Judgment and Decision Making (2025), **vol. 20**:e22 1–22
doi:10.1017/jdm.2025.8

EMPIRICAL ARTICLE

Using conventional framing to offset bias against algorithmic errors

Hamza Tariq, Jonathan A. Fugelsang, and Derek J. Koehler

Department of Psychology, University of Waterloo, Waterloo, ON, Canada

Corresponding author: Hamza Tariq; Email: h33tariq@uwaterloo.ca

Received: 7 November 2024; **Revised:** 13 March 2025; **Accepted:** 13 March 2025

Keywords: algorithm aversion; artificial intelligence; error judgments; status quo; conventionality

Abstract

Prior research has shown that people judge algorithmic errors more harshly than identical mistakes made by humans—a bias known as algorithm aversion. We explored this phenomenon across two studies (N = 1199), focusing on the often-overlooked role of conventionality when comparing human versus algorithmic errors by introducing a simple conventionality intervention. Our findings revealed significant algorithm aversion when participants were informed that the decisions described in the experimental scenarios were conventionally made by humans. However, when participants were told that the same decisions were conventionally made by algorithms, the bias was significantly reduced—or even completely offset. This intervention had a particularly strong influence on participants' recommendations of which decision-maker should be used in the future—even revealing a bias against human error makers when algorithms were framed as the conventional choice. These results suggest that the existing status quo plays an important role in shaping people's judgments of mistakes in human–algorithm comparisons.

Public significance statement

This research highlights the importance of considering conventionality when evaluating people's biases in judging mistakes. In particular, these studies help explain why people might judge errors from alternative options, like algorithmic technologies, more critically compared to more conventional, human alternatives.

1. Introduction

a) *Everything that's already in the world when you're born is just normal;*
b) *anything that gets invented between then and before you turn 30 is incredibly exciting and with any luck you can make a career out of it;*
c) *anything that gets invented after you're 30 is the end of civilization as we know it until it's been around for about 10 years when it gradually turns out to be alright really.*

—Douglas Adams (1999), *A Hitchhiker's Guide to the Internet*

Throughout human history, the ‘normal’ or conventional ways of doing things have always evolved with the advent of new inventions. Manual scribing gave way to the printing press, magnetic compasses replaced celestial navigation, steam engines revolutionized how we built and traveled, and pocket calculators transformed complex manual calculations. What might have initially been perceived as ‘against the natural order of things’ almost always—over time—turned out, as Douglas Adams put it 25 years ago, to be “alright really”.

Over recent decades, algorithms have moved from novelty to near ubiquity in many domains: medicine, navigation, finance, criminal justice, and more (Rainie and Anderson, 2017). Despite their demonstrated strengths—algorithms frequently outperform humans in forecasting and decision-making tasks (Brzezicki et al., 2020; Dawes, 1979; Dawes et al., 1989; Meehl, 1954)—research has observed a persistent human distrust of algorithms, which has been reported as far back as the 1950s (Dietvorst et al., 2015; Eastwood et al., 2012; Jussupow et al., 2020; Meehl, 1954; Önkal et al., 2009; Promberger and Baron, 2006; Shaffer et al., 2013). This bias against algorithms, often called algorithm aversion, appears in various contexts. In a series of five studies, Dietvorst et al. (2015) asked participants to predict outcomes from real data, such as MBA students’ success and the number of airline passengers departing from US states. Participants first observed predictive forecasts made by either an algorithm or a human (some participants saw both or neither as controls), and then chose which forecaster’s predictions to rely on for future tasks. According to the authors, their results showed that seeing an algorithm err sharply reduced participants’ confidence in it. Consequently, participants who had witnessed the algorithm’s mistakes became much less likely to choose it over a human forecaster going forward—even when that human was objectively inferior, and the algorithm had actually outperformed the human overall. The authors concluded that algorithm aversion is at least partly driven by people’s experience with the algorithm—particularly when they witness it making mistakes.

But what is it about algorithms that causes this mistrust? It has been speculated that: people expect algorithms to be near perfect and free of biases; algorithms are unable to learn from experience unlike human decision-makers; algorithms would continue to make systematic errors; algorithms may not consider important qualitative information or unique circumstances; that many modern algorithmic systems (i.e., artificial intelligence) are black boxes; and algorithms might be dehumanizing or unethical to use for important decisions (Bigman and Gray, 2018; Bonezzi et al., 2022; Bonezzi and Ostinelli, 2021; Carabantes, 2020; Castelo et al., 2019; Dawes, 1979; Einhorn, 1986; Grove and Meehl, 1996; Highhouse, 2008; Longoni et al., 2019).

However, these explanations often overlook a critical contextual factor: Which option—human or algorithm—is seen as the convention in a particular setting. We propose that algorithm aversion may partly reflect a more general bias against whichever agent is the ‘alternate’ choice, rather than the convention or status quo. Our goal with this work has been to move beyond the properties of the algorithm itself and explore the role of context—an aspect we feel has not been sufficiently addressed in current theories. We posit that as the context of our interactions with algorithms evolves, many of the stated reasons for algorithm aversion may become less relevant. If humans are usually viewed as the established, conventional decision-makers, then an algorithm’s mistakes may be judged more harshly simply because it lacks that conventional status. By contrast, when an algorithm is regarded as the norm—common, longstanding, and integral to the system—algorithmic errors in that setting might be perceived as less severe, and human errors might appear more suspect.

1.1. Conventionality and alternate aversion

By ‘conventional,’ we refer to the default or status quo options that are trusted and accepted within a given context or domain. Our interpretation of ‘conventionality’ draws on the status quo bias (Samuelson and Zeckhauser, 1988), omission bias (Baron and Ritov, 2004; Ritov and Baron, 1992), default options in choice architecture (Johnson and Goldstein, 2003; Johnson et al., 2012), endowment effects, and loss aversion (Dinner et al., 2011; Kahneman et al., 1991; Thaler, 1980). Omission bias suggests that people are more forgiving of errors that arise from inaction. Thus, a mistake made by the conventional option might be viewed more leniently since it is a choice that would be made most of the time anyway—similar to an error of inaction. However, choosing a non-conventional option and then seeing it make a mistake might be seen as an error resulting from action and, therefore, less forgivable. Choice architecture research shows that presenting options as defaults can influence perceptions and decisions in their favor. Hence, when the conventional option is seen as the default choice, identical errors by the 'non-default' or alternate option could be judged as relatively worse.

Additionally, there might be a tendency to overestimate the importance of the conventional option to the entire system and how disruptive and costly it might be to replace it. This bias might be considered a form of 'reference dependence' in the default effects literature (Dinner et al., 2011). The default option acts as a reference point, influencing how alternative or non-default options are evaluated as gains or losses. Dinner et al. (2011) suggested that defaults could function as 'instant endowments', meaning decision-makers may perceive themselves as having already chosen the default option—and use that as their reference point. This aligns with the endowment effect, where people value an object more once they own it (Thaler, 1980). This effect can increase the perceived value of the default option and activate loss aversion—where the pain of losing something is much more powerful than the pleasure of gaining something of equal value (Tversky and Kahneman, 1991). The bias in the evaluation of conventional and alternative options may result from this interplay of reference dependence and loss aversion (Kahneman et al., 1991). Drawing on this literature and people's changing relationships with algorithms (Logg, 2022; Logg et al., 2019), we hypothesize that alternate aversion occurs when the presence of a conventional option leads people to have a stronger aversion to the non-conventional or alternate option. Simultaneously, when a decision-making option is established as the convention or the status quo, it is often judged less severely, even for identical mistakes.

2. The present research

We conducted two studies to examine whether framing a human or algorithmic decision-maker as the conventional option influences judgments and decisions regarding identical errors and whether this affects the phenomenon of algorithm aversion. Specifically, we predicted that (1) errors made by non-conventional decision-makers would be perceived as more severe and concerning; and (2) despite making identical mistakes, conventional decision-makers would be retained and recommended for future tasks more frequently than non-conventional ones. The conventional status of the error maker (i.e., whether it was framed as conventional or alternative) was varied across both studies. In Study 1, participants observed human and algorithmic decision-makers making errors while screening applicants in a college admissions scenario. In Study 2, participants saw decision-makers making errors while evaluating the quality of sound speakers in a product quality control scenario.

