{\bar{x}_4}{\bar{y}_4}$, where $\bar{x}_4$ is the proportion of the reviewers endorsing the fourth output as their belief, and $\bar{y}_4$ is the geometric mean of the endorsement predictions for the fourth output. In other words, reviewers maximize their information scores when their endorsements are surprisingly common (i.e., when they are more common than what is collectively predicted). Mathematically, this translates into $\bar{x}_4$ being greater than $\bar{y}_4$. The prediction score ensures that each reviewer's forecast is also reported honestly. In particular, it is a penalty term proportional to the Kullback-Leibler divergence between the empirical distribution of endorsed outcomes and one's prediction of that distribution. Putting the above together, each reviewer's BTS score is:

$$\sum_{i=1}^{n} x_{\mathrm{i}} \times \log \frac{\bar{x}_i}{\bar{y}_i} + \sum_{i=1}^{n} \bar{x}_{\mathrm{i}} \times \log \frac{y_i}{\bar{x}_i}$$

Two changes to traditional peer-review models adopted by IS conferences are necessary if the BTS method is to be used appropriately. First, in addition to reporting their outcome recommendation, reviewers must report what they believe others will recommend. Second, scores must be meaningful so that reviewers will indeed try to maximize them in expectation, and, thus, truth-telling becomes the best reporting strategy. In the context of peer review for IS conferences, that score can naturally be attached to awards, such as best reviewer awards for each track. An additional benefit of doing so is that it removes the subjective nature of selecting awardees. Although the original formulation of the BTS method requires a large population of experts, adaptations have been made for small groups (Witkowski & Parkes, 2012) that would enable BTS to be used in peer-review settings. Interestingly, the BTS method has been used to elicit truthful information from researchers regarding questionable research practices (John et al., 2012). Moreover, its use has been linked to greater accuracy (i.e., higher quality responses) in qualitative content analysis than several alternative methods (Shaw et al., 2011). We discuss next how blockchain can be the foundational technology to implement incentive-compatible scoring methods, such as BTS, and quid pro quo rules in a trusted manner.

3.3 Tokens, Smart Contracts, and Zero-Knowledge Proofs

Although often associated with cryptocurrencies, blockchain technology has been proposed to solve a range of problems requiring immutable and decentralized information storage (Carvalho et al., 2020). For example, in the context of peer review, Avital (2018) has suggested that blockchain technology can be the backbone of a market-based peer review process. This market would operate on peer-review coins, which authors would use to pay submission fees. These coins would then be awarded to peers whenever they review papers. Moreover, coins could be traded (purchased and sold) among researchers. The idea of financially rewarding reviewers has received some attention. For example, at the time of writing, journals affiliated with the American Economic Association pay US$100 to each reviewer.[1] Often, such payments come from paper submission/processing fees. Similar schemes can offset the massive time-related cost associated with peer reviews, estimated to be US$1.5 billion annually in the United States alone (Aczel et al., 2021). Some advantages of paying reviewers include increasing the review pool, speed, and potential quality (Cheah & Piasecki, 2022). In contrast, some concerns include an increase in the total research budget due to article processing charges and, as Moustafa (2022) suggested, the rise of "*new commercial peer review agencies, cronyism or nepotism reviewing activities*." There is also the issue of how much money reviewers/authors should receive/pay (Diamandis, 2015). Naturally, a tiered system charging less from authors from less economically developed nations or even junior researchers without well-established funding sources should be in place. Besides ethical and economic considerations resulting from incorporating market mechanisms into the peer review process, we argue that such an approach might not be entirely feasible for IS conferences as they often happen in different places, sometimes virtually, sporadically, and at different times. That means the demand for financial assets, such as peer-review coins, might fluctuate wildly, leading to possible speculatory and even illiquid markets.

