Fall 12-1-2023

Robot umpire vs. human umpire: The spectators' perception of algorithm errors in baseball games

Sunyoung Choi
Korea Advanced Institute of Science and Technology, South Korea, sunyung97@kaist.ac.kr

Jooyoung Kim
Korea Advanced Institute of Science and Technology, South Korea, jyk1095@kaist.ac.kr

Donghwan Kwon
Korea Advanced Institute of Science and Technology, South Korea, dh_kwon@kaist.ac.kr

Gimok Yun
Korea Advanced Institute of Science and Technology, South Korea, ygm5239@kaist.ac.kr

Hangjung Zo
Korea Advanced Institute of Science and Technology, South Korea, joezo@kaist.edu

Choi, S.Y., Kim, J.Y., Kwon, D.H., Yun, G.M., & Zo, H.J. (2023). Robot umpire vs. Human umpire: The spectators' perception of algorithm errors in baseball games. In Li, E.Y. *et al.* (Eds.) *Proceedings of The International Conference on Electronic Business, Volume 23* (pp. 744-749). ICEB'23, Chiayi, Taiwan, October 17-23, 2023

Robot Umpire vs. Human Umpire: The Spectators' Perception of Algorithm Errors in Baseball Games

Sunyoung Choi [1]
Jooyoung Kim [2]
Donghwan Kwon [3]
Gimok Yun [4]
Hangjung Zo [5,*]

*Corresponding author
[1] Ph.D. Student, Korea Advanced Institute of Science and Technology, South Korea, sunyung97@kaist.ac.kr
[2] Ph.D. Student, Korea Advanced Institute of Science and Technology, South Korea, jyk1095@kaist.ac.kr
[3] Master Student, Korea Advanced Institute of Science and Technology, South Korea, dh_kwon@kaist.ac.kr
[4] Master Student, Korea Advanced Institute of Science and Technology, South Korea, ygm5239@kaist.ac.kr
[5] Professor, Korea Advanced Institute of Science and Technology, South Korea, joezo@kaist.edu

ABSTRACT

Advancements in technology have allowed for robot umpires in baseball, replacing human umpires to minimize errors in judgment: strike/ ball decision. Yet, when watching games with robot umpires in pilot operations, there are some instances where the robot umpire also commits erroneous calls. Hence, this study would conduct a scenario-based online experiment to investigate how spectators perceive robot umpires in baseball games depending on the subject of error (robot umpire vs. human umpire). The experimental results would indicate that people demonstrate a lower degree of trust, attitude, and call credibility on robot umpires. Furthermore, the effects of robot umpires would be moderated by the familiarity of the robot umpire and by the phase of error (early vs. late in the game). These findings would contribute to the understanding of the spectator's perception of new technologies replacing humans and have practical implications for sports industry practitioners.

Keywords: Baseball, Robot umpire, Algorithm error, Serial position effect, Familiarity.

INTRODUCTION

In an effort to minimize errors in judgment, sports are increasingly embracing new technologies. Sports officials are charged with the important responsibility of accurately and fairly overseeing high-level sports competitions (Mascarenhas, Collins, & Mortimer, 2005). However, the decisions of umpiring may be impacted by various contextual dimensions, such as crowd noise, home advantage, a team's aggressive reputation, or physical appearance (Jones, Paul, & Erskine, 2002; Nevill, Balmer, & Williams, 2002; Unkelbach & Memmert, 2010; Van Quaquebeke & Giessner, 2010).

The baseball league has been testing robot umpires, starting with the lower leagues, to replace human umpires who make incorrect judgments in strike/ball decision. In the year 2019, the Atlantic League in the United States became the first to implement the usage of robot umpires, employing a trackman system to make ball/strike decisions. Additionally, the Triple-A league in the United States plans to implement two distinct types of robot umpires during the 2023 season. One of half the games will feature all calls made by a robot umpire, while the other half will utilize a challenge system, enabling teams to challenge up to three calls per game. Following these lower leagues' pilot operation, Major League Baseball (MLB) officially announced the plan to adopt robot umpires in calling balls and strikes in 2024.