2.1. Setting up conventionality

The so-called conventional or status quo framing of the decision-maker was set up using three characteristics: historic use, prevalence, and perceived system dependence.

2.1.1. Historic use

The conventional decision-maker—human or algorithmic—was described as having been in the role for the past 10 years. This history may carry potential implications: A long tenure might indicate familiarity, stability, reliability, and effectiveness over time. Similar to the status quo bias (Samuelson and Zeckhauser, 1988), it might suggest a pre-existing familiarity and trust in the decision-maker, in contrast to the unfamiliarity and uncertainty associated with a newer decision-maker who lacks a historical track record.

https://doi.org/10.1017/jdm.2025.8 Published online by Cambridge University Press

2.1.2. Prevalence

The conventional decision-making option was described as the most commonly used and widespread approach in similar contexts, such that 85% of operators or the industry used the conventional option. We assume that people are likely to trust existing methods that are widely recognized and adopted as standard practice. As such, widespread use might indicate general social acceptance and collective endorsement of the decision-maker by a majority. This reinforces its position as the conventional method and stands in contrast to an alternative option that is not in common use.

2.1.3. System Dependence

The conventional decision-maker was framed as being essential to the continued functioning of the system of which it is part. Because the stability or survival of the conventional option is seen as vital, people may tend to be more forgiving of its errors. In this context, the conventional method serves as the default reference point, making individuals reluctant to penalize it, while the alternate option is viewed as more expendable. Consequently, harsh penalties for the conventional option may be perceived as more costly and a threat to the overall system function, whereas similar penalties for the alternate option pose little risk.

3. Study 1

In Study 1, we created a scenario where both human and algorithmic decision-makers made errors while screening college applicants, focusing specifically on prediction-based errors. These errors occurred when either human admissions officers or admissions algorithms failed to accurately predict future outcomes (college performance) based on available information (the admissions application). We utilized the three ‘convention’ characteristics—history, prevalence, and system dependence—to frame which decision-making agent was the conventional option and which was the alternate in each condition (Figure 1).

3.1. Method

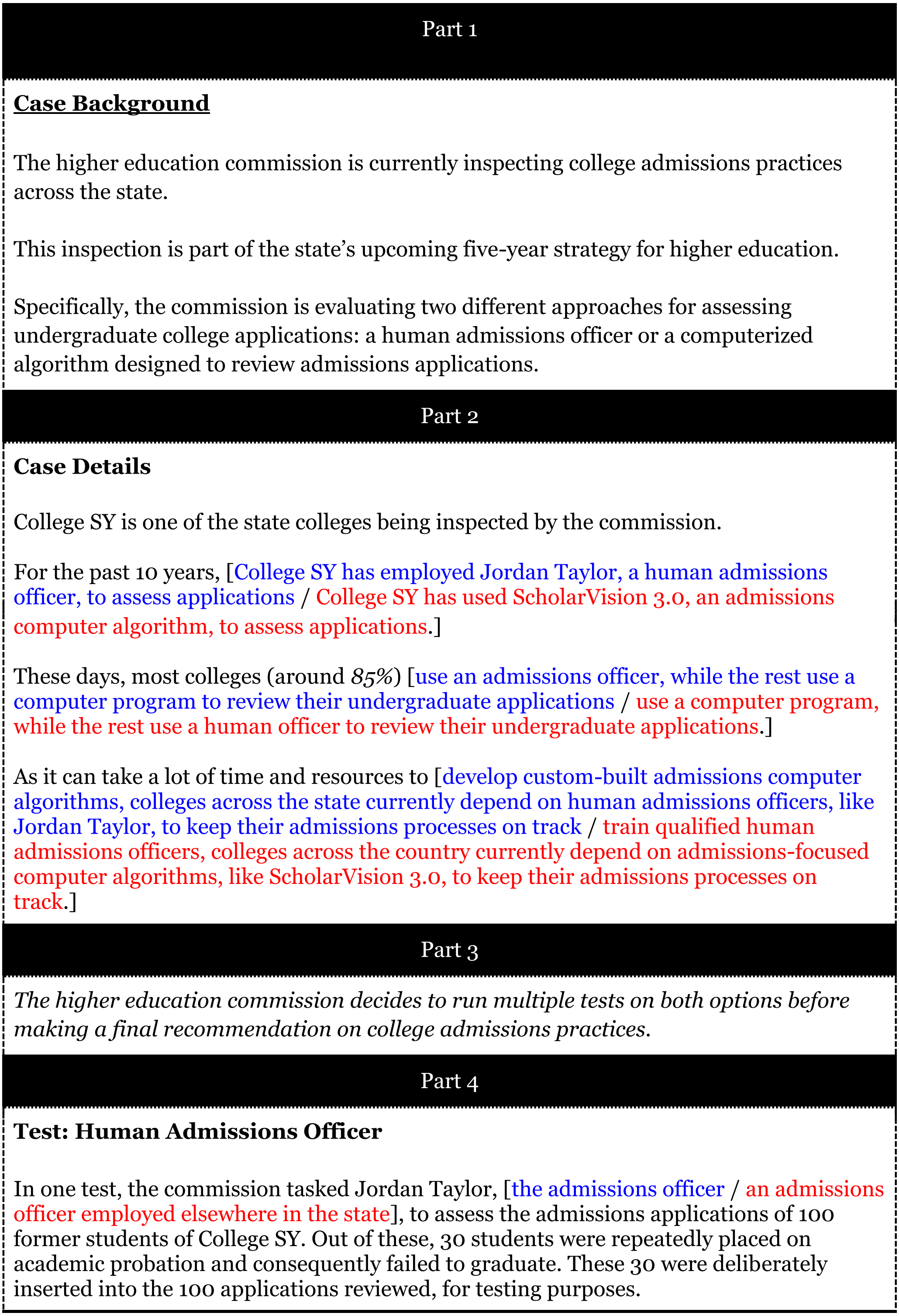
Data and materials for this experiment are available on the Open Science Framework (https://osf.io/29sj7/). This study was not preregistered. We recruited participants for this study (N = 594; 308 women, 279 men, 6 non-binary, and 1 other; $Mdn_{\text{age group}}$ = 35–44) through CloudResearch’s MTurk Toolkit. Participants were recruited simultaneously for both studies in this paper but could only participate in one. The recruitment criteria included: US residency; completion of at least 100 prior HITs; an HIT approval rating of 99% or higher; no participation in an earlier pilot study of this project; and two comprehension checks before starting the study.[1]

3.1.1. Procedure

Participants were instructed to review a case study before being asked to make their judgments. The fictitious case study in this experiment involved college admissions. All participants were briefed on the scenario where a state higher education commission was evaluating admissions practices. The review focused on the performance of human admission officers versus computerized admissions algorithms in processing applications.

Participants were randomly assigned to one of four conditions in a 2 × 2 between-subjects design. First, they learned that either a human officer or a computer algorithm was the conventional method for reviewing admissions applications at a hypothetical college. Next, the commission described a test where 100 former student applications were reviewed—30 of which were deliberately inserted from students who had been repeatedly placed on academic probation and ultimately failed to graduate. The

[1] We included two pre-study comprehension check screening questions. Due to a programming error, participants were permitted to correct their answers to these checks even if they initially responded incorrectly.



Part 1

Case Background

The higher education commission is currently inspecting college admissions practices across the state.

This inspection is part of the state's upcoming five-year strategy for higher education.

Specifically, the commission is evaluating two different approaches for assessing undergraduate college applications: a human admissions officer or a computerized algorithm designed to review admissions applications.

Part 2

Case Details

College SY is one of the state colleges being inspected by the commission.

For the past 10 years, [College SY has employed Jordan Taylor, a human admissions officer, to assess applications / College SY has used ScholarVision 3.0, an admissions computer algorithm, to assess applications.]

These days, most colleges (around *85%*) [use an admissions officer, while the rest use a computer program to review their undergraduate applications / use a computer program, while the rest use a human officer to review their undergraduate applications.]

