
Information Systems Research

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

# Research Spotlights

To cite this article:
(2025) Research Spotlights. Information Systems Research 36(4):iv-x. https://doi.org/10.1287/isre.2025.resspot.v36.n4

https://pubsonline.informs.org/journal/

INFORMATION SYSTEMS RESEARCH
Vol. 36, No. 4, December 2025, pp. iv–x
ISSN 1047-7047 (print), ISSN 1526-5536 (online)

# Research Spotlights

https://doi.org/10.1287/isre.2025.resspot.v36.n4





**Healthcare Cost Prediction for Heterogeneous Patient Profiles Using Deep Learning Models with Administrative Claims Data** (p. 1968)
Mohammad Amin Morid, Olivia R. Liu Sheng

Accurate and equitable patient cost prediction is essential for informing health management policies and optimizing resource allocation, directly impacting government agencies, private insurers, and healthcare providers. This study highlights the importance of addressing disparities in prediction outcomes, particularly for high-need patients with complex chronic conditions, to ensure more effective economic and clinical decision making. By introducing a novel deep learning framework that segments administrative claims data into multiple channels, this research enhances both predictive accuracy and fairness, reducing overpayments and underpayments while mitigating bias in cost estimation. The findings underscore the potential of channel-wise modeling to support fair reimbursement structures, improve budget planning, and foster policies that better accommodate the diverse needs of patient populations. Policymakers and healthcare organizations can leverage these insights to design more efficient risk adjustment strategies, ensuring that vulnerable patients receive appropriate care without financial inefficiencies. The study provides a roadmap for integrating advanced machine learning approaches into healthcare decision making, promoting a more just and sustainable system.

**Beyond Complements and Substitutes: A Graph Neural Network Approach for Collaborative Retail Sales Forecasting** (p. 1993)
Jing Liu, Gang Wang, Huimin Zhao, Mingfeng Lu, Lihua Huang, Gang Chen

This paper proposes a novel approach for enhanced sales forecasting by leveraging multifaceted product relations, disentangled on the ground of the cross-category choice dependence theory. With superior forecasting performance over state-of-the-art alternatives and a deep understanding of product relations, the proposed approach has significant practical implications for various stakeholders (e.g., retail store managers, inventory department, purchasing department, operational staff, marketers, and retail platforms). On the one hand, improved forecasting could provide solid data-driven decision support for supply chain management, resource planning, inventory control, and purchasing planning. The semblance of predictive power in sales forecasting demonstrates operational utility. On the other hand, derived insights on product relations could facilitate reasonable pricing and promotion strategies, enhance the relevance and diversity of recommendation systems, and provide benefits for assortment planning, cross-selling, and shelf space allocation.

**Routine and Innovative Use of Enterprise System: Intricacy of Change Management Levers, System Characteristics, and Regulatory Focus** (p. 2017)
Shaobo Wei, Weiling Ke, Chuan-Hoo Tan

This study explores how organizational contexts, system design, and individual differences shape employees' use of enterprise systems (ESs) and their impact on job outcomes. By examining leadership styles (transactional versus transformational), support structures (impersonal versus personal), and system characteristics (modularity and complexity), we highlight the role of employees' situational regulatory focus (prevention versus promotion) in determining routine and innovative use of ES. Our findings show that leadership and support structures affect ES usage through employees' regulatory focus, with transactional leadership and impersonal support fostering routine use and transformational leadership and personal support encouraging innovative use. System modularity enhances the relationship between prevention focus and routine use, whereas complexity strengthens the link between promotion focus and innovative use but weakens the relationship between prevention focus and routine ES use. Both routine and innovative ES use influence the employees' job outcomes, including job performance and satisfaction. For organizations, this means that effective change management and ES design should align with employees' regulatory focus and leadership strategies to maximize system adoption and improve performance. This study offers practical insights for designing more effective ES implementation strategies and tailoring support to boost employee engagement and outcomes.