A robot umpire is one of the algorithms defined as "a set of encoded instructions for converting input data into a desired output based on predetermined calculations (Gillespie, 2014, p. 1)". According to the phenomena referred to as 'algorithm aversion', people frequently refuse and criticize algorithms despite the fact algorithms outperform human decision-making (Castelo, Bos, & Lehmann, 2019; Dietvorst, Simmons, & Massey, C., 2015; Jussupow, Benbasat, & Heinzl, 2020; Longoni, Bonezzi, & Morewedge, 2019). In contrast, algorithm appreciation, as demonstrated in computer science research, refers to the participants showing more agreement with participants showing more agreement with the same argument when it is presented as coming from an "expert system" rather than a "human" (Dijkstra, Liebrand, & Timminga, 1998).

Thus, this article aims to provide further insights into resolving conflicting results within the realm of baseball. Furthermore, the purpose of the article is to direct attention towards the perception of algorithm error (vs. human error), rather than solely engaging in comparisons between algorithms and human performance. Therefore, this research would examine how spectators perceive umpires in baseball games depending on the subject of an error: robot umpire's error vs. human umpire's error. The evaluations of umpires would be measured by trust, fairness, and other behavioral outcomes.

Another interesting inquiry is how people perceive a robot umpire's error (vs. a human error) according to the phase of error. According to the serial position effect, an umpire's error occurring early in the decision-making process (primacy effect) would have a greater negative impact than the error that occurs later (recency effect) (Manzey, Reichenbach, & Onnasch, 2012). Therefore, this research is planning to conduct 2 (algorithm error, human error) x 2 (early phase, late phase of error occurrence) between-subjects design on an online survey platform.

In addition, this research would explore the influence of familiarity with robot umpires on the evaluation of robot umpires. Familiarity is referred to as an individual's exposure to an algorithm carrying out a certain activity (Castelo, Bos, & Lehmann, 2019). Hence, this research expects the evaluation of robot umpires would be moderated by familiarity with robot umpires.

This research would make theoretical and practical contributions. First, this study would extend both sports and Human-Robot Interaction literature by analyzing how baseball spectators perceive robot umpires from a new perspective through algorithmic error. Second, this finding would also suggest that the managers of sports officials understand how to ensure that spectators' perceptions are positive as the sports industry introduces new technologies to replace humans.

LITERATURE REVIEW

The Introduction of Robot Umpire in Baseball

An umpire is essential because they enforce the rules of the game by deciding on match play in accordance with the rules of the sport. In baseball, an umpire's primary role is to determine whether a pitch thrown by a pitcher is in or out of the strike zone; According to MLB, the official strike zone is defined as the area over home plate between the midpoint of a batter's shoulders and the top of the uniform pants, when the batter is in his stance and ready to swing at a pitched ball, and a point just below the kneecap. However, there are frequent erroneous calls made by human umpires when determining whether the pitch is a strike or a ball. In the 2018 season, for instance, an average of 14 incorrect calls was made by human umpires per game (Hille, 2019).

Sports officials have a vital duty to impartially and accurately supervise elite sporting competitions (Mascarenhas, Collins, & Mortimer, 2005). Therefore, sports are progressively utilizing new technology in an effort to reduce human error in judgment. For example, Hawk-Eye was unveiled in 2001 as the first decision-aid technology; It was a complex ball trajectory system that was mostly used in cricket and tennis to assist umpires in making decisions. Likewise, baseball is set to adopt robot umpires in the 2024 season for strike/ball determination to minimize errors in judgment.

Algorithm Aversion and Algorithm Appreciation

An algorithm is defined as "a set of encoded instructions for converting input data into a desired output based on predetermined calculations (Gillespie, 2014, p. 1)". As technology evolves, the constraints of human judgment and decision-making may be "cognitively treated" by algorithms (Burton et al., 2020). In other words, algorithms are being utilized to complement or replace people in many industries including sports. Despite the fact that algorithms frequently exceed human decision-making, people frequently oppose and criticize algorithms (Castelo, Bos, & Lehmann, 2019; Dietvorst, Simmons, & Massey, 2015; Jussupow, Benbasat, & Heinzl, 2020; Longoni, Bonezzi, & Morewedge, 2019). Dietvorst, Simmons, and Massey (2015) refer to this phenomenon as algorithm aversion.