As it can take a lot of time and resources to [develop custom-built admissions computer algorithms, colleges across the state currently depend on human admissions officers, like Jordan Taylor, to keep their admissions processes on track / train qualified human admissions officers, colleges across the country currently depend on admissions-focused computer algorithms, like ScholarVision 3.0, to keep their admissions processes on track.]

Part 3

The higher education commission decides to run multiple tests on both options before making a final recommendation on college admissions practices.

Part 4

Test: Human Admissions Officer

In one test, the commission tasked Jordan Taylor, [the admissions officer / an admissions officer employed elsewhere in the state], to assess the admissions applications of 100 former students of College SY. Out of these, 30 students were repeatedly placed on academic probation and consequently failed to graduate. These 30 were deliberately inserted into the 100 applications reviewed, for testing purposes.

Figure 1. *College admissions scenario as presented to participants.*

Note: This includes all scenario parts of Study 1 in sequence, as seen by the participants, barring the comprehension questions. Colored text within square brackets shows variations made per condition (blue = human convention and red = algorithm convention).

https://doi.org/10.1017/jdm.2025.8 Published online by Cambridge University Press

Taylor was then asked to **pick the 30 students from the list of 100** who were unlikely to graduate based on their expected academic performance in college.

When the test results were reviewed, the commission found that Taylor had ***failed to include 16 of the 30 students who could not graduate*** *due to poor academic results.*

OR

Test: Computerized Admissions Algorithm

In one test, the commission tasked ScholarVision 3.0, [an admissions algorithm used elsewhere in the state / the admissions algorithm], to assess the admissions applications of 100 former students of College SY. Out of these, 30 students were repeatedly placed on academic probation and consequently failed to graduate. These 30 were deliberately inserted into the 100 applications reviewed, for testing purposes.

ScholarVision was then asked to **pick the 30 students from the list of 100** who were unlikely to graduate based on their expected academic performance in college.

When the test results were reviewed, the commission found that ScholarVision had **failed to include 16 of the 30 students who could not graduate** due to poor academic results.

Figure 1. *(Continued)*

officer or algorithm was instructed to pick these 30 students from the full list as those unlikely to graduate based on their expected academic performance. When the test results were reviewed, 16 of the 30 non-graduating students were missed—that is, a failure to predict 16/30 (just over 50%) of the cases. In the scenario, only one of the decision-making systems, either the human or the algorithm, was said to have been tested and found to make errors. This factor was varied independently of what was said to be the convention; that is, in some conditions, the convention was tested and found to make errors, while in other conditions, it was the alternative, non-conventional option that was tested and found to make errors. In summary, the experiment had a two (convention: human vs. algorithm) by two (error maker: human vs. algorithm) between-subject design. The complete information regarding the conventional method and the error, as presented to the participants, is detailed in Figure 1.

Participants completed comprehension and memory checks both after learning about the conventional method and again after the error simulation to ensure they understood and retained the provided details. If participants answered incorrectly, they were given the relevant information again and were required to select the correct answer to proceed further.[2]

Finally, participants were asked to evaluate: (1) the severity of the mistake on a 1 (not at all serious) to 6 (extremely serious) scale; (2) their level of concern regarding the continued use of the error-prone method on a 1 (not all concerned) to 6 (extremely concerned); (3) whether the fictional college should maintain its conventional admissions approach on a 1 (definitely not) to 6 (definitely) scale; and (4) which admissions option—human or algorithm—they believed should be preferred by the state in the future. Table 1 provides details on each of these questions. The experiment concluded with a series of demographic questions.

[2] The comprehension checks used in this and the following study are detailed in the Materials documents on the OSF page for this paper (https://osf.io/29sj7/).

https://doi.org/10.1017/jdm.2025.8 Published online by Cambridge University Press

Table 1. *All measures presented to participants following the error.*

Measure	Study	Questions	Scale or options
Mistake severity	Study 1	According to your judgment, how serious is this error by the [human admissions officer OR admissions algorithm]?	1 (Not serious at all) to 6 (extremely serious)
	Study 2	According to your judgment, how serious is this error by the [sound-quality analyst OR sound-quality algorithm]?	
Level of concern	Study 1	How concerned should the commission be, given this mistake, about the use of the [the human officer OR computerized algorithm] in statewide college admissions?	1 (Not at all concerned) to 6 (extremely concerned)
	Study 2	How concerned should the company be, given this mistake, about the use of the [human analyst OR computerized algorithm] in quality control for its speakers?	
Retention	Study 1	Should College SY's [admissions officer Or computer program], [Jordan Taylor OR ScholarVision 3.0], be retained after the results of this test?	1 (Definitely not) to 6 (definitely)
	Study 2	Should the company's [analyst OR computer program], [Jordan Taylor Or SonicVerifier 3.0], be retained after the results of this test?	
Future choice	Study 1	Which college admissions approach do you think the commission should recommend for the state's future five-year higher education plan?	Human admissions officers OR Customized admissions algorithms
	Study 2	Which sound-quality testing method do you think the company should choose for their next five-year cycle?	A human sound-quality analyst OR A sound-quality algorithm

Note: Depending on the condition, each participant in both studies was asked only about the option within square brackets that made the error. The order of choices in the last measure (future choice) was randomized for all participants.

3.2. Results

3.2.1. Mistake severity and concern measure

We first conducted a correlation analysis to assess the relationship between the mistake severity and level of concern ratings. There was a strong, positive correlation between the two variables, $r(592) = 0.798, p < 0.001$, 95% CI [0.767, 0.825]. The high correlation suggested that the variables were capturing similar aspects of participants' judgment—that is, negative reactions to errors—and hence we combined the two to avoid redundancy. This was done by creating a new variable by averaging the two variables across all participants.

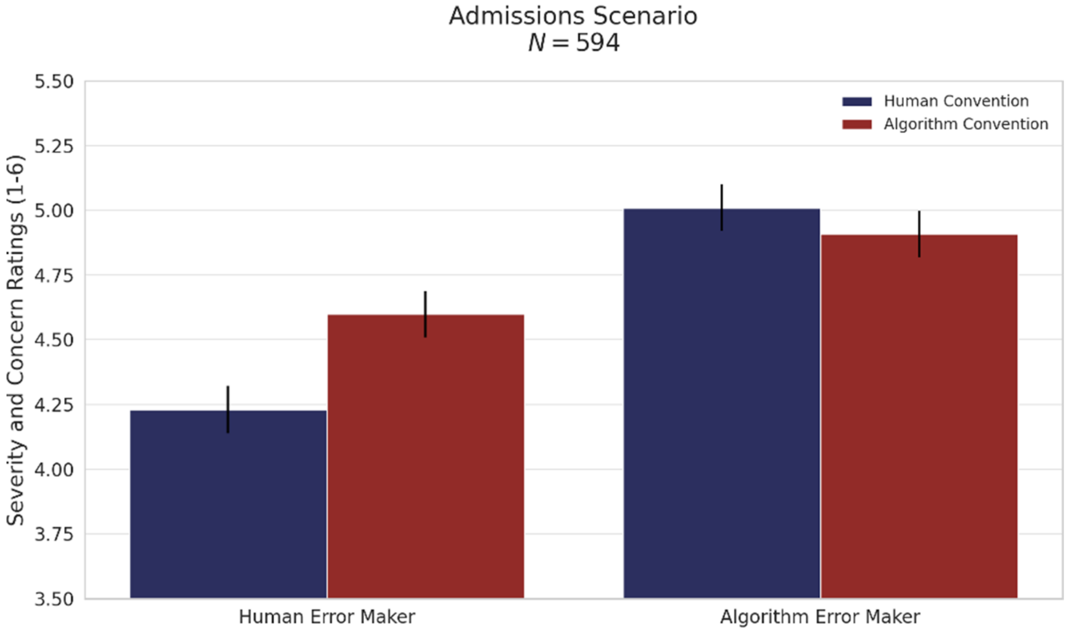


Figure 2. *Study 1: Combined severity and level of concern ratings for admissions scenario by error maker and convention. When the admissions algorithms were framed as convention (red bars), algorithm mistakes were judged more severely than human errors, but less so than when human admissions officers were presented as the convention (blue bars). Error bars represent standard error throughout this paper.*

We examined participants' responses using a convention (human convention and algorithm convention) by error maker (human admissions officer and computerized admissions algorithm) ANOVA (Figure 2). There was a main effect of the error maker ($F(1, 590) = 36.34$, $p < 0.001$, $\eta^2_P = 0.06$), indicating that overall, participants judged the computerized admissions algorithm more severely when it was presented as the error maker ($M = 4.96$, $SD = 1.00$) than they did the human admissions officer ($M = 4.42$, $SD = 1.20$). There was no main effect of convention ($F(1, 590) = 2.19$, $p = 0.14$).