Instead, we suggest using blockchain as a neutral, transparent, and trusted technology to implement the previously proposed ideas to boost the quantity and fix some misaligned incentives in peer reviews. Due to its technological and governance complexities, a common question in applications of blockchain technology is whether blockchain is indeed needed. Some decision models have been proposed to guide one in answering that question. For example, Figure 1 shows an adapted version of the model by Wüst and Gervais (2018). The first step in that model is to answer whether one needs to store data. The answer is unequivocally yes in the context of peer review processes in IS conferences. Specifically, designing incentives to attract reviewers requires data about peer reviews, such as the decision outcomes and reviewers' scores and suggestions. The second question regards the existence of multiple data writers. This is also true since different reviewers, associate editors, and chairs enter their own peer-review-related data into a storage structure via submission management systems. The third question regards trust in a third party. It is unrealistic and unreasonable to assume a third party can write data on behalf of reviewers, editors, or chairs. The fourth question is whether all the data writers are known.

[1] https://www.aeaweb.org/journals/aer/reviewers

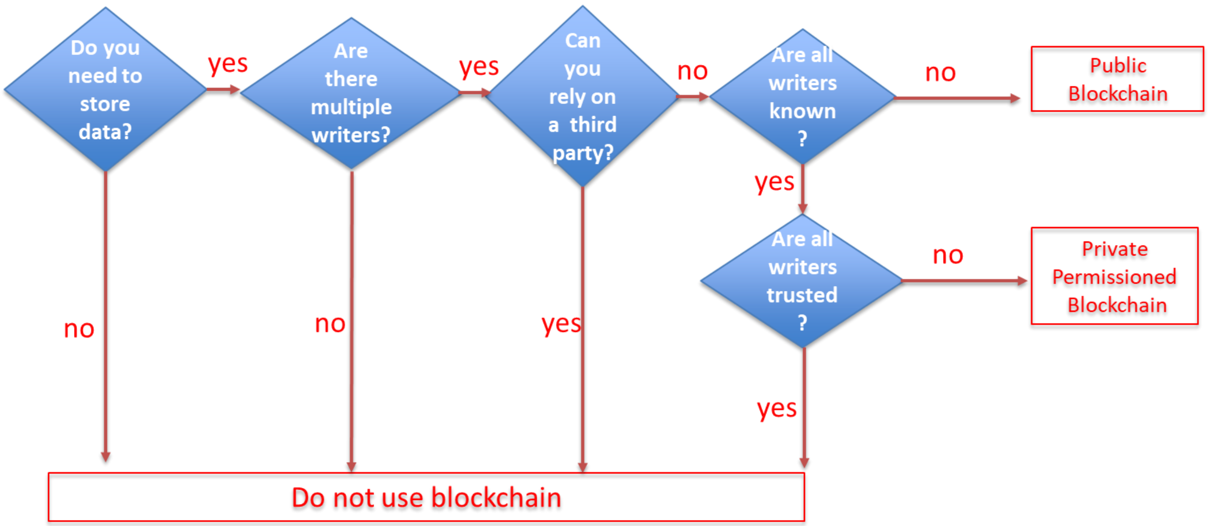

Figure 1. Decision Model Regarding Blockchain Adoption (Adapted from: Wüst and Gervais (2018)).

We argue that this is partially the case as editors and chairs tend to know the reviewers' identities and vice versa, although reviewers might not know each other's identities. Lastly, as we have elaborated throughout the paper, conflicts of interest can lead reviewers to not report their honest opinions. Thus, one cannot assume that all data writers are trusted.

The result of this decision model is that *permissioned blockchains* are suitable candidates to form the infrastructure required to instantiate our ideas. Unlike *public blockchains*, in which anyone can at any time read and create transactions or join the blockchain network as a computational node, permissioned blockchains are business networks of known, vetted participants operating under a well-defined governance model. Knowing the participants' identities allows permissioned blockchains to be more efficient, e.g., Bitcoin's onerous proof-of-work consensus mechanism, often criticized for being energy-intensive (Foteinis, 2018; Howson, 2019), can be replaced by more traditional and less energy-intensive Byzantine fault tolerance schemes. Regarding the governance of such a network, we envision a consortium of associations, such as the Association for Information Systems (AIS), and publishers operating the IT infrastructure.