While the majority of empirical data shows algorithm aversion, some research has discovered an appreciation for algorithmic guidance or even an excessive dependence on AI-based solutions (Dijkstra, Liebrand, & Timminga, 1998; Dijkstra, 1999; Logg, Minson, & Moore, 2019; Wagner, Borenstein, & Howard, 2018). For example, algorithmic financial advice tends to receive greater attention from individuals compared to crowdsourced advice, as it is generally preferred by people (Gunaratne, Zalmanson, & Nov, 2018). Depending on the nature of the task in which the algorithm is used, people may show an algorithm aversion or an algorithm appreciation. Therefore, the purpose of this article is to contribute to untangling these contradictory findings in the context of baseball games.

HYPOTHESES DEVELOPMENT

Although the introduction of robot umpires in the sports industry has been mentioned due to human error, the robot umpires themselves also demonstrate instances of making mistakes during the trial operation. Error in decision-making refers to making a judgment that ultimately proves to be less accurate than intended, which is typical for judgments made in unclear situations. According to Dietvorst, Simmons, and Massey (2015), people are less forgiving of erring algorithms than they are of erring humans because they ignore the notion that algorithms might improve and learn from their flaws. Therefore, if people identify an algorithm's error, their reliability of the algorithm's advice tends to diminish. This phenomenon may be attributed to the presence of unfulfilled performance expectations (Dzindolet et al. 2002; Prahl and Van Swol 2017). Therefore, this research proposes how people perceive a robot umpire's error (vs. human error) in the context of calling strike/ball decisions in a baseball game. Hence, this research hypothesizes the following.

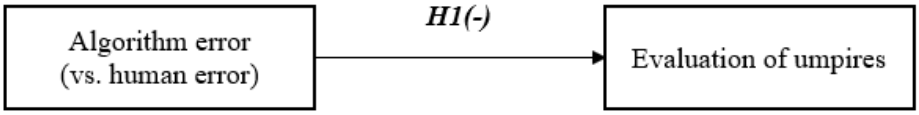

Figure 1: A research model of the negative effect of algorithm error on the evaluation of umpires (H1)

Hypothesis 1. Spectators would evaluate robot umpires more negatively compared to human umpires when they saw a robot umpire's error (vs. human error)

Serial position effect is commonly identified as an individual recall pattern: Items shown first or last in a sequence have a recall advantage over items provided in the middle, which is known as a primacy or recency effect, respectively (Glanzer & Cunitz, 1966; Murdock, 1962); Furthermore, the primacy effect is stronger than the recency effect (Worchel et al., 1988). The identification of the decision-making stage upon the discovery of an algorithmic error is imperative due to its evident significance. Hence, the degree of algorithm aversion differs based on whether the error happens early or late in the decision-making process. As a result, an algorithmic error occurring early in the decision-making process (primacy effect) has a greater negative impact than the error that occurs later (recency effect) (Manzey, Reichenbach, & Onnasch, 2012). This study would display a robot umpire's error (vs. a human umpire's error) in the second and eighth innings of a baseball game. The aim of this study is to investigate the influence of the error phase on algorithm aversion, concerning errors made by either a robot umpire or a human umpire. Specifically, this study seeks to determine how the extent of algorithm aversion is impacted by the phase of the error (early vs late). Therefore, this research proposes the following.

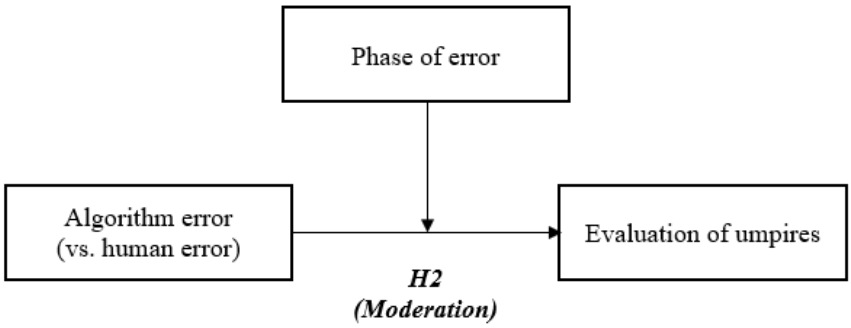

Figure 2: A research model of the moderating effect of the phase of an error (H2)

Hypothesis 2. The effect of the type of umpire error on spectators' evaluation of an umpire would be more negative when they see an umpire error in the early phase of a baseball game