However, there was a significant interaction between convention and error maker, $F(1, 590) = 6.71$, $p = 0.01$, $\eta^2_P = 0.01$. Judgments were harsher when the algorithm was the error maker compared to the human officer—particularly when the human officer was framed as the conventional method. In contrast, when the algorithm was presented as the conventional method, although its errors were still judged more harshly, the difference was considerably smaller. Specifically, when the human admissions officers were framed as the conventional option, participants judged the algorithmic error maker (admissions algorithm) significantly ($F(1, 292) = 36.08$, $p < 0.001$) more severely ($M = 5.01$, $SD = 0.99$) compared to erring human officers ($M = 4.23$, $SD = 1.22$). When admissions algorithms were framed as the convention, participants still judged the algorithmic error maker ($F(1, 298) = 6.08$, $p = 0.01$) more severely ($M = 4.91$, $SD = 1.01$) than human error makers ($M = 4.60$, $SD = 1.16$), but this difference was smaller than when the human decision-maker was the convention, as indicated by the significant interaction. In short, when the human decision-maker was framed as the convention (as in past studies of algorithm aversion), an erring algorithm was judged much more severely than a human error maker when it made identical mistakes (difference between the blue bars in the figure); but this bias against the algorithm was much less pronounced when the algorithm was framed as the convention (difference between the red bars).

https://doi.org/10.1017/jdm.2025.8 Published online by Cambridge University Press

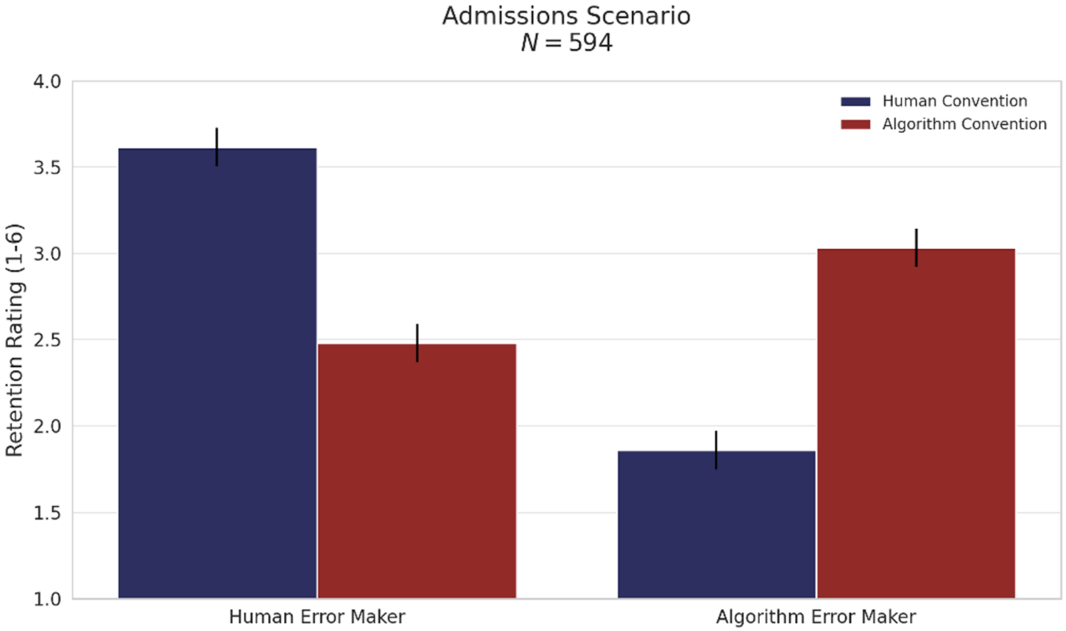


Figure 3. *Preference to retain admissions review method after it made an error.*

Note: Study 1: When admissions algorithms were framed as the convention (red bars), participants preferred retaining the algorithm more than the human after identical mistakes. When human admissions officers were framed as the convention (blue bars), participants preferred retaining the human officer over the algorithm.

3.2.2. Retention measure

The retention question (Table 1) assessed participants' preference for keeping the conventional option used by College SY. For instance, it asked whether the college's human admissions officer, Jordan Taylor (in human convention conditions), should be retained, both in the condition where the human was tested and found to make errors and, in the condition, where the algorithm was tested and found to make errors. As would be expected, participants were much less favorable to retaining the human [algorithm] when the human [algorithm] was tested and found to have made errors (compared to when the human [algorithm] was not tested and the algorithm [human] was tested instead). The main interest here, however, as captured in the analysis above of the severity and concern ratings, was of judgments of the error maker (and how those judgments are influenced by whether the error maker is a human or algorithm, and whether or not the error maker is the convention). For consistency in analysis, we recoded the retention measure so that it consistently reflected a judgment directed toward the error maker. To do this, the retention measure was reverse coded for the conditions in which the error was made by the non-conventional option. Following this recoding, in all conditions, higher scores indicate a stronger tendency to retain the error maker, despite it having been observed to make errors.

We examined participants' responses using a convention (human convention and algorithm convention) by error maker (human admissions officer and computerized admissions algorithm) ANOVA (Figure 3). We observed a main effect of the error maker ($F(1, 590) = 29.64$, $p < 0.001$, $\eta^2_P = 0.05$), indicating that overall, participants had a stronger preference for retaining the human admissions officer ($M = 3.04$, $SD = 1.39$) to retaining the admissions algorithm ($M = 2.45$, $SD = 1.54$) when they made identical errors. This suggested that, overall, participants were more forgiving of the human admissions officer error maker compared to the algorithmic error maker. There was no main effect of convention ($F(1, 590) = 0.03$, $p = 0.86$).

https://doi.org/10.1017/jdm.2025.8 Published online by Cambridge University Press

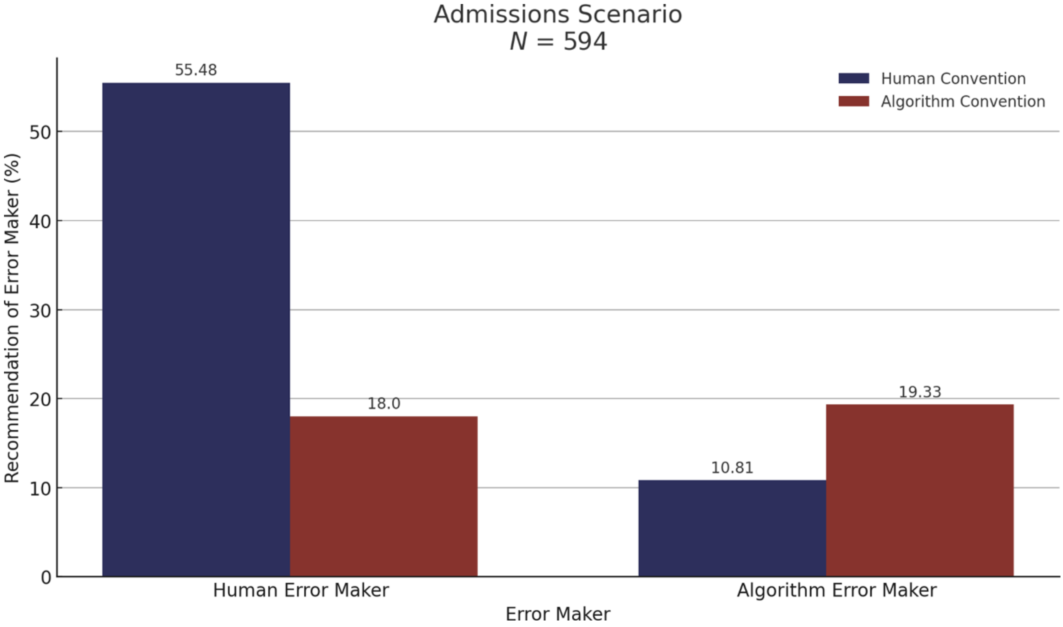


Figure 4. *Proportion of recommendations for error makers under different conventions in the admissions scenario.*

Note: Study 1: When admissions algorithms were framed as the convention (red bars), participants recommended both the humans and the algorithms at similar rates after identical mistakes. But when human admissions officers were framed as the convention (blue bars), participants recommended humans more than algorithms.