**Beyond Digital vs. IT: The Untold Story of Their Relationship from an Organizing Logic Perspective** (p. 2039)
Abayomi Baiyere, Markus P. Zimmer, Kalina S. Staykova, Jan Jöhnk

Organizations often feature digital and information technology (IT) units, suggesting that managers perceive digital technology as different from IT. However, these units do not coexist in silos; rather, they interact in pursuit of organizational goals. In this study, we investigate the interactions between the digital and IT units of three organizations undergoing digital transformation. We find that these interactions reflect three relationships with varying dynamics. The interplay of these dynamics shape organization's digital transformation efforts. We outline three considerations for managing these dynamics productively. First, we outline conditions that determine the dynamic (e.g., synergistic or conflicting) more likely to manifest. These conditions depend on the (in)compatibility between the units' need to interact and the strategic, routine or technology rationales they draw on. We caution against seeking compatibility at all costs. Rather, managers should consider the context of each interaction before deciding how to influence these dynamics. Second, we found that interpersonal relationships, hiring talent with diverse perspectives, or establishing idea exchange forums can help to foster productive collaborations. However, they alone do not determine these dynamics. Third, cross-unit relationships is important for accomplishing organizational goals like digital transformation. Managers should proactively nurture such relationships to foster collaboration among units.

**Lost in the Crowd: How Group Size and Content Moderation Shape User Engagement in Live Streaming** (p. 2076)
Keran Zhao, Yili Hong, Tengteng Ma, Yingda Lu, Yuheng Hu

Live streaming platforms such as Twitch and YouTube Live now play a central role in digital engagement, offering real-time interaction that grows increasingly complex with larger audiences. Our study leverages the exogenous viewer influx from Twitch's Raid function to examine how increases in group size affect user engagement. Analyzing chat histories from more than 7,000 playbacks using a difference-in-differences approach, we find that although attracting more viewers, larger group sizes also lead to reduced engagement among existing participants. This decline is linked to increased topic incoherence and heightened emotional volatility—or comment polarity—in live chats. Importantly, our results highlight that targeted moderation strategies can mitigate these negative effects. Bot moderators are particularly effective in maintaining coherence during large-scale raids, whereas human moderators better manage emotional surges when the incoming group is smaller. These findings reveal a congestion effect in synchronous digital environments and offer clear practice- and policy-oriented implications: Online synchronous platforms should consider the scalability of viewers and commenters and use flexible moderation strategies to sustain user engagement and foster healthier, more constructive real-time interactions.

**Algorithms to the Rescue: Market Mechanisms for Consensual Trading of Unbiased Individual Data** (p. 2096)
Brian Birkhead, Ashkan Eshghi, Ram D. Gopal, Hooman Hidaji, Raymond A. Patterson

This paper proposes a novel algorithmic market mechanism to address key challenges in individual data markets. Current data collection practices lack transparency and proper compensation, leading privacy-conscious users to opt out and creating biased data sets. Our proposed mechanism enables an intermediary platform to obtain unbiased samples of individual-level data while appropriately compensating users for privacy loss. Through theoretical analysis and simulations using both synthetic and real-world data sets, the authors demonstrate that their mechanism provides unbiased data samples at near-optimal cost compared with benchmark approaches. The mechanism outperforms both fixed-compensation methods and centralized-optimization approaches, even when platforms have partial information about user privacy preferences. Surprisingly, platforms achieve better outcomes by using this market mechanism rather than relying on estimated privacy preferences from user behavior. The approach is practical to implement, using straightforward sampling and conventional compensation mechanisms rather than complex techniques, like differential privacy. The mechanism enables creation of effective data markets that benefit both data subjects and buyers while ensuring compliance with regulations requiring transparency and consent. The findings are particularly relevant as new privacy regulations emerge globally and third-party tracking faces increased constraints, providing a viable solution for improving data quality and fairness in digital markets.

**Artificial Intelligence and Firm Resilience: Empirical Evidence from Natural Disaster Shocks** (p. 2116)
Miaozhe Han, Hongchuan Shen, Jing Wu, Xiaoquan (Michael) Zhang

Artificial intelligence (AI) has been increasingly deployed in business operations over the past decade, whereas direct evidence of its effectiveness in uncertain contexts is limited. Our work examines the contribution of AI to corporate resilience under natural disaster shocks, particularly concentrating on AI-using and

goods-producing firms. We measure firm AI investment by the cumulative AI-relevant skills extracted from a comprehensive job posting database and firm resilience by the changes in corporate valuation in response to operational shocks. Evidence suggests that AI generates resilience: An average firm that equips 2.4% of total jobs to be AI-related could approximately recover the full damage of disasters reflected in corporate valuation over a short event window. From the product function test, we find that resilience is attributable to the moderating effect of AI on the damaged input responsiveness under the volatile production environment. Our analyses further reveal a pressing phenomenon: Although underperforming firms could benefit more from an additional unit of AI investment, the realized productivity is notably restrained due to a lack of complementary organizational designs. Our findings provide managerial implications regarding the interplay between environmental conditions and firm investments in both AI technology and complementary infrastructures.