With that blockchain network storing peer-review data, quid pro quo rules can be implemented through non-transferable tokens, where the submission rules are defined by smart contracts (i.e., algorithms running on the blockchain). A potential implementation of an extreme submission rule would be to give authors generic non-transferable tokens whenever they finish reviewing a paper. Then, when the review deadline is passed, the papers from authors whose token balances are not sufficiently above a threshold — meaning they did not review enough papers — would be automatically rejected. The downside to this rule might be a decrease in the quality of reviews as authors complete reviews simply to meet the review threshold. We return to this point later.

Ideally, the IS community should collectively define the quid pro quo rules. Blockchain provides an excellent opportunity for the community to try different governance mechanisms, such as voting on proposals through *decentralized autonomous organizations*. The above non-transferable tokens naturally translate, at least partially, into voting power. Given the immutable nature of blockchains, this solution is desirable not only for the operational regulation of peer reviews but also to prove that authors performed external service by reviewing papers. This information, in turn, can be used in various ways, such as in tenure and promotion packets, grant applications, and more creative ones, such as discounts in future conferences or opportunities to chair tracks. The use of non-transferable tokens effectively implements the idea by Urquhart et al. (2017), who suggested that "*Reviews that are completed on time and have sufficient quality would result in some sort of token being exchanged. Authors would then be required to be in possession of some number of tokens (perhaps three or four) in order to submit a paper*".

For these incentives to work, reviewers' true identities must be linked to a blockchain identifier (address), which can bring some privacy issues depending on what is stored on the blockchain. For example, suppose a reviewer's recommendation for a given paper is publicly available on the blockchain. In that case, that information will be indirectly linked to the reviewer's identity, thus violating the blind-review policies of many IS conferences. At the same time, reviewers' recommendations are required by the BTS or similar truth-inducing scoring systems. A potential solution to this dilemma is to avoid storing sensitive information on the blockchain, such as reviewers' recommendations, while using cryptographic primitives, such as zero-knowledge proofs, for the truth-inducing scoring systems. Our vision is that reviewers should be the ones reporting their received truth-telling scores to the blockchain alongside verifiable proofs that the scores are indeed accurate, all of that without disclosing any other information to the community (i.e., "zero knowledge"), such as the components of the BTS scoring system or which papers were reviewed. In practice, many technical challenges must be solved to make such a vision feasible. For example, BTS requires calculating the average of reviewers' reports, which in turn requires private communications among reviewers or with a trusted computational entity. Despite the challenges, we believe the proper development and implementation of the above blockchain-based solution can help IS conferences continue to prosper in a more transparent and trusted way.

3.4 Discussion

We have discussed two solutions to current issues faced by the peer review process as IS conferences currently implement it. First, we proposed implementing quid pro quo rules to handle the scarcity of qualified reviews. Second, we elaborated on review quality issues as well as potential conflicts of interest that may arise when authors also review papers in the same conference track and proposed the use of truth-inducing mechanisms, such as the BTS method, to address these issues. Finally, we explained how blockchain technology, in tandem with non-transferable tokens and zero-knowledge proofs, can form the foundational information technology to bring our ideas to life.

We acknowledge that our proposed solutions are not without potential unintended consequences. For example, while quid pro quo rules would increase the number of reviewers, Chua et al. (2018) note that "*both good and bad reviewers would review under a mandatory system*." Specifically, authors with limited reviewing experience could end up reviewing more papers under a quid pro quo system than if reviewing was voluntary. To understand why, let us return to the example where a submission would require the paper's authors to collectively perform two reviews. In many cases, conference papers are co-authored by some combination of senior faculty, junior faculty, and doctoral students. Senior faculty overloaded with teaching, research, and service tasks could pass the reviewing responsibility to their junior colleagues and doctoral students. As a result, conference chairs and AEs would be further challenged to balance the review teams with both novice and experienced reviewers. Despite this potential drawback and the need for careful implementation, the idea of quid pro quo rules has some acceptance within the IS community. For example, Chua et al. (2018) reported that around 90% of the nearly 70 participants in a peer review panel at the 2017 ICIS reported agreeing with the idea that a journal's author should be required to review for that journal.