Figure 3 shows that a significant interaction was found between convention and error maker ($F(1, 590) = 109.09$, $p < 0.001$, $\eta^2{}_P = 0.16$). Participants preferred retaining the human admissions officer more ($M = 3.62$, $SD = 1.30$) than they did the admissions algorithm ($M = 1.86$, $SD = 1.37$) after both made identical mistakes in the human convention condition ($F(1, 292) = 127.96$, $p < 0.001$; difference between the blue bars in the figure). Conversely, when admissions algorithms were framed as the convention, participants retained the admissions algorithm significantly more ($M = 3.03$, $SD = 1.48$) than the human admissions officer after observing them make the same mistakes ($M = 2.48$, $SD = 1.24$) ($F(1, 298) = 12.35$, $p < 0.001$; difference between the red bars in the figure).

3.2.3. Recommendation for the future

For the final question (Table 1), participants were given a binary choice between recommending human admissions officers or computerized admissions algorithms for use in the future of the state's college admissions. Again, our main interest here—in line with the earlier analyses—was in recommendations for the error maker and how those recommendations were influenced by whether the error maker was a human or an algorithm, and whether or not the error maker was the convention. Hence, to maintain consistency in our analysis, we recoded the recommendation measure to focus on the error maker. This involved creating a new variable that reflected whether the participant had recommended the error maker (coded as 1) or did not recommend the error maker for the future (coded as 0). Following this recoding, Figure 4 shows the proportion of participants who, despite observing mistakes, still recommended the error-making agent for future use.

Of the 594 participants, 153 (25.76%) recommended retaining the error maker for future use. Of these, 108 recommendations (36.49% of 296 cases) were for human error makers and 45 recommendations (15.1% of 298 cases) were for algorithmic error makers. A Chi-square analysis of the collapsed data confirmed that this difference was significant (X^2 $(1, N = 594) = 35.51$, $p < 0.001$),

https://doi.org/10.1017/jdm.2025.8 Published online by Cambridge University Press

showing that participants were, overall, significantly more likely to recommend human error makers over algorithmic ones.

A Chi-square test was conducted to examine the relationship between the type of error maker (human or algorithm) and participants' recommendation of the error maker (recommended vs. not recommended), specifically within the conditions where the human is the convention (difference between the blue bars in the figure). We found a significant association between the type of error maker and recommendations for error makers in these conditions (X^2 (1, $N = 294$) = 66.33, $p < 0.001$). This showed that when the human admissions officer was presented as the conventional method for evaluating college applications, participants were significantly more likely to recommend a human error maker (55.48%) compared to an algorithmic error maker (10.81%).

A second Chi-square test was run between the type of error maker and participants' recommendation of the error maker within conditions where the algorithm was the convention (difference between the red bars in the figure). Here, we did not find a significant association between the type of error maker and recommendations for error maker (X^2 (1, $N = 300$) = 0.09, $p = 0.767$). This result indicated that when the admissions algorithm was framed as the conventional way to screen college applications, participants' likelihood of recommending the error maker did not significantly differ whether the error maker was a human (18.0%) or an algorithm (19.33%).

3.2.4. Study 1: Summary of findings

In summary, overall results from Study 1 show algorithm aversion across both convention conditions and all three measures—that is, when collapsed across conditions, participants generally judged and treated algorithmic error makers more harshly than human error makers. As expected, this effect was most pronounced when human admissions officers were presented as the conventional option: Participants not only judged algorithmic error makers more severely but also preferred retaining the erring human officer and recommended human error makers for future college admissions. These findings align with our predictions and prior research on algorithm aversion.

When the admissions algorithm was framed as the conventional choice, however, this pattern no longer held. Specifically, while algorithmic error makers were still judged more severely than human error makers, the difference was considerably reduced ($M_{\text{difference}} = 0.31$, $SE = 0.13$) compared to when humans were shown as the conventional option ($M_{\text{difference}} = 0.78$, $SE = 0.13$). Moreover, as seen in Figure 2, human error makers were now judged significantly more harshly than they were when human officers were described as the convention. The main finding here is that people's typical aversion to algorithmic errors was substantially offset by the intervention of framing the algorithm as the convention.

For the retention variable, the hypothesis that conventional decision-makers would have higher retention for future tasks than non-conventional ones was strongly supported. Participants showed a higher preference for retaining the admissions algorithm over the human officer when the algorithm was framed as the conventional option—indicating a reversal of the algorithm aversion effect in the other direction. This pattern provided support for our alternate aversion account, suggesting that aversion may be driven not only by the algorithm itself but also by its status as the non-conventional option. Specifically, the drop in retention ratings for the human error maker when it was the conventional option ($M_{difference} = –1.14$, $SE = 0.16$) was similar to the drop for the algorithmic error maker when it was not the conventional option ($M_{difference} = –1.18$, $SE = 0.16$).

Finally, for **future use recommendations**, we observed that when the **participants were informed that the admissions algorithm** was the **conventional** option, they **no longer favored** the **human** admissions officer over the algorithm (18.0% vs. 19.33%). This pattern suggests that the **conventional framing** completely offsets algorithm aversion in the context of future choices. Furthermore, as shown in Figure 4, the recommendation rate for the humans dropped significantly—from 55.48% (when humans were framed as conventional) to 18.0% (when the algorithm was presented as the convention).

Changing the context by presenting the algorithm as the conventional option altered participants' judgments and preferences from what we commonly observe in conventional human–algorithm comparisons. Algorithm aversion appeared to be significantly offset by the conventionality intervention in judgments of error and reversed or eliminated when participants were asked to make future choices. Overall, these findings provide strong support for our hypothesis that conventionality influences judgments of identical mistakes between humans and algorithms.

4. Study 2

In Study 2, we developed a different hypothetical scenario to test the generalizability of our earlier results. The new scenario involved the quality control of sound speakers, where human and algorithmic agents failed to detect faulty products. In contrast to the forecasting errors in the first study, mistakes made in the present study are detection-based errors. That is, mistakes were shown to be made when either humans or algorithms failed to accurately detect issues within existing products as opposed to forecasting future performance. As in Study 1, the same three characteristics—history, prevalence, and system dependence—were used to establish the conventional and alternate options.

4.1. Method

Data and materials for this experiment are available on OSF (https://osf.io/29sj7/). This study was not preregistered. Participants for this study (N = 605; 308 women, 279 men, 6 non-binary, and 1 other; $Mdn_{\text{age group}}$ = 35–44) were recruited simultaneously with those for Study 1 through CloudResearch's MTurk Toolkit, using identical recruitment criteria and screening. Participants could only take part in one of the two studies.