Familiarity refers to individuals' exposure to an algorithm carrying out a certain activity (Castelo, Bos, & Lehmann, 2019). Previous psychological research (e.g. Monin, 2003) has found strong evidence for the beneficial relationship between familiarity and one's emotional reactions to an object since it fosters pleasant emotions toward the object (Garcia-Marques, 2000; Patterson and Mattila, 2008). According to previous research, people who have received training in statistical methods and algorithms are more likely to employ them (Araujo et al., 2020; Önkal, Gönül, & De Baets, 2019). Additionally, algorithm aversion can also be lessened via practical algorithm practice in simulated contexts (Li, Rau, & Huang, 2020). Existing literature has demonstrated that such familiarity will ultimately result in one's favorable attitude since humans frequently judge based on their affection (Monin, 2003; Patterson and Mattila, 2008). Therefore, this research aims to investigate the applicability of the aforementioned findings within the context of a baseball game, even if the robot umpire commits errors. Hence, this study hypothesizes the following.

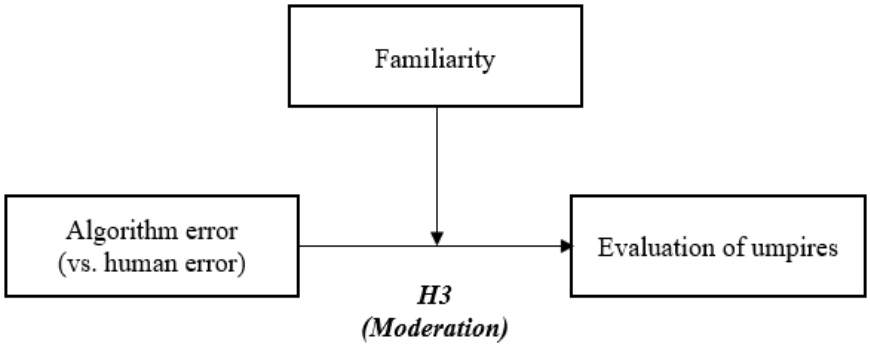

Figure 3: A research model of the moderating effect of familiarity (H3)

Hypothesis 3. The effect of the type of umpire error on spectators' evaluation of an umpire would be lessened by the familiarity of robot umpire

METHODS

Participants & Procedure

A total of 200 participants would be recruited from the survey platform. In this study, both the scenarios and the survey questions are written in English. Therefore, the participant pool would be restricted to participants in the United States whose first language is English. Firstly, the participants read the basic information about baseball and the process of strike and ball determination executed by an umpire. They also read a scenario in which a robot umpire or a human umpire makes an on-field error in a strike/ball call. After that, participants would be asked about their evaluation of the robot umpire (vs. human umpire) by using these factors: fairness, trust, attitude, call credibility, and expertise. Finally, they would answer questions about the baseball viewing experience and demographic questions. To prevent their inattention and improve the validity of response, Catcha and one attention check question would be included in the experiment material (Aguinis, Villamor, & Ramani, 2021).

The participants would be randomly assigned to a 2 x 2 between-subject design. The first two-level factor is a subject of an on-field error in a strike/ball call. Half of the participants would be assigned to a match where a robot umpire makes an error in a call. The other half of the participants would be assigned to a match where a human umpire makes an error in a call.

The second two-level factor is the phase of error occurrence: early phase vs. late phase. Half of the participants would read a scenario about a game in which the robot umpire (or human umpire) makes an error in a call early in the game. There are nine regular innings in a baseball game and it is divided into 3 innings. The early phase of a game would be set as the 2nd inning, which is the middle of the 1st and 3rd innings. The other half of the participants would read a scenario about a game in which the robot umpire (or a human umpire) makes an error in a call late in the game.

Measures

This study would use fairness, trust, attitude, call credibility, and expertise as the evaluation of an umpire in a baseball game. This research would measure fairness (Waddell, 2018), trust (Johnson and Grayson's research, 2005), call credibility (Roberts, 2010), and expertise (Ohanian, 1990) using four-item scale (1= "strongly disagree", 7 = "strongly agree"). This research would measure the attitude toward using an umpire randomly assigned to each participant (Heerink et al., 2010) using three-item scale (1= "strongly disagree", 7 = "strongly agree"). In this study, familiarity would be employed as a moderator variable. Specifically, the degree of familiarity that participants had with a robot umpire in the context of baseball would be assessed using a three-item scale (1= "strongly disagree", 7 = "strongly agree"). The measurement items are shown in Table 1.

Variable name	Items
Fairness	
Fairness1	If the umpire makes the call, then the call was done unbiased
Fairness2	If the umpire makes the call, the call was error-free
Fairness3	If the umpire makes the call, then the call was done accurately
Fairness4	If the umpire makes the call, then the call was done objectively
Trust	
Trust1	I trust the umpire
Trust2	I can rely on the umpire
Trust3	I have no doubt about the umpire
Trust4	I don't need to be cautious about the umpire
Attitude	
Attitude1	I think it is good to use an umpire to determine a strike/ball in baseball
Attitude2	The use of an umpire is beneficial for the judgment of a strike/ball
Attitude3	In my opinion, the use of an umpire in baseball will have a positive impact
Call credibility	
Call credibility1	The umpire's call is accurate
Call credibility2	The umpire's call is believable
Call credibility3	The umpire's call is fair
Call credibility4	The umpire's call is trustworthy
Expertise	
Expertise1	The umpire's call is expert
Expertise2	The umpire's call is knowledgeable
Expertise3	The umpire's call is qualified
Expertise4	The umpire's call is skilled
Familiarity	
Familiarity1	I am familiar with the concept of a robot umpire in baseball
Familiarity 2	I am familiar with the actions and decisions of a human umpire in person
Familiarity 3	I am familiar with the use of technology in baseball umpiring, such as automated strike zones

Table1: Measurement items of evaluation of umpire and familiarity

EXPECTED OUTCOME

In order to evaluate hypothesis 1, this research would conduct an analysis using t-test methodology. The expected results of the analysis would indicate that trust in umpires, attitude, call credibility, expertise, and fairness would be different depending on the agent of error. For each measurement, the t-test would indicate that the evaluation of the robot umpire will be more negative than the human umpire.

Additionally, in order to evaluate hypothesis 2, Hayes's PROCESS macro (Model 1) would be used to test the moderating effects of the phase of the error. The moderating effect of the phase of error would be significant for all measurement factors: trust, attitude, call credibility, expertise, and fairness. Furthermore, Hayes's PROCESS macro (Model 1) was employed to test the moderating effects of familiarity with the robot umpire (Hypothesis 3). The results of the analysis would indicate that trust, attitude, call credibility, expertise, and fairness would be significantly moderated by the familiarity of the robot umpire.

Additionally, this research would use Johnson-Neyman (J-N) technique (Johnson & Fay, 1950; Johnson & Neyman, 1936; Potthof, 1964) to investigate significant interactions. In contrast to the pick-a-point method, Johnson-Neyman (J-N) technique has been used to compute a "region of significance" for the simple slope of a focused predictor based on the value of the continuous moderator. After identifying the moderating effect on each measurement, this research would examine the effect of algorithmic error * the phase of the error and the effect of algorithmic error * Familiarity on each dependent variable: trust, attitude, call credibility, and expertise by a moderator.

EXPECTED DISCUSSION

Although research on algorithm aversion was extensively tested in various domains, there exists a dearth of studies examining this phenomenon in the context of the sports industry. The findings of the research would reveal the presence of algorithm aversion in robot umpires in baseball games. Despite the errors of judgment, participants would perceive higher levels of trust, attitude, call credibility, and expertise in human umpires. Moreover, participants would maintain a positive attitude toward the utilization of human umpires rather than robot umpires in the context of baseball games. These expected outcomes would highlight the significance of considering algorithm aversion and its implication for the introduction of new technologies, such as robot umpires, into the sports industry. Furthermore, the findings of this study would hold potential implications for decisionmakers involved in the consideration of robot umpires as a replacement for human umpires in the realm of baseball.

Moreover, this study is planning to explore whether spectators' algorithm aversion toward robot umpires is influenced by the phase of error during a baseball game. Hence, this research would investigate the relative impact of the primary effect compared to the recency effect based on the serial position effect. Additionally, this paper would examine whether familiarity with robot umpires influences the level of algorithm aversion. The expected results would demonstrate a significant effect on factors such as trust, attitude, call credibility, expertise, and fairness. If familiarity with a robot umpire is employed as a moderator to reduce algorithm aversion for a robot umpire, the findings would emphasize the significance of familiarity to the spectators' perception of the robot umpire. As the implementation of robot umpires is planned for 2024 in the MLB league, it would be imperative for baseball league authorities to continue educating their audience about these technological innovations.

Ultimately, this research would be critical for enhancing spectators' perceptions of the introduction of new technologies. And for this to happen, sports industry stakeholders would need to proactively engage in effective promotion well in advance of introducing these new technologies, thereby facilitating a process of demystification. Hence, the result of this research would contribute to fostering a favorable reception and understanding among spectators.

ACKNOWLEDGMENT

This research was supported by the MSIT (Ministry of Science and ICT), Republic of Korea, under the ITRC (Information Technology Research Center), Republic of Korea, support program (IITP-2020-0-01787) supervised by the IITP (Institute of Information & Communications Technology Planning & Evaluation), Republic of Korea.

REFERENCES

Aguinis, H., Villamor, I., & Ramani, R. S. (2021). MTurk research: Review and recommendations. *Journal of Management, 47*(4), 823-837. https://doi.org/10.1177/0149206320969787

Araujo, T., Helberger, N., Kruikemeier, S., & De Vreese, C. H. (2020). In AI we trust? Perceptions about automated decision-making by artificial intelligence. *AI & Society, 35*, 611-623. https://doi.org/10.1007/s00146-019-00931-w

Burton, J. W., Stein, M. K., & Jensen, T. B. (2020). A systematic review of algorithm aversion in augmented decision making. *Journal of Behavioral Decision Making, 33*(2), 220-239. https://doi.org/10.1002/bdm.2155

Castelo, N., Bos, M. W., & Lehmann, D. R. (2019). Task-dependent algorithm aversion. *Journal of Marketing Research, 56*(5), 809-825. https://doi.org/10.1177/0022243719851788

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General, 144*(1), 114. https://doi.org/10.1037/xge0000033

Dijkstra, J. J. (1999). User agreement with incorrect expert system advice. *Behaviour & Information Technology, 18*(6), 399-411. https://doi.org/10.1080/014492999118832

Dijkstra, J. J., Liebrand, W. B., & Timminga, E. (1998). Persuasiveness of expert systems. *Behaviour & Information Technology, 17*(3), 155-163. https://doi.org/10.1080/014492998119526.

Dzindolet MT, Pierce LG, Beck HP, Dawe LA (2002) The perceived utility of human and automated aids in a visual detection task. *Human Factors, 44*(1), 79–94. https://doi.org/10.1518/0018720024494856

Garcia-Marques, T. (2000). The Positive Feeling of Familiarity: Mood as an Information Processing Regulation mechanism. In *The message within: The role of subjective experience in social cognition and behavior* (pp. 240). Psychology Press, Philadelphia.

Gillespie, T. (2014). The relevance of algorithms. *Media Technologies: Essays on Communication, Materiality, and Society*, *167*(2014), 167. https://doi.org/10.7551/mitpress/9780262525374.003.0009

Glanzer, M., & Cunitz, A. R. (1966). Two storage mechanisms in free recall. *Journal of Verbal Learning and Verbal Behavior*, *5*(4), 351-360. https://doi.org/10.1016/S0022-5371(66)80044-0

Gunaratne, J., Zalmanson, L., & Nov, O. (2018). The persuasive power of algorithmic and crowdsourced advice. *Journal of Management Information Systems*, *35*(4), 1092-1120. https://doi.org/10.1080/07421222.2018.1523534

Heerink M, Kröse B, Evers V, Wielinga B (2010) Assessing acceptance of assistive social agent technology by older adults: the Almere model. *International Journal of Social Robotics*. doi:10.1007/s12369-010-0068-5

Hille, B. (2019). MLB umpires missed 34,294 ball-strike calls in 2019, study shows. Retrieved from: https://www.sportingnews.com/us/mlb/news/mlb-umpires-missed-34294-ball-strike-calls-2018-studyshows/ieco5qe561lj1w2cmv3accu5w

Johnson, D., & Grayson, K. (2005). Cognitive and affective trust in service relationships. *Journal of Business Research*, *58*(4), 500-507. https://doi.org/10.1016/S0148-2963(03)00140-1

Johnson, P. O., & Fay, L. C. (1950). The Johnson-Neyman technique, its theory and application. *Psychometrika*, *15*(4), 349-367.