It is also important to acknowledge some limitations involving the BTS and similar methods. First, the BTS method relies on the *surprisingly common criterion* to promote truthful reporting, as that action maximizes the score received by *risk-neutral* experts. Other incentive engineering techniques rely on different cognitive biases, such as the *false consensus effect* — i.e., the general tendency of experts to overestimate the degree of agreement that others have with them (Ross et al., 1977) — to induce truthfulness (Carvalho et al., 2014). However, the extent to which these biases are prevalent in the peer review process is an open question. In particular, cultural differences, unique or uncommon beliefs, and, more broadly, personal differences might impact such cognitive biases. Second, most truth-inducing elicitation techniques rely on the assumption that the experts are risk-neutral, and truthfulness properties may no longer hold true under different risk attitudes (Carvalho et al., 2018). These two concerns open the doors to meaningful research into the behavior and preferences of peer reviewers, particularly in IS conferences.

For the BTS or similar scoring system to be trusted, its implementation must be transparent and verifiable. That observation led to our suggestion for using blockchain technologies. However, blockchain is also associated with potential usage barriers due to the complexity of, for example, setting up wallets, defining governance structures, and scalability and performance issues.

We conclude this subsection by noting that our suggestions for improvements in IS conferences' peer reviews assume a double-blind peer review process. That is, the identities of both the authors and the reviewers are kept confidential from each other, which is currently standard in IS conferences. It has been suggested that a double-blind peer review "*minimizes positive and negative biases between editors, reviewers, and authors*" (Chua et al., 2018). That process can be made more transparent in different ways. For example, reviewers' comments and/or identities can be disclosed to the authors and the community during and/or after the peer review process. Besides increasing transparency, one can argue that a more productive and constructive revision process may arise if the authors know the reviewers' identities and areas of expertise. At the same time, some issues can result from such an open practice, (e.g., reviewers might be less willing to review papers, thus exacerbating the issue of the lack of reviewers). For example, only 8.1% of the reviewers agreed to reveal their identities in a large-scale study consisting of 9,220 submissions and 18,525 reviews by the publisher Elsevier (Bravo et al., 2019). Of those who disclosed their information, less than 10% of the reviewers suggested rejecting the submission (i.e., the vast majority reported either a revision or an acceptance decision). The above result indicates that reviewers are sensitive to possible retaliation by authors and would prefer to opt out when reporting on a weak manuscript. Recent simulation studies further suggest that retaliatory practices may arise if referees are sensitive to competition and status (Bianchi & Squazzoni, 2022). Disclosing reviews to the community can also bring unintended consequences, even when the reviewer's name is masked. For example, modern machine learning algorithms can identify the most likely author of a manuscript with an accuracy of over 90% (Anwar et al., 2019). This discussion highlights the need to protect the reviewers' identities and, thus, our suggestion for strict measures to preserve anonymity, such as using zero-knowledge proofs.

4 Conclusion

In response to the call for debate, we have offered suggestions to improve the peer-review processes implemented by IS conferences. Specifically, we directly address the discussion point: "*Are there changes to the review system that are warranted* (*or worth experimenting with*)?" Our proposed solutions are based on quid pro quo rules and incentive-compatible scoring systems implemented via a blockchain-based system. Our ultimate goal is to increase the number of peer reviewers and the quality/trustworthiness of reviews while boosting the transparency and reliability of the peer-review process.

The above solutions are responses to our concerns regarding the pool of reviewers in IS conferences having too many inexperienced researchers as well as to observed practices that may create conflicts of interest by requesting authors who submit papers to a track to also review papers for that track. We argue that the inherently competitive nature of conferences may incentivize those reviewers to blindly reject papers to increase their own chances of acceptance. We acknowledge that the above issues could benefit from better evidence (data) to quantify them, and that is our suggestion for a first step toward bettering the peer-review process. For example, future IS conference organizers should be able to collect data on the proportion of reviewers who are doctoral students. Moreover, the organizers should be able to collect and analyze data to determine the difference between the recommendations by reviewers who are authors of submitted papers and those who are not. A statistical difference between those recommendations may help (in)validate our hypothesized conflicts of interest. Having validated the issues, our innovative blockchain-based solution presents an opportunity for the IS community to maintain its role as stewards of the digital revolution and the creators of innovative artifacts that can impact knowledge creation and, consequently, society.