4.1.1. Procedure

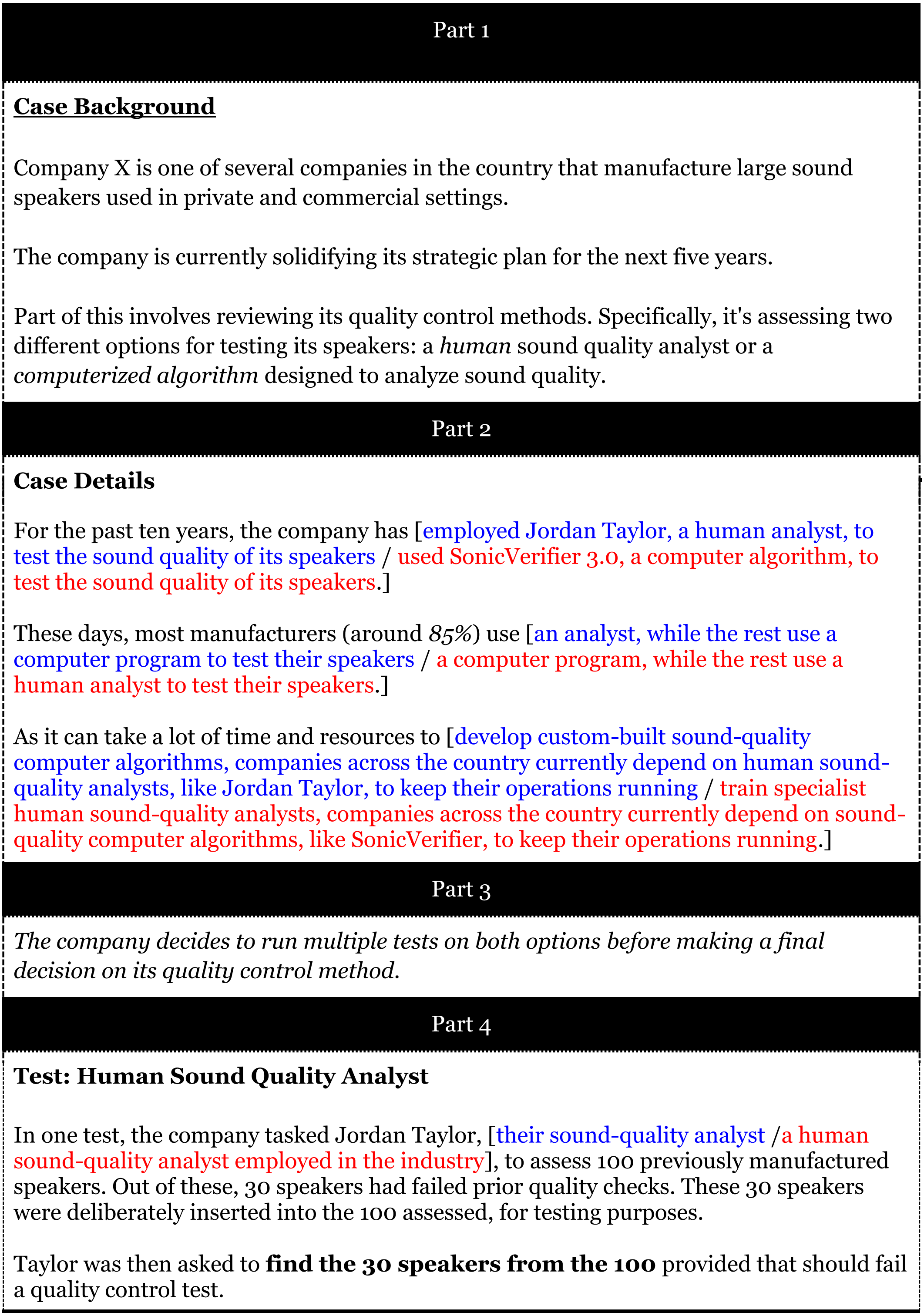
This experiment was conducted similarly to Study 1, using a fictitious case study focused on the quality control of sound speakers. Participants were briefed about a fictional company specializing in manufacturing sound speakers, which was evaluating two quality control options: a human sound quality analyst and a sound quality computer algorithm. As in Study 1, participants were informed that either a human or an algorithm was the conventional method for quality control (see Figure 5). Next, the company described a test in which 100 manufactured speakers were reviewed—30 of which were deliberately inserted as defective. The human analyst or algorithm was instructed to identify these 30 faulty speakers from the full batch. When the test results were examined, 12 of the 30 defective speakers went undetected—that is, 40% of the defective products were missed. The rest of the experimental procedure mirrored Study 1, including comprehension checks, judgment questions (Table 1), and demographic questions.

4.2. Results

4.2.1. Mistake severity and concern measure

Recall that in Study 1, we conducted a correlation analysis to examine the relationship between mistake severity and level of concern ratings. Due to a high correlation, we created a new combined variable by averaging the two variables across all participants. Similarly, in Study 2, there was a strong correlation between mistake severity and concern measures, $r(603) = 0.735$, $p < 0.001$, 95% CI [0.696, 0.770], prompting the creation of an identical combined variable for this analysis.

We examined participants' responses using a convention (human convention, algorithm convention) by error maker (human analyst, sound quality algorithm) ANOVA (Figure 6). There was a main effect of the error maker ($F(1, 601) = 44.21$, $p < 0.001$, $\eta^2_P = 0.07$), indicating that overall, participants judged the computerized quality control algorithm when it was framed as the error maker more severely



Part 1

Case Background

Company X is one of several companies in the country that manufacture large sound speakers used in private and commercial settings.

The company is currently solidifying its strategic plan for the next five years.

Part of this involves reviewing its quality control methods. Specifically, it's assessing two different options for testing its speakers: a *human* sound quality analyst or a *computerized algorithm* designed to analyze sound quality.

Part 2

Case Details

For the past ten years, the company has [employed Jordan Taylor, a human analyst, to test the sound quality of its speakers / used SonicVerifier 3.0, a computer algorithm, to test the sound quality of its speakers.]

These days, most manufacturers (around *85%*) use [an analyst, while the rest use a computer program to test their speakers / a computer program, while the rest use a human analyst to test their speakers.]

As it can take a lot of time and resources to [develop custom-built sound-quality computer algorithms, companies across the country currently depend on human sound-quality analysts, like Jordan Taylor, to keep their operations running / train specialist human sound-quality analysts, companies across the country currently depend on sound-quality computer algorithms, like SonicVerifier, to keep their operations running.]

Part 3

The company decides to run multiple tests on both options before making a final decision on its quality control method.

Part 4

Test: Human Sound Quality Analyst

In one test, the company tasked Jordan Taylor, [their sound-quality analyst /a human sound-quality analyst employed in the industry], to assess 100 previously manufactured speakers. Out of these, 30 speakers had failed prior quality checks. These 30 speakers were deliberately inserted into the 100 assessed, for testing purposes.

Taylor was then asked to **find the 30 speakers from the 100** provided that should fail a quality control test.

Figure 5. *Sound speakers scenario as presented to participants.*

Note: This includes all scenario parts of Study 2 in sequence, as seen by the participants, barring the comprehension questions. Colored text within square brackets shows variations made per condition (blue = human convention, red = algorithm convention).

When the test results were reviewed, the company found that Taylor had **missed 12 of the 30 malfunctioning speakers** that had failed previous quality control tests.

OR

Test: Sound Quality Computer Algorithm

In one test, the company tasked SonicVerifier 3.0, [a sound-quality computer algorithm used in the industry / their sound-quality algorithm], to assess 100 previously manufactured speakers. Out of these, 30 speakers had failed prior quality checks. These 30 speakers were deliberately inserted into the 100 assessed, for testing purposes.

SonicVerifier was then asked to **find the 30 speakers from the 100** provided that should fail a quality control test.