Johnson, P. O., & Neyman, J. (1936). Tests of certain linear hypotheses and their application to some educational problems. *Statistical Research Memoirs*.

Jones, M. V., Paull, G. C., & Erskine, J. (2002). The impact of a team's aggressive reputation on the decisions of association football referees. *Journal of Sports Sciences*, *20*(12), 991–1000. https://doi.org/10.1080/026404102321011751

Jussupow, E., Benbasat, I., & Heinzl, A. (2020). Why are we averse towards algorithms? A comprehensive literature review on algorithm aversion. Retrieved from https://aisel.aisnet.org/ecis2020_rp/168

Logg, J. M., Minson, J. A., & Moore, D. A. (2019). Algorithm appreciation: People prefer algorithmic to human judgment. *Organizational Behavior and Human Decision Processes*, *151*, 90-103. https://doi.org/10.1016/j.obhdp.2018.12.005

Longoni, C., Bonezzi, A., & Morewedge, C. K. (2019). Resistance to medical artificial intelligence. *Journal of Consumer Research*, *46*(4), 629-650. https://doi.org/10.1093/jcr/ucz013

Li, Z., Rau, P. L. P., & Huang, D. (2020). Who should provide clothing recommendation services: artificial intelligence or human experts?. *Journal of Information Technology Research (JITR)*, *13*(3), 113-125. https://doi.org/10.4018/JITR.2020070107

Manzey, D., Reichenbach, J., & Onnasch, L. (2012). Human performance consequences of automated decision aids: The impact of degree of automation and system experience. *Journal of Cognitive Engineering and Decision Making*, *6*(1), 57-87. https://doi.org/10.1177/1555343411433844

Mascarenhas, D. R., Collins, D., & Mortimer, P. (2005). Elite refereeing performance: Developing a model for sport science support. *The Sport Psychologist*, *19*(4), 364–379. https://doi.org/10.1123/tsp.19.4.364

Monin, B. (2003). The warm glow heuristic: When liking leads to familiarity. *Journal of Personality and Social Psychology*, *85*(6), 1035. https://doi.org/10.1037/0022-3514.85.6.1035.

Murdock, B. B. (1962). The serial position effect of free recall. *Journal of Experimental Psychology*, *64*(5), 482–488. http://dx.doi.org/10.1037/h0045106

Nevill, A. M., Balmer, N. J., & Williams, A. M. (2002). The influence of crowd noise and experience upon refereeing decisions in football. *Psychology of Sport and Exercise*, *3*(4), 261–272. https://doi.org/10.1016/S1469-0292(01)00033-4

Önkal, D., Gönül, M. S., & De Baets, S. (2019). Trusting forecasts. *Futures & Foresight Science*, *1*(3-4), e19. https://doi.org/10.1002/FFO2.19

Patterson, P. G., & Mattila, A. S. (2008). An examination of the impact of cultural orientation and familiarity in service encounter evaluations. *International Journal of Service Industry Management*, *19*(5), 662–681. https://doi.org/10.1108/09564230810903514.

Potthoff, R. F. (1964). On the Johnson-Neyman technique and some extensions thereof. *Psychometrika*, *29*(3), 241-256.

Prahl, A., & Van Swol, L. (2017). Understanding algorithm aversion: When is advice from automation discounted?. *Journal of Forecasting*, *36*(6), 691-702. https://doi.org/10.1002/for.2464

Roberts, C. (2010). Correlations among variables in message and messenger credibility scales. *American Behavioral Scientist*, *54*(1), 43-56. DOI: 10.1177/0002764210376310

Unkelbach, C., & Memmert, D. (2010). Crowd noise as a cue in referee decisions contributes to the home advantage. *Journal of Sport and Exercise Psychology*, *32*(4), 483–498. https://doi.org/10.1123/jsep.32.4.483

Van Quaquebeke, N., & Giessner, S. R. (2010). How embodied cognitions affect judgments: Height-related attribution bias in football foul calls. *Journal of Sport and Exercise Psychology*, *32*(1), 3–22. https://doi.org/10.1123/jsep.32.1.3

Wagner, A. R., Borenstein, J., & Howard, A. (2018). Overtrust in the robotic age. *Communications of the ACM*, *61*(9), 22-24. https://doi.org/10.1145/3241365