References

Aczel, B., Szaszi, B., & Holcombe, A. O. (2021). A billion-dollar donation: Estimating the cost of researchers' time spent on peer review. *Research Integrity and Peer Review*, *6*(1), 1-8.

Anderson, M., Ronning, E., De Vries, R., & Martinson, B. (2007). The perverse effects of competition on scientists' work and relationships. *Science and Engineering Ethics*, *13*, 437-461.

Anwar, W., Bajwa, I. S., & Ramzan, S. (2019). Design and implementation of a machine learning-based authorship identification model. *Scientific Programming*, *2019*.

Avital, M. (2018). Peer review: Toward a blockchain-enabled market-based ecosystem. *Communications of the Association for Information Systems, 42*, 646-653.

Balietti, S., Goldstone, R. L., & Helbing, D. (2016). Peer review and competition in the Art Exhibition Game. *Proceedings of the National Academy of Sciences*, *113*(30), 8414-8419.

Baxt, W. G., Waeckerle, J. F., Berlin, J. A., & Callaham, M. L. (1998). Who reviews the reviewers? Feasibility of using a fictitious manuscript to evaluate peer reviewer performance. *Annals of Emergency Medicine, 32*(3), 310-317.

Bianchi, F., & Squazzoni, F. (2022). Can transparency undermine peer review? A simulation model of scientist behavior under open peer review. *Science and Public Policy, 49*(5), 791-800.

Bravo, G., Grimaldo, F., López-Iñesta, E., Mehmani, B., & Squazzoni, F. (2019). The effect of publishing peer review reports on referee behavior in five scholarly journals. *Nature Communications, 10*(1), 322.

Carvalho, A. (2016). An overview of applications of proper scoring rules. *Decision Analysis, 13*(4), 223-242.

Carvalho, A., Dimitrov, S., & Larson, K. (2014). The output-agreement method induces honest behavior in the presence of social projection. *ACM SIGecom Exchanges, 13*(1), 77-81.

Carvalho, A., Dimitrov, S., & Larson, K. (2018). On proper scoring rules and cumulative prospect theory. *EURO Journal on Decision Processes, 6*(3-4), 343-376.

Carvalho, A., Sambhara, C., & Young, P. (2020). What the history of Linux says about the future of cryptocurrencies. *Communications of the Association for Information Systems, 46*, 18-29.

Cheah, P. Y., & Piasecki, J. (2022). Should peer reviewers be paid to review academic papers? *The Lancet, 399*(10335), 1601.

Chua, C. E. H., Thatcher, J. B., Niederman, F., Chan, Y. E., & Davidson, E. J. (2018). ICIS 2017 panel report: Break your shackles! Emancipating information systems from the tyranny of peer review. *Communications of the Association for Information Systems*, *43*(1), 442-465

Diamandis, E. P. (2015). Peer review as a business transaction. *Nature*, *517*(7533), 145-145.

Eve, K., Mehmani, B., & Wilson, K. (2021). The case for (more) diversity in peer review. *Elsevier Connect*. Retrieved from https://www.elsevier.com/connect/reviewers-update/the-case-for-more-diversity-in-peer-review.

Foteinis, S. (2018). Bitcoin's alarming carbon footprint. *Nature*, *554*(7690), 169-170.

Frachtenberg, E., & McConville, K. S. (2022). Metrics and methods in the evaluation of prestige bias in peer review: A case study in computer systems conferences. *PLOS One, 17*(2), e0264131.

Gardner, A., Willey, K., Jolly, L., & Tibbits, G. (2012). Peering at the peer review process for conference submissions. In *Proceedings of the 2012 Frontiers in Education Conference.*

Howson, P. (2019). Tackling climate change with blockchain. *Nature Climate Change*, *9*(9), 644-645.

Jennex, M. (2016). No free lunch: Suggestions for improving the quality of the review process. *Communications of the Association for Information Systems, 38*, 307-313.