When the test results were reviewed, the company found that SonicVerifier had ***missed 12 of the 30 malfunctioning speakers*** *that had failed previous quality control tests.*

Figure 5. *(Continued)*

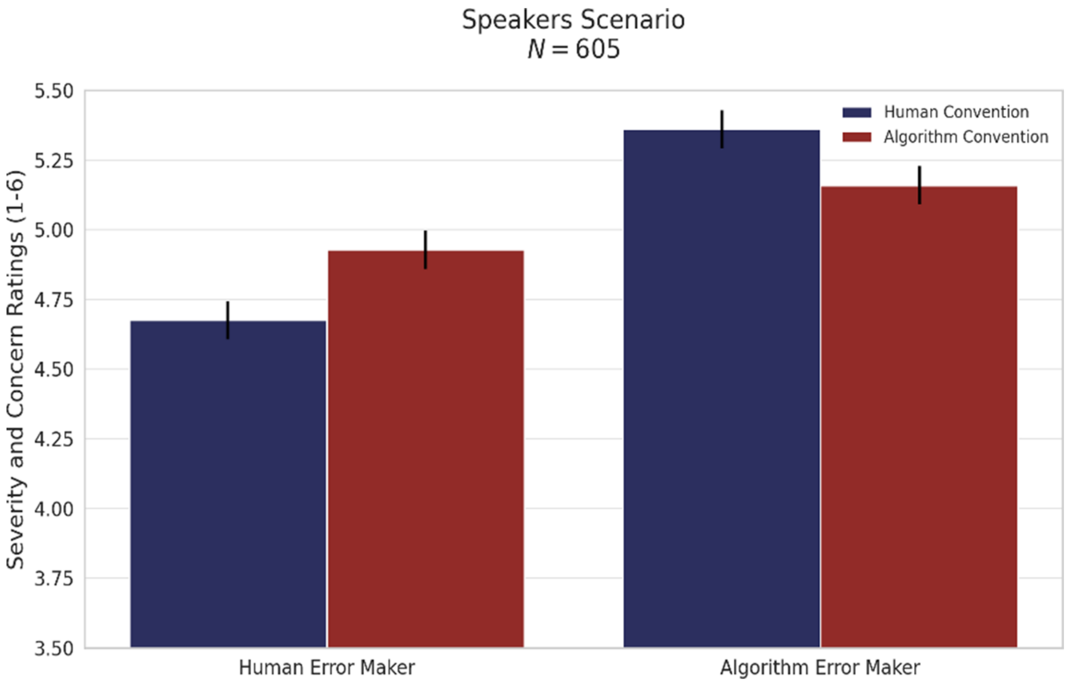


Figure 6. *Combined severity and level of concern ratings for speakers scenario by error maker and convention.*

Note: Study 2: As in Study 1, when the sound quality algorithms were framed as the convention (red bars), algorithm mistakes were judged more severely than human errors, but less so than when human analysts were presented as the convention (blue bars).

(M = 5.26, SD = 0.73) than they did the human analyst (M = 4.80, SD = 0.97). Again, there was no main effect of convention (F(1, 601) = 0.14, p = 0.71).

As in Study 1, there was a significant interaction between convention and error maker (F(1, 601) = 10.81, p = 0.001, $\eta^2{}_p$ = 0.02). Similarly, judgments were harsher when the error maker was the

https://doi.org/10.1017/jdm.2025.8 Published online by Cambridge University Press

sound quality algorithm than the human analyst, particularly when the human analyst was presented as the conventional method. In contrast, when the algorithm was framed as the conventional method, algorithmic error makers were still judged more harshly, but the difference was smaller. Specifically, when the human sound quality analysts were presented as the convention, participants judged the erring sound quality algorithm significantly ($F(1, 300) = 50.39$, $p < 0.001$) more severely ($M = 5.36$, $SD = 0.65$) compared to human error makers ($M = 4.68$, $SD = 0.99$). When sound quality algorithms were framed as the convention, participants still judged the algorithmic error makers ($F(1, 301) = 5.54$, $p = 0.012$) more severely ($M = 5.16$, $SD = 0.78$), than mistakes made by the human officer ($M = 4.93$, $SD = 0.93$). However, similar to Study 1, this difference was smaller than when the human analyst was the convention, as indicated by the significant interaction noted above. Overall, we saw again that when the human decision-maker was framed as the convention (as in past work on algorithm aversion), erring algorithms were judged much more severe than human error makers who made identical mistakes (difference between the blue bars in Figure 6); but this bias against the algorithm was much less pronounced when the algorithm was framed as the convention (difference between the red bars).

4.2.2. Retention measure

The retention question (Table 1) assessed participants' preference for keeping the conventional option used by Company X. For instance, it asked whether the college's human sound quality analyst, Jordan Taylor (in human convention conditions), should be retained, both in the condition where the human was tested and found to make errors and in the condition where the algorithm was tested and found to make errors. As was expected for the retention measure in Study 1, participants were much less favorable to retaining the human [algorithm] when the human [algorithm] was tested and found to have made errors (compared to when the human [algorithm] was not tested and the algorithm [human] was tested instead). The main interest here remained the judgments of the error maker (and how those judgments are influenced by whether the error maker is a human or algorithm, and whether or not the error maker is the convention). To maintain consistency in the analysis, we adjusted the retention measure to uniformly reflect judgments directed at the error maker. This involved reverse coding the retention measure for conditions where the error was made by the non-conventional option. After this recoding, higher scores across all conditions indicated a stronger tendency to retain the error maker, despite it having been observed to make mistakes.

We examined participants' responses using a convention (human convention, algorithm convention) by error maker (human analyst, sound quality algorithm) ANOVA (Figure 7). We observed a main effect of the error maker ($F(1, 601) = 42.70$, $p < 0.001$, $\eta^2{}_P = 0.04$), indicating that overall, participants preferred retaining the human analyst ($M = 2.70$, $SD = 1.48$) more than they preferred retaining the sound quality algorithm ($M = 2.18$, $SD = 1.41$) when they made identical errors. This time, we observed an unexpected main effect of convention ($F(1, 601) = 5.45$, $p = 0.02$, $\eta^2{}_P = 0.01$), which indicated that participants had a slightly higher preference for retaining error makers in the human convention conditions ($M = 2.56$, $SD = 1.57$) than they did in the algorithm convention conditions ($M = 2.32$, $SD = 1.35$). We do not consider this relevant to our current analyses.

A significant interaction between convention and error maker was found ($F(1, 601) = 192.0$, $p < 0.001$, $\eta^2{}_P = 0.24$). Participants preferred retaining the human analyst ($M = 3.54$, $SD = 1.29$) significantly more than the admissions algorithm ($M = 1.59$, $SD = 1.17$) after both made identical mistakes—when human analysts were framed as the convention ($F(1, 300) = 187.44$, $p < 0.001$; difference between the blue bars in Figure 7). Conversely, in the algorithm convention condition ($F(1, 301) = 36.33$, $p < 0.001$; difference between the red bars in the figure), participants were more likely to retain the sound quality algorithm ($M = 2.77$, $SD = 1.39$) than the human analyst ($M = 1.88$, $SD = 1.16$) after observing the same mistakes.

https://doi.org/10.1017/jdm.2025.8 Published online by Cambridge University Press

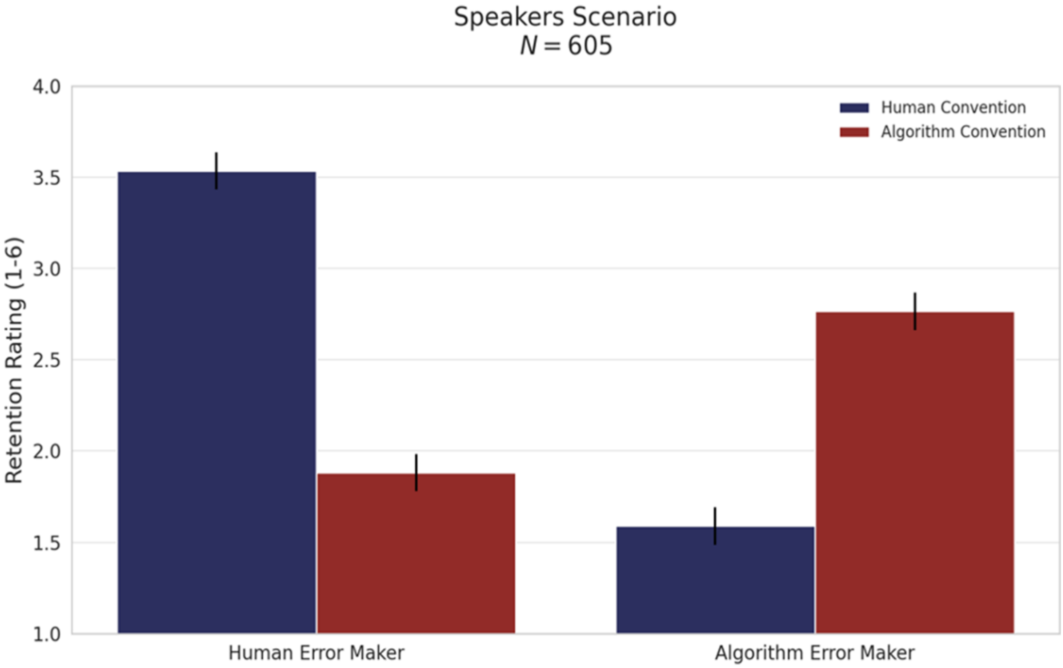


Figure 7. *Preference to retain sound testing method after it made an error.*

Note: Study 2: When sound quality algorithms were the framed convention (red bars), participants retained the algorithm more than the human after identical mistakes. When human analysts were the framed convention (blue bars), participants preferred retaining the human over the algorithm.