**Understanding and Mitigating the Robot Disadvantage in Luxury Services: The Role of Desire for Superiority** (p. 2134)
Shuang Ma, Lin Ge, He (Michael) Jia, Yonggui Wang

In recent years, the use of service robots in the lodging, dining, and retail sectors has become more prevalent. Service robots offer various benefits, such as reducing operational costs, improving service efficiency, enabling 24-hour support, and ensuring consistent service quality. However, our research indicates that service firms do not uniformly benefit from service robot deployment across all service settings. Specifically, we show that customers' relative preference for service robots over human service providers is lower in luxury services than in mainstream services. We call it the robot disadvantage in luxury services. This disadvantage arises because service robots—as nonhuman service providers—cannot make customers feel that they are being served as a prestigious customer of superior social status over service providers. To address this challenge, we propose two effective strategies for luxury service firms. First, luxury service firms may emphasize the concept of social equality in their communications. Second, luxury service firms may frame the use of service robots as an exclusive privilege. Both approaches can mitigate customers' resistance to robots in luxury services.

**A Study of Ride-Hailing Platforms' Business Models in the Presence of Surge Demand** (p. 2151)
Haiyang Feng, Nan Feng, Ling Zhang, Zhengrui Jiang, Minqiang Li

This study analyzes whether a ride-hailing platform should adopt a closed business model, that is, serving riders only with platform-owned vehicles, or an open business model, that is, allowing private vehicles to utilize the platform to provide services. Analytical results reveal that the open business model leads to more service supply (supply-augmenting effect) and a lower price increase (price discrimination–hindering effect) in the surge period, both benefiting riders. It also reduces the number of idle vehicles in the normal period, thus resulting in more efficient utilization of vehicle resources. Interestingly, the platform does not always benefit from adopting the open business model. In sum, the open business model always benefits the riders, private drivers, and the society as a whole, but it may not be more profitable for the platform. This interesting result calls for policymakers to incentivize the platform so that it adopts the open business model instead of the closed one.

**To Split or to Merge? How Partitioning Affects Consumption and Engagement with Digital Content** (p. 2170)
Heeseung Andrew Lee, Angela Aerry Choi, Wonseok Oh, Tianshu Sun

This study reveals how partitioning serialized e-books into many short episodes (small partitioning (SP)) or fewer but lengthier segments (large partitioning (LP)) shapes reading behavior and economic outcomes. Using actual consumption data, we find that SP fosters higher total word consumption but can reduce overall progression through a title. Conversely, LP boosts completion rates, encourages deeper engagement through annotations and reviews, and spurs subsequent purchases. These positive effects are stronger when a book is popular or high in quality. Our results challenge the conventional view that shorter episodes automatically increase user attention, highlighting the need for a tailored approach. Stakeholders can implement LP for well-regarded or lengthy titles to maximize reader satisfaction and stimulate both reviews and further sales. Smaller partitions, however, might be beneficial for shorter works under a pay-per-word scheme or for certain audience preferences. By aligning partitioning strategies with content characteristics, publishers and platform owners can maintain robust consumer engagement, manage possible drawbacks of segmentation, and support sustainable growth. This research provides clear, actionable insights for digital publishing professionals seeking to optimize user experiences and drive profitability. These findings also encourage further inquiry into how partitioning strategies interact with rapidly evolving consumer preferences and reading technologies.

**Post-Earnings-Announcement Drift Prediction: Leveraging Postevent Investor Responses with Multitask Learning** (p. 2191)
Yu Zhu, Xiao Liu, Olivia R. Liu Sheng

Post-earnings-announcement drift (PEAD) remains one of the most persistent market anomalies, yet traditional

models struggle to predict it effectively. Prior research has relied on *single-task learning* (STL), which treats PEAD prediction as an isolated task, overlooking key *postevent investor responses*—such as analyst forecast revisions and institutional trading—that drive stock price movements. However, incorporating these signals directly as model inputs introduces look-ahead bias, making real-world predictions impractical. Our study proposes a *multitask learning* (MTL) framework that circumvents this issue by treating postevent investor responses as *auxiliary tasks* rather than direct inputs. This enables the model to learn from these critical signals without "cheating." Additionally, we introduce GradPerp, a novel adaptive task weighting method that *prioritizes diverse, meaningful training signals,* further improving predictive performance. A key insight from our research is that leveraging MTL in real-world contexts requires deep domain knowledge and novel designs. More importantly, our MTL framework opens new opportunities for practitioners to enhance deep learning models by *incorporating their financial expertise through carefully chosen auxiliary tasks.* Unlike traditional AI models that rely solely on automated feature selection, our approach provides a structured way for investment professionals to embed domain-driven signals into predictive modeling, unlocking new potential in quantitative finance.

**Irrationality-Aware Human Machine Collaboration: Mitigating Alterfactual Irrationality in Copy Trading** (p. 2213)
Zhe Shen, Wei Jiang, Zhiqiang (Eric) Zheng

Artificial intelligence (AI) algorithms are trained on human-generated data, but what if that data reflects irrational human decision making? To tackle this challenge, Shen et al. developed a new irrationality-aware human-machine collaboration (IA-HMC) framework, designed to help AI recognize and adapt to human irrationality. A key concept introduced in this framework is "alterfactual irrationality"—a term used to describe human decisions influenced by irrelevant alternatives. The researchers applied this idea to copy trading, a popular investment strategy where everyday investors (followers) mimic the trades of expert traders. They identified two major irrational behaviors affecting followers: herding behavior—blindly following others without independent analysis; and identity bias—making investment choices based on who made the trade rather than its actual merit. By developing irrationality-aware machine learning methods, the study showed that AI can help followers make better trading decisions. Their approach led to a 49% improvement in success rates compared to human decisions alone and a 10.2% improvement over previous AI-driven methods. This research presents an innovative approach in human-AI collaboration, showing that for AI to truly align with human needs, it must first learn to account for and correct human irrationality.

**Framing of Seeker-Generated Information and New Solver Participation in Open Innovation Contests: An Empirical Analysis of the Temporal Effects** (p. 2235)
Jiahui Mo, Nila Zhang

In today's highly competitive global market, firms recognize innovation as a key driver of success. To foster innovation, firms (or seekers) increasingly leverage open innovation contests to engage external solvers in addressing challenges. These contests are dynamic, allowing new solvers to enter at any point before the deadline, with their participation influenced by seeker-generated information (*SGI*), including contest descriptions, clarifications to solvers' questions, and solution reviews. When communicating requirements, firms can frame *SGI* by specifying the extent of preference disclosure and defining boundaries for unacceptable design elements. This study examines how *SGI* framing impacts new solver entry and how its effects evolve as the contest progresses. Using a comprehensive data set from an open innovation contest platform, we find that extensive preference disclosure attracts more new solvers daily, with its effect intensifying as the deadline approaches. Additionally, clear and precise rejection criteria initially deter but later encourage solver participation as the deadline nears. When the deadline is distant, overly restrictive requirements can overshadow other important factors, such as solvers' need for exploration, and deter participation. These findings provide actionable insights for firms on how to strategically frame *SGI* at different contest stages to optimize solver participation in open innovation contests.

**Team-Enacted Use vs. Developer-Needed Use of Agile Practices: How Perceptual (In-)Congruence and Team Feedback-Seeking Shape Developer Well-Being** (p. 2253)
Alexander Benlian, Marc Pinski, Martin Adam

Are agile practices always beneficial for developers? Although agile methods, such as daily stand-ups or pair programming, are designed to enhance flexibility and productivity, developers often experience a mismatch between the team-enacted use of these practices and their own individual needs. Using daily survey responses from 149 agile developers (1,510 observations), our study uncovers a crucial yet overlooked factor; perceived (in-)congruence between team-enacted and developer-needed agile practices significantly impacts developer well-being. We show that alignment between the use and need of agile practices is associated with higher well-being, whereas mismatches—both excessive and insufficient agile practices—can be detrimental. Interestingly, we find that frequent team feedback seeking amplifies the negative effects of

misalignment but does not enhance well-being when alignment is achieved. These findings challenge the assumption that the use of agile practices is inherently beneficial and highlight the need for a tailored, developer-centric approach. For organizations, the key takeaway is clear. A "one-size-fits-all" approach to agile practices can backfire. Instead, fostering alignment through adaptive work environments, flexible agile practice use, and targeted interventions can promote sustainable developer well-being and long-term agile team effectiveness.

**Exaptation in Platforms: A Theory of Origins, Mechanisms, and Consequences** (p. 2278)
Amrit Tiwana

Beyond gradual refinement ("adaptation"), IT systems evolve through "exaptation"—the emergence of novel functionality when apps interact with evolving technological and competitive environments. This nine-year study of 625 Android apps across 63 countries shows that real options embedded in app design fuel exaptation, enhancing market performance and user retention. We conceptualize IT exaptation as distinct from adaptation. Whereas adaptation refines, exaptation transforms evolutionary trajectories. Strategic options and operational options enable developers to capitalize on emergent opportunities. For practitioners, this study explains why seemingly similar apps develop divergent trajectories when exposed to identical platform innovations. It suggests that app developers should prioritize embedding flexibility in initial designs—creating not just adaptable systems, but ones with intrinsic leapfrogging potential via exaptation.

**Gains from Product Variety: Evidence from a Large Digital Platform** (p. 2298)
Erik Brynjolfsson, Long Chen, Xijie Gao

Digital platforms can increase product variety and consumer choice by facilitating the discovery and availability of new products. In this paper, we document the massive growth of new products on the largest digital platform in China and quantify the welfare implications for consumers. Using sales data on three categories of books from 2015 and 2019, we find that the number of product titles almost doubles, whereas prices fall somewhat. Most of the new products are niche offerings that exhibit less elastic demand. Accounting for the niche nature of new products generates welfare gains 40 times larger than those from lower prices and 30% higher than existing estimates that do not distinguish between mass and niche products. We also examine the geographic variation in these gains and find that consumers in rural and low-income regions enjoy greater benefits from increased variety. The findings emphasize that expanding niche product offerings may outperform price reductions in generating consumer benefits. Policymakers are encouraged to support e-commerce development in underserved regions to harness the inclusive growth enabled by digital platforms.

**When Top-Down Meets Bottom-Up: Legislative Signals and Online Crowdfunding** (p. 2309)
Anqi Wu, Aravinda Garimella, Ramanath Subramanyam

Over the last two decades, online crowdfunding platforms have facilitated fundraising efforts to alleviate resource shortages across various sectors. As these platforms become increasingly important channels for bottom-up resource mobilization, it is essential to understand their vulnerability to shifts in public perception caused by top-down, off-platform events. We examine the effect of legislative signals on online giving behavior in the context of public education. We analyze how the ratification of the prominent Every Student Succeeds Act (ESSA) impacted giving behavior on a leading education crowdfunding platform and find two contrasting donor tendencies. Although donors contributed more to local initiatives following the ratification, they contributed less to nonlocal initiatives, leading to a substantial net decline in the funds raised. These effects manifest through two mechanisms: information push and information pull. This hyper-localization exacerbates the resource-divide problem in education; schools with students of lower socioeconomic status experienced a sharper decline in the proportion of projects funded. This finding suggests the usefulness of timed and targeted nudges by platform managers, especially around the announcement of major external events. If matching processes are designed carefully, online platforms can mitigate, at least to some extent, the consequences of education inequity.

**Predicting Instructor Performance in Online Education: An Interpretable Hierarchical Transformer with Contextual Attention** (p. 2327)
Wen Wang, Mi Zhou, Beibei Li, Honglei Zhuang

Online education is a vital consumer industry that is undergoing rapid technological change. This paper develops a deep learning model to predict instructor performance on online education platforms from a content-based perspective. Specifically, we design an interpretable hierarchical transformer with contextual attention to predict instructor rating and course rating using textual data. Our model captures the inherent hierarchical structure of online courses and the sequential dependency of lectures within a course. Moreover, it goes beyond prediction and enables interpretability analysis, which provides additional insights into potential reasons behind the predictions made by the model. Extensive experiments demonstrate that our model outperforms classic machine learning models as well as state-of-the-art deep learning models. Furthermore, we conduct in-depth interpretability analysis to explore what factors might predict the

success of an online course at the lecture, sentence, and word level. We also showcase the value and applicability of our model through a randomized experiment. Our findings and methods provide managerial implications for instructors and online education platforms to improve course creation and delivery in this vitally important emerging market.

**Ephemeral State-Dependent Recommendation for Digital Content** (p. 2344)
Lanfei Shi, Jin Liu, Yongjun Li, Natasha Zhang Foutz

Building upon recent advances in consumption theories, we propose an ephemeral state-dependent framework for digital content recommendations. The framework accentuates a critical, yet understudied, interplay between a firm's recommendation *strategy* (assimilation or diversification) and a consumer's *ephemeral state* (fixation or foraging). The framework adaptively recommends either assimilated or diversified content based on a consumer's ephemeral state. Through a randomized field experiment, we provide compelling evidence that *state-dependent* schemes can enhance engagement and revenue. Although the *congruent* scheme (i.e., assimilation when fixation, diversification when foraging) generally outperforms the *incongruent* one, contributing a 7.3% ($19.73 million) annual revenue lift for the platform, our findings underscore the necessity for nuanced personalization. Specifically, consumers with broader, more fluid preferences benefit more from the *incongruent* scheme (i.e., assimilation when foraging, diversification when fixation), challenging the prevailing assumption that congruence is always optimal. Our research not only adds to the theoretical understanding of consumer behavior in digital content consumption, but also offers actionable insights for designing more effective recommender systems. By accounting for consumer heterogeneity and considering the broader implications of our proposed recommendation framework, including spillover effects, our findings have the potential to influence industry practices and future academic inquiry in this rapidly evolving field.

**Artificial Intelligence (AI) Assistant in Online Shopping: A Randomized Field Experiment on a Livestream Selling Platform** (p. 2358)
Lingli Wang, Ni Huang, Yumei He, De Liu, Xunhua Guo, Yan Sun, Guoqing Chen

Livestream selling is an innovative form of online shopping that supports real-time interactions between streamers and consumers. However, a key challenge remains: Streamers have limited capacity to answer individual inquiries, whereas shoppers expect fast, personalized responses. This study investigates whether an AI-powered streaming assistant can address this tension by providing interactive, chat-based support to help consumers access and process information. Through a large-scale randomized field experiment on a leading livestream selling platform, we find that the AI assistant increases sales by 3.00% and reduces product return rates by 12.55%. Our analysis suggests that the AI assistant helps consumers feel more informed and confident in their purchases, thereby reducing uncertainty. At the same time, the AI assistant can occasionally disrupt the consumers' livestream experience. Overall, the benefits of uncertainty reduction outweigh the negative influence of interruptions. For platform managers and policymakers, these findings evidence the potential of AI technology to enhance online commerce. The AI assistant is particularly effective for high-uncertainty products and for streamers with large audiences, offering implications for strategic deployment. Our research provides actionable insights for integrating AI into livestream selling and other digital commerce scenarios where real-time, AI-powered support can facilitate both consumer satisfaction and business growth.

**Social Media and Political Affiliation: How Expressing Hot-Button Opinions Affects Raters' Assessments of Job Applicants** (p. 2375)
Teagen Nabity-Grover, Kevin D. Matthews, Philip L. Roth, Youngjin Kwon, Jason B. Thatcher

People frequently post their views on hot-button issues, such as immigration, on social media platforms on which the content is often publicly accessible to broad audiences. One of these audiences is hiring managers as more than 50% of organizations report viewing job applicants' accounts as part of social media screening during the hiring process. Our research shows that applicants' social media activity on hot-button issues can significantly impact their hireability. In a study of 377 professionals with hiring experience, we find that viewing applicants' social media posts about immigration influenced judgments of similarity, likability, and two dimensions of hireability: expected task performance and expected organizational citizenship behaviors. Surprisingly, the hot-button posts outweighed job-relevant information regardless of where (Facebook or Twitter) the content was viewed; this indicates hiring managers generalized their impressions across social media platforms. A post hoc qualitative applicability check with 79 hiring managers confirmed these findings. Our study highlights the risks of using hedonic social media for screening job applicants, emphasizing the need for organizations to be mindful of how personal disclosures online can unintentionally influence hiring decisions.

**Strategic Best-Response Fairness Framework for Fair Machine Learning** (p. 2391)
Hajime Shimao, Warut Khern-Am-Nuai, Karthik Kannan, Maxime C. Cohen

This study introduces a framework called "strategic best-response fairness" (SBR-fairness) to address discrimination

perpetuated by machine-learning (ML) algorithms. It challenges the conventional focus on fairness solely in prediction results, arguing that this approach ignores how individuals affected by the predictions may alter their behavior in response to algorithmic decisions. The framework considers whether an algorithm, trained on potentially biased data, leads to identical equilibrium behaviors across different subpopulations that are ex ante identical. The study finds that common fair-ML algorithms, such as those relying on color-blindness and demographic parity fairness criteria, do not always achieve SBR fairness. This means that they may not eliminate disparities in effort and outcomes. Equalized odds (EO), however, have been shown to achieve SBR fairness, but they suffer from several practical limitations. The study proposes that SBR fairness is a necessary condition for breaking cycles of discrimination in ML. It also argues that SBR fairness offers a complementary way to assess other fairness criteria and understand behavioral responses. The findings suggest a need for policy and practical focus on designing SBR-fair algorithms that promote equitable outcomes at both the prediction and behavioral level.