John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological Science, 23*(5), 524-532.

Lee, J. C., Sugimoto, R. C., Zhang, G., & Cronin, B. (2012). Bias in peer review. *Journal of the American Society for Information Science and Technology, 64*(1), 2-17.

Moustafa, K. (2022). No to paid peer review. *The Lancet*, *400*(10347), 160.

Pang, M. S., & Thatcher, J. B. (2023). A practical guide for successful revisions and engagements with reviewers. *Journal of the Association for Information Systems, 24*(2), 317-327.

Prelec, D. (2004). A Bayesian truth serum for subjective data. *Science, 306*(5695), 462-466.

Rai, A. (2016). Writing a virtuous review. *MIS Quarterly, 40*(3), iii-x.

Ross, L., Greene, D., & House, P. (1977). The “false consensus effect”: An egocentric bias in social perception and attribution processes. *Journal of Experimental Social Psychology, 13*(3), 279–301.

Shaw, A. D., Horton, J. J., & Chen, D. L. (2011). Designing incentives for inexpert human raters. In *Proceedings of the ACM 2011 Conference on Computer Supported Cooperative Work* (pp. 275-284).

Smith, R. (2006). Peer review: A flawed process at the heart of science and journals. *Journal of the Royal Society of Medicine, 99*(4), 178-182.

Stafford, T. (2018). Reviews, reviewers, and reviewing: The “Tragedy of the Commons” in the scientific publication process. *Communications of the Association for Information Systems, 42*, 624-629.

Urquhart, C., Carte, T., & Heinzl, A. (2017). Time for some changes to ICIS? Reflections on our highest-quality conference. *Communications of the Association for Information Systems, 41*, 179-197.

Witkowski, J., & Parkes, D. C. (2012). A robust Bayesian truth serum for small populations. In *Proceedings of the 26th AAAI Conference on Artificial Intelligence* (pp. 1492-1498).

Wüst, K., & Gervais, A. (2018). Do you need a blockchain? In *Proceedings of the 2018 Crypto Valley Conference on Blockchain Technology* (pp. 45-54).

About the Authors

Arthur Carvalho is the Dinesh & Ila Paliwal Innovation Chair and an Associate Professor of Information Systems & Analytics at Farmer School of Business, Miami University. Dr. Carvalho's research interests lie at the intersection of Artificial Intelligence, Information Systems, and Decision Analysis. In particular, Dr. Carvalho works on the development and deployment of artificial intelligence techniques, statistical methods, and blockchain-based systems to improve business processes and decision-making. Dr. Carvalho has collaborated with several companies through research work, consulting, and/or joint student supervision, including Microsoft, KPMG, IBM, PwC, Cardinal Health, and Unilever. His research has been published in academic journals such as Decision Support Systems, Decision Analysis, International Journal of Information Management, among many others. In addition, Dr. Carvalho's views have been featured in outlets such as Bloomberg, Yahoo Finance, Inside Higher ED, and Poets & Quants.

Chad Anderson is an Assistant Professor of Information Systems & Analytics at the Farmer School of Business, Miami University. He received his Ph.D. in Computer Information Systems from Georgia State University. His research focuses on the role of information technology in the delivery of health care, including issues of clinician burnout, regulatory compliance, and information security and privacy. His work has been published in Information and Organization, Journal of the Association for Information Systems, Journal of Management Information Systems, MIS Quarterly, and other academic outlets.

Liudmila Zavolokina is a postdoctoral fellow and group leader at the Digital Society Initiative, University of Zurich and an incoming Assistant Professor at the Faculty of Business and Economics (HEC Lausanne), University of Lausanne. She holds a Ph.D. from the University of Zurich. Dr. Zavolokina's research interests include blockchain, digital platforms and ecosystems, data-driven decision-making, and public value creation. Dr. Zavolokina has published research in high-quality information systems outlets and received several best paper awards and nominations, such as the recent best paper award at the HICSS Digital Government Track. Dr. Zavolokina has collaborated with public authorities in Switzerland and Germany and the private sector, including the automotive industry and crypto startups.