4.2.3. Recommendation for the future

For the final question of Study 2 (Table 1), participants were asked to indicate whether Company X should use a human sound quality analyst or a computerized sound quality algorithm for the next 5 years. The analysis was identical to that used for the recommendation measure in Study 1. Figure 8 displays the proportion of participants who, despite observing mistakes, still recommended the error-making agent for future use.

Of the 605 participants, 139 (23.0%) recommended retaining the error maker for future use. Of these, 84 recommendations (27.6% of 304 cases) were for human error makers and 55 recommendations (18.3% of 301 cases) were for algorithmic error makers. A Chi-square analysis of the collapsed data confirmed that this difference was statistically significant (X^2 (1, $N = 605$) = 7.49, $p = 0.006$), reflecting yet again that participants were, overall, significantly more likely to recommend human error makers over algorithmic ones.

As in Study 1, a Chi-square test was conducted to examine the relationship between the type of error maker (human or algorithm) and participants' recommendation of the error maker (recommended vs. not recommended), specifically within the conditions where the human is the convention (difference between the blue bars in Figure 8). There was a significant association between the type of error maker and recommendations for error maker in these conditions (X^2 (1, $N = 302$) = 68.57, $p < 0.001$). Consistent with Study 1 results, this showed that when the human analyst was presented as the conventional method for testing sound speakers, participants were significantly more likely to recommend a human error maker (48.34%) compared to an algorithmic error maker (5.96%).

A second Chi-square test was run between the type of error maker and participants' recommendation of the error maker within conditions where the algorithm was the convention (difference between the red bars in the figure). This time, we found a significant association between the type of error maker and recommendations for error makers under these conditions (X^2 (1, $N = 303$) = 27.33, $p < 0.001$).

https://doi.org/10.1017/jdm.2025.8 Published online by Cambridge University Press

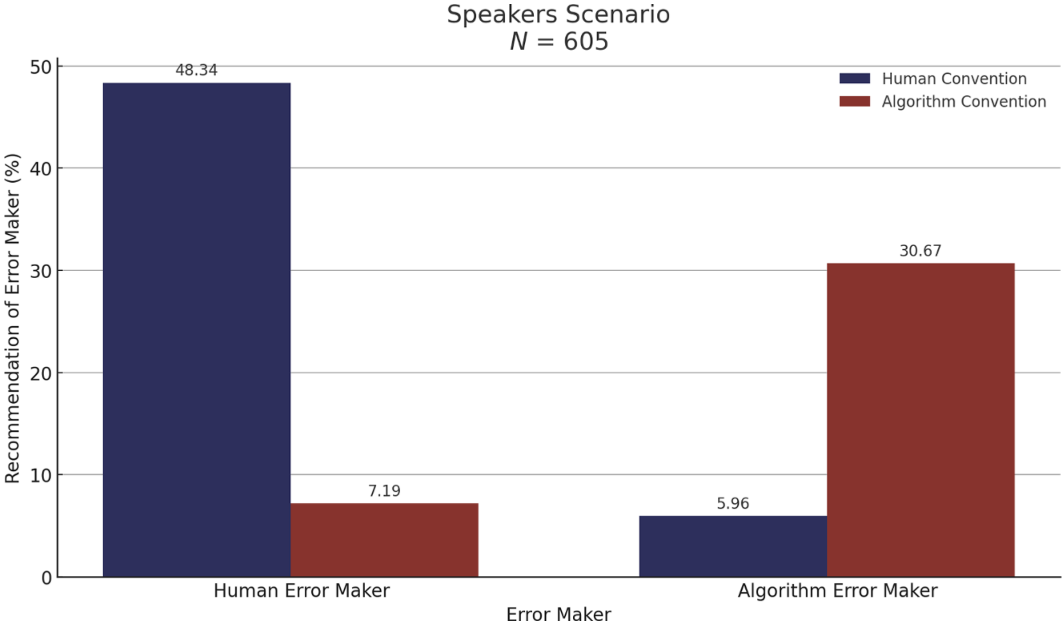


Figure 8. *Proportion of recommendations for error makers under different conventions in the speakers' scenario.*

Note: Study 2: When sound quality algorithms were framed as the convention (red bars), participants recommended the algorithm more than the human for future use after identical mistakes. When human analysts were the framed convention (blue bars), participants recommended humans more than algorithms.

As in Study 1, we saw the aversion to algorithms was completely offset when the algorithm was framed as the convention, with human analysts no longer recommended more often than the algorithms. In fact, in this experiment, the effect reversed, such that the algorithms were favored over the human analysts (30.67% vs. 7.19%) for future recommendations.

4.2.4. Study 2: Summary of findings

In line with Study 1, overall results from Study 2 reflected algorithm aversion across both error conditions; that is, participants judged algorithmic error makers more harshly than human error makers when data were collapsed across convention framing and all three measures. As expected, this effect was strongest when humans were portrayed as the conventional option: Participants judged algorithmic error makers more severely, retained erring human analysts more frequently, and recommended humans for future use. Again, these findings align with our predictions and prior research on algorithm aversion.

In Study 2, although algorithmic error makers were still judged more harshly than human error makers, the difference was considerably reduced ($M_{\text{difference}}$ = 0.23, SE = 0.10) compared to when humans were the conventional option ($M_{\text{difference}}$ = 0.69, SE = 0.10). Moreover, the drop in retention ratings for the human error maker when the human analyst was the conventional option ($M_{difference}$ = –1.66, SE = 0.14) was larger than the drop in retention for the algorithmic error maker when the sound quality algorithm was the conventional option ($M_{difference}$ = –1.18, SE = 0.15). This large decline in retention for the human error maker indicates that when humans were the alternate option, tolerance for their mistakes decreased substantially.

Like Study 1, when the algorithm was presented as the convention, human analysts were no longer recommended more often than the algorithms (18.0% vs. 19.33%). In Study 2, this difference became starker as we saw algorithms preferred over human analysts (30.67% vs. 7.19%) for future recommendations after making identical mistakes. In fact, in Study 2, this difference became starker,

https://doi.org/10.1017/jdm.2025.8 Published online by Cambridge University Press

with algorithms being preferred over human analysts for future recommendations (30.67% vs. 7.19%) after identical mistakes. Overall, the results of Study 2 suggest that our findings from Study 1 are generalizable to a different contextual situation, and provide strong evidence that the conventionality intervention can significantly offset algorithm aversion.

5. General discussion

It has long been known that the status quo really matters (Samuelson and Zeckhauser, 1988): default choices can shape subsequent judgments and behaviors (Johnson and Goldstein, 2003). This work provides evidence that biases stemming from the status quo also go on to have a consequential influence on people's evaluations of identical mistakes. Specifically, in this paper, we operationalized the status quo as conventional options and investigated how the conventional status of the error maker influences judgments and decisions. In both studies, we observed that when human admissions officers or human analysts were considered the conventional choice, identical errors made by algorithms were judged more harshly compared to those made by humans. Put simply, participants generally judged algorithmic mistakes more severely than human mistakes. This aligns with previous research on algorithm aversion, where people exhibit a distrust of algorithmic decision-making.

However, it is also relevant to consider that people have an inherent bias toward maