Likelihood of Questioning AI-Based Recommendations Due to Perceived Racial/Gender Bias

Carlos M. Parra, Manjul Gupta, and Denis Dennehy

Abstract—Advances in artificial intelligence (AI) are giving rise to a multitude of AI-embedded technologies that are increasingly impacting all aspects of modern society. Yet, there is a paucity of rigorous research that advances understanding of when, and which type of, individuals are more likely to question AI-based recommendations due to perceived racial and gender bias. This study, which is part of a larger research stream contributes to knowledge by using a scenario-based survey that was issued to a sample of 387 U.S. participants. The findings suggest that considering perceived racial and gender bias, human resource (HR) recruitment and financial product/service procurement scenarios exhibit a higher questioning likelihood. Meanwhile, the healthcare scenario presents the lowest questioning likelihood. Furthermore, in the context of this study, U.S. participants tend to be more susceptible to questioning AI-based recommendations due to perceived racial bias rather than gender bias.

Index Terms—AI-based recommendations, artificial intelligence (AI), gender bias, racial bias.

I. INTRODUCTION

ARTIFICIAL intelligence (AI)-embedded digital technologies continue to permeate and reshape our daily working and social lives, with household names such as Netflix using machine-learning algorithms to analyze billions of records to suggest films that users might like based on previous reactions and choices of films [1]. The hype surrounding AI is breath taking, with International Data Corporation (IDC) predicting that the global market will exceed U.S. $500 billion by 2024 [2]. This hype has fueled an expectation that "over the next decade, AI will not replace managers, but managers who use AI will replace those who do not" [3, p. 20]. Even though, by some estimates, AI would eliminate 75 million jobs, it would also help create 133 million new emerging roles, more adapted to a new division of labor between AI and humans [4].

It is widely accepted that inclusive policies aimed at promoting and advancing diversity are important catalyzers of innovation and economic development [5]. However, private and public sector organizations adopting AI-embedded technologies that use AI-based recommendations, may be perpetuating and exacerbating society's implicit biases. Not only imposing limits on the advancement and cultivation of diversity and inclusion but perhaps even forestalling it.

Unfortunately, this is not a new phenomenon, as it was first recognized in the 1980s when the United Kingdom's Commission for Racial Equality found that a British medical school was using a computer program to help identify applicants to be interviewed that was biased against women and minority applicants [6]. This is especially troubling when it occurs in various important realms of daily life. For example, in the realm of criminal justice and crime prevention, an algorithm in Broward County (Florida) used to identify convicted criminals likely to reoffend produced a double number of false positives for African Americans than for whites. In essence, it incorrectly identified black defendants as "high risk" at twice the rate it labeled white defendants as such [7]. In the realm of hiring and recruitment, Amazon stopped using a human resource (HR) algorithm that showed high-paying jobs to men much more than to women, as it favored applicants' resumes containing words that were more commonly found on men's [8]. In the context of healthcare, an algorithm used to predict healthcare needs was found to exhibit significant racial discrimination as it considered health expenses and costs as a proxy for healthcare needs, without having first controlled for evident inequalities in access to healthcare services [9].

Although the existence of algorithmic bias (and of subsequent biased AI-based recommendations) has been acknowledged, extant research has just started to systematically investigate the conditions under which individuals might question AI-based recommendations due to perceived racial/gender bias. A recent study [10] examined which national cultural values espoused by individuals, might lead them to question AI-based recommendations when perceived as exhibiting racial or gender biases and reported that individuals who espouse national cultural values associated with collectivism, masculinity, and uncertainty avoidance are more likely to question biased AI-based recommendations [10]. As our study builds upon extant literature, it contributes to the tradition of accumulative building of knowledge as we consider the same scenarios utilized in the previous study. To this end, the aim of this study is to "examine when or under which daily life circumstances individuals would be (on average) more likely to question biased AI-based recommendations."

Carlos M. Parra and Manjul Gupta are with the Information Systems and Business Analytics (ISBA) Department, Florida International University, Miami, FL 33199 USA.

Denis Dennehy is with the Business Information Systems Department, the National University of Ireland Galway, Galway, H91WN80 Ireland (e-mail: denis.dennehy@nuigalway.ie).

The remainder of this article is structured as follows. First, a synthesis of the extant literature is presented. Next, the research methodology is outlined. Then, the results are presented, followed by a discussion. This article ends with limitations, future research, and conclusions.

II. LITERATURE REVIEW

Keeping in mind that "scholars have been posing critically important questions regarding ubiquitous computing since at least the last quarter century" [10, p. 3]. During the past decade, AI-based recommendations have given academics additional justification for their misgivings. For instance, geographers have expressed doubts about the adequacy of land use/land cover classification models [11]. Similarly, doctors have raised concerns about the accuracy of AI-based diagnostic tools [12]. All of which, understandably, has led lawyers to probe the laws and policies that should govern automation in general [13].

In addition, automated decision making, by minimizing human involvement, encounters, and interactions, limit the sense of otherness, and of common humanity, essential to strengthening the social fabric of communities [14]. In turn, arguably, making it easier to depersonalize (and even to vilify) out-groups, for instance, by means of social-media-facilitated interactions riddled with polarizing recursive performative affordances. All of which, may lead to associated worldview gaps that strain social ties [15]. Alternatively, it may be the case that academics and advocates of AI and AI-based recommender systems need to be aware of algorithmic bias as such instances are being discussed beyond traditional media and academic journals, such as social media.

Regardless of the reasons why the topic at hand may be top of mind to readers, we do subscribe to the notion that information systems in general, and AI in particular, ought to be used for the economic and public good. Indeed, AI can play an essential role in helping achieve the United Nations Sustainable Development Goals [16]. Specifically, as mentioned previously, AI ought to help advance an institutions' diversity and inclusion efforts and not obstruct.

III. METHODS

It is considering the above that, here, we aim at helping elucidate which settings would make biased AI recommendations more noticeable to a sample of U.S. participants, by comparing the average likelihood of questioning biased AI-based recommendations in different contexts. Furthermore, we do this, on the one hand, to complement findings related to which type of individuals are more likely to question AI-based recommendations when perceived as biased due to race or gender. On the other hand, to help set a prioritized agenda for tackling instances in which AI may be inhibiting the advancement of diversity and inclusion (by helping perpetuate societies' biases). Thus, in this section, we describe our participants, as well as the scenarios considered.

A. Participants

A sample of U.S. participants was recruited using Amazon Mechanical Turk (MTurk). Mturk was chosen because it has been proven to be an effective data collection tool in many fields, producing results equivalent to studies performed in laboratory settings [17]. The sample of 387 U.S. participants included 237 male and 150 female respondents. The mean age of participants was 38 years. Participants received U.S. $1.50 for completing the survey. Survey responses that failed attention checks were omitted from the analysis.

B. Survey Scenarios

The survey was designed to gauge individuals' likelihood of questioning biased AI-based recommendations in seven different scenarios, namely: 1) security screening (e.g., airport immigration kiosks); 2) HR recruitment; 3) blocking user-generated online content; 4) procurement of financial products/services; 5) booking hotel online; 6) booking travel online; and 7) healthcare (see Table I). The seven scenarios used in this study are based on examples of racially and gender-biased AI-based recommendations that occurred in real life [10]. References to academic research and news items involving biased AI-based recommendations in the seven scenarios considered are also listed in Table I. In addition, the readability and clarity of these seven scenarios were calibrated using a pilot survey of 60 participants.

Before answering questions concerning *racial bias*, participants first read the heading "Let us imagine that you and a friend have the same gender, age, as well as practically identical educational and professional achievements but have a different race. How likely are you to question the following outcomes?" They then were asked to rate on a 5-point Likert-type scale where 1 meant Highly Unlikely, and 5 meant Highly Likely the likelihood of them questioning the AI-based outcome in each of the seven scenarios. Similarly, before answering questions concerning *gender bias*, participants first read the heading "Let us imagine that you and a friend have the same race, age, as well as practically identical educational and professional achievements but have a different gender. How likely are you to question the following outcomes?" And, once again, participants were asked to rate the likelihood of them questioning the AI-based outcome in each scenario.

IV. RESULTS

The influence that individual characteristics (associated to espoused national cultural values) may exert on the likelihood of questioning biased AI-based recommendations has been elucidated elsewhere [10]. Here, we gauge (on average) how likely a sample of 387 U.S. participants would be to question biased AI-based recommendations in different scenarios. Fig. 1 shows that on average our sample of 387 U.S. participants was more likely to question AI-based recommendations due to perceived racial bias when scenarios involved security screenings or HR recruitment situations.

TABLE I
AI-BASED RECOMMENDATION SCENARIOS EXHIBITING
GENDER/RACIAL BIAS

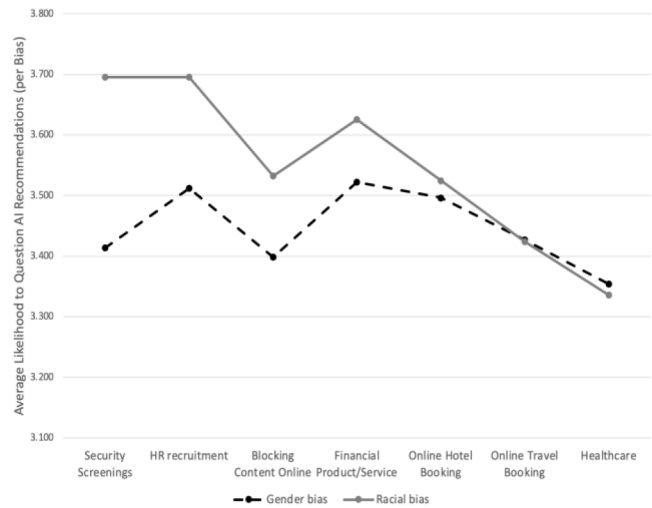| Scenario (related sources)	How likely are you to question the following outcomes? (1 = Highly Unlikely; 5 = Highly Likely)
Bias in financial services [18,19]	You are both applying for the same financial product (such as a credit card or home mortgage/loan) on the same bank app/website using your own devices. You notice the products that are offered to your friend charge higher interest rates than those offered to you.
Bias in recruitment [20, 21]	You are both looking for similar jobs on the same employment app/website using your own devices. You notice the jobs that are offered to your friend usually have lower-paying salaries than those offered to you.
Bias in facial recognition software used to screen travelers [22]	You both have the same nationality and are at the airport going through the same automated immigration kiosk that uses face recognition technology to verify travelers' identity. The automated immigration kiosk directs your friend to see an immigration officer while you are cleared to go through.
Bias in online hotel bookings [23]	You are both booking a similar hotel room using the same hotel booking app/website using your own devices. Hotel rooms offered to your friend have higher prices than those offered to you.
Bias in online hotel bookings [23]	You are both booking the same flight using the same travel booking app/website on your own devices. Flights offered to your friend are costlier than those offered to you.
Bias in blocking online content [24]	You both regularly write posts on similar topics on the same social network service (for instance, Facebook). Your friend's posts are found objectionable (that is, flagged for removal) more often by the social network service than those posted by you.
Bias in healthcare [25,26]	You both have similar diets, daily routines, and are feeling just fine. You are both using an automated health assessment app on your own devices that involves interacting and answering questions using voice recognition. The automated health assessment app suggests that your friend is at a higher risk of contracting the flu, but not you.

Fig. 1. Survey respondents' likelihood of questioning AI-based recommendations due to perceived gender/racial bias in seven scenarios.

TABLE II
MEANS AND PAIRED MEAN DIFFERENCES IN LIKELIHOOD OF
QUESTIONING AI-BASED RECOMMENDATIONS DUE TO PERCEIVED
GENDER/RACIAL BIAS IN SEVEN SCENARIOS

Scenario	Average likelihood to question due to perceived **gender bias**	Average likelihood to question due to perceived **racial bias**	Paired means differences (gender-racial)	Paired means difference (p-value)
Security Screening	3.413	3.695	-0.282	0.00
HR Recruitment	3.512	3.695	-0.183	0.00
Blocking User-Generated Online Content	3.398	3.532	-0.134	0.02
Financial Product/ Service	3.522	3.625	-0.103	0.10
Online Hotel Booking	3.496	3.525	-0.028	Not significant
Online Travel Booking	3.426	3.424	0.003	Not significant
Healthcare	3.354	3.336	0.018	Not significant

While survey respondents were least likely to question AI-based recommendations due to perceived racial bias in healthcare settings. Regarding gender bias, survey respondents were most likely to question AI-based recommendations when scenarios entailed procuring financial products/services and HR recruitment, and least likely to do so in a healthcare scenario. Thus, considering perceived racial and gender bias, healthcare situations offered the lowest questioning likelihood, while HR recruitment ones exhibited higher questioning likelihood.

We also conducted paired *t*-tests to compare the likelihood of questioning AI-based recommendations due to perceived racial to that of perceived gender bias per scenario. In Table II, mean differences for security screenings and HR recruitment scenarios are statistically significant (with $p < 0.01$), also for when user-generated online content is blocked (with $p < 0.05$), as well as for when individuals procure financial products/services (with $p \leq 0.10$). This indicates that when both gender and racial bias are perceived at the same time in these four scenarios, our sample of 387 U.S. participants appeared to be more susceptible to questioning AI-based recommendations due to perceived racial bias than gender bias. For the remaining three scenarios (i.e., when individuals are booking travel/lodging services or in health care settings), since the difference is not statistically significant, both racial and gender bias are equally likely to trigger questioning of AI-based recommendations.

V. DISCUSSION

From a theoretical perspective, biased AI-based recommendations (emerging out of unintentionally ill-conceived biased algorithms), provides additional justification for having misgivings about ubiquitous computing, specifically, the detrimental effects that the perpetuation of societies' biases, by means of AI-embedded digital technologies, may have on the promotion of diversity and inclusion should be explored and quantified. These adverse impacts could include the inadvertent attenuation of efforts aimed at mitigating risks associated with misguided/biased viewpoint reinforcement, and associated worldview gaps.

From a practice perspective, it is evident that algorithmic bias can arise in the day-to-day use of AI-embedded technologies or other mediating technologies. For example, HR recruitment is a business function in which organizations have been increasingly employing AI-based recommendations for screening candidates [27]. It would, however, not be counterintuitive for potential employees, to feel discouraged, disenfranchised, and perhaps even betrayed if they perceived that the AI systems used by organizations they wish to work for, were biased against them because of their race or gender. Such situations could be perceived as corporate infractions that could in turn initiate negative word-of-mouth transmission chains as well as associated protests against the organizations in question, and reputational damage [28].

Tradeoffs between failure to adopt AI-embedded technologies to augment organizational capabilities and exposure to adverse situations in which individuals (i.e., users) or groups (i.e., consumers, employees, suppliers, etc.) may feel discriminated against can be managed. For example, by implementing algorithmic fairness initiatives that should be reported on (and clearly communicated) by developers of AI-based recommender systems, as well as by organizations adopting them. Specific adjustments could entail modifying algorithmic learning objectives such that special treatment is afforded to protected groups or incorporating mechanisms that help compensate for outcome inequities between protected and unprotected groups. In addition, managers deciding to adopt AI-based recommender systems should be aware of the possibility that these may lead to biased outcomes. And thus, these managers should make special efforts to understand why and how their AI-embedded technologies could end up helping perpetuate and exacerbate society's implicit biases. These possibilities and understandings should then be communicated and explained to their employees (as well as to other stakeholders), along with relevant algorithmic fairness measures being implemented to help mitigate associated risks.

Critically, investigating the conditions under which individuals might question AI-based recommendations due to perceived racial and gender bias, could help multilateral organizations, governments, as well as private enterprises set a prioritized agenda for addressing situations in which AI-embedded technologies may be more noticeably forestalling their diversity and inclusion efforts. Specifically, mitigating possible racial bias should be the focus of organizations and individuals involved in security screenings, HR recruitment, and the procurement/facilitation of financial products/services.

Similarly, gender bias mitigation should be prioritized by organizations/individuals working on HR recruitment, as well as the procurement/facilitation of financial products/services.

VI. LIMITATIONS AND FUTURE RESEARCH

As with all research, this study has limitations, which also offer directions for future research. For instance, it seems puzzling that while booking online hotel/travel and in healthcare scenarios, study participants might be equally susceptible to questioning AI-based recommendations due to racial and gender bias. Similarly, psychological/behavioral or perhaps even sociological reasons for why, during security screenings, our participants could be more susceptible to questioning AI-based recommendations due to racial bias than gender bias should be studied. Insofar as this outcome might be the consequence of a uniquely racially charged juncture in the U.S.

In addition, we did not include a situational outcome associated with a racially biased criminal justice scenario (e.g., sentencing recommender systems). However, considering the social unrest and upheaval that have transpired throughout the U.S. with protests against racial injustice, it would not be counter-intuitive to assume that such scenario would have grouped along with security screenings and HR recruitment and exhibited a high likelihood of being questioned. In addition, the questioning of AI-based recommendations due to perceived gender bias in HR recruitment situations and while individuals procure financial products/services, when considered together, may be pointing to the need to explore equal pay for equal work issues in different industries. However, it may be challenging to succinctly portray a situational outcome associated with an AI-related wage gap for co-workers of the same race, age, as well as practically identical educational and professional achievements but who have different gender. Future research could include more scenarios and explore how individuals with different personalities react to biased AI-based recommendations. Finally, as all participants in this study live in the U.S., future research may consider including participants from other countries to broaden the generalizability of the study findings.

VII. CONCLUSION

The study aimed to gauge the likelihood of individuals-based in the U.S., to question biased AI-based recommendations in seven different scenarios. This study highlights the importance of private and public sector organizations not blindly adopting potentially biased AI-based recommendations. Not just to avoid damaging news coverage but perhaps also associated class action ligation. Implementing awareness training on biased AI-based recommendations, in contexts, such as security screenings, HR recruitment, and financial product/service offerings would help promote algorithmic fairness, which in turn would advance and cultivate a diverse and inclusive American society.

The findings have implications for strengthening fairness in AI-based recommender systems, the use of which has become ubiquitous in nearly all business functions. Finally, research on the people side of AI-based recommender systems and

how organizations can unknowingly exacerbate the unintended consequences of such systems, as well as proactively mitigating such risks, requires greater scrutiny by the academic community.

REFERENCES

[1] L. Columbus, *Roundup of Machine Learning Forecasts and Market Estimates, 2020*, Forbes, Jersey City, NJ, USA, Jan. 2020. [Online]. Available: https://www.forbes.com/sites/louiscolumbus/2020/01/19/roundup-of-machine-learning-forecasts-and-market-estimates-2020/?sh=386f4ddf5c02

[2] *Worldwide Semiannual Artificial Intelligence Tracker*, Int. Data Corp. (IDC), Needham, MA, USA, Feb 2021.

[3] E. Brynjolfsson and A. Mcafee, *The Business of Artificial Intelligence, What it Can—And Cannot—Do for an Organization*, Harvard Bus. Rev., Brighton, MA, USA, 2017.

[4] *The Future of Jobs Report 2018*, World Econ. Forum, Geneva, Switzerland, 2018.

[5] C. R. Østergaard, B. Timmermans, and K. Kristinsson, "Does a different view create something new? The effect of employee diversity on innovation," *Res. Policy*, vol. 40, no. 3, pp. 500–509, 2011.

[6] J. Manyika, J. Silberg, and S. Presten, *What Do We Do About the Biases in AI?* vol. 25, Harvard Bus. Rev., Brighton, MA, USA, 2019.

[7] J. Angwin, J. Larson, S. Mattu, and L. Kirchner, *Machine Bias*, ProPublica, New York, NY, USA, May 2016. Accessed: Apr. 20, 2019. [Online]. Available: https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

[8] I. A. Hamilton, *Why It's Totally Unsurprising That Amazon's Recruitment AI Was Biased Against Women*, Bus. Insider, New York, NY, USA, Oct. 2018. Accessed: Apr. 20, 2019. [Online]. Available: https://www.businessinsider.com/amazon-ai-biased-against-women-no-surprise-sandra-wachter-2018-10

[9] Z. Obermeyer, B. Powers, C. Vogeli, and S. Mullainathan, "Dissecting racial bias in an algorithm used to manage the health of populations," *Science*, vol. 366, pp. 447–453, Oct. 2019.

[10] M. Gupta, C. M. Parra, and D. Dennehy, "Questioning racial and gender bias in AI-based recommendations: Do espoused national cultural values matter?" *Inf. Syst. Front*, to be published.

[11] Y. Grinblat, M. Gilichinsky, and I. Benenson, "Cellular automata modeling of land-use/land-cover dynamics: Questioning the reliability of data sources and classification methods," *Ann. Amer. Assoc. Geograph.*, vol. 106, no. 6, pp. 1299–1320, 2016.

[12] A. Jain, B. Khanna, R. Dubey, and S. Agarwal, "A comprehensive study of artificial intelligence-based medical diagnosis," presented at the IEEE Int. Conf. Innovat. Technol. (INOCON), Bangluru, India, 2020, pp. 1–4.

[13] W. Hartzog, "On questioning automation," *Cumberland Law Rev.*, vol. 48, no. 1, pp. 1–8, 2017.

[14] A. Barker, A. Crawford, N. Booth, and D. Churchill, "Everyday encounters with difference in urban parks: Forging 'openness to otherness' in segmenting cities," *Int. J. Law Context*. vol. 15, no. 4, pp. 495–514, 2019.

[15] C. M. Parra, M. Gupta, and P. Mikalef, "Information and communication technologies (ICT)-enabled severe moral communities and how the (Covid19) pandemic might bring new ones," *Int. J. Inf. Manage.*. vol. 57, Apr. 2021, Art. no. 102271.

[16] A. Di Vaio, R. Palladino, R. Hassan, and O. Escobar, "Artificial intelligence and business models in the sustainable development goals perspective: A systematic literature review," *J. Bus. Res.*, vol. 121, pp. 283–314, Dec. 2020.

[17] G. Paolacci, J. Chandler, and P. G. Ipeirotis, "Running experiments on Amazon mechanical Turk," *Judg. Decis. Making*, vol. 5, no. 5, pp. 411–419, 2010.

[18] E. Hamilton, *AI Perpetuating Human Bias in the Lending Space*, Tech Times, New York, NY, USA, 2019. [Online]. Available: https://www.techtimes.com/articles/240769/20190402/ai-perpetuating-human-bias-in-the-lending-space.htm

[19] A. Fuster, P. Goldsmith-Pinkham, T. Ramadorai, and A. Walther, *Predictably Unequal? The Effects of Machine Learning on Credit Markets*, CEPR, Washington, DC, USA, 2017.

[20] J. Dastin, *Amazon Scraps Secret AI Recruiting Tool That Showed Bias Against Women*, Guardian, London, U.K., 2018. [Online]. Available: https://www.theguardian.com/technology/2018/oct/10/amazon-hiring-ai-gender-bias-recruiting-engine

[21] A. Lambrecht and C. E. Tucker, "Algorithmic bias? an empirical study of apparent gender-based discrimination in the display of STEM career ads," *Manage. Sci.*, vol. 65, no. 7, pp. 2966–2981, 2019.

[22] C. Tate, *Racial Bias in Facial Recognition Software: What Travelers Should Know as TSA, CBP Expand Programs*, USA Today, McLean, VA, USA, 2019. [Online]. Available: https://www.usatoday.com/story/travel/news/2019/12/30/facial-recognition-racial-bias-traveler-info-tsa-cbp-expand-use/2755613001/

[23] A. Hannak, G. Soeller, D. Lazer, A. Mislove, and C. Wilson, "Measuring price discrimination and steering on E-commerce Web sites," presented at the Conf. Internet Meas. Conf., 2014, pp. 305–318.

[24] S. Ghaffary, *The Algorithms That Detect Hate Speech Online Are Biased Against Black People*, Vox Media, Inc., Washington, DC, USA, 2019. [Online]. Available: https://www.vox.com/recode/2019/8/15/20806384/social-media-hate-speech-bias-black-african-american-facebook-twitter

[25] D. Gershgorn, *If AI Is Going to Be the World's Doctor, it Needs Better Textbooks*, Quartz, New York, NY, USA, 2018. [Online]. Available: https://qz.com/1367177/if-ai-is-going-to-be-the-worlds-doctor-it-needs-better-textbooks/

[26] Z. Obermeyer and S. Mullainathan, "Dissecting racial bias in an algorithm that guides health decisions for 70 million people," presented at the Conf. Fairness Accountability Transparency, 2019, p. 89.

[27] S. Bogle and S. Sankaranarayanan, "Job search system in android environment—Application of intelligent agents," *Int. J. Inf. Sci. Techn.*, vol. 2, no. 3, pp. 1–17, 2021.

[28] S. Romani, S. Grappi, and R. P. Bagozzi, "My anger is your gain, my contempt your loss: Explaining consumer responses to corporate wrongdoing," *Psychol. Marketing*, vol. 30, no. 12, pp. 1029–1042, 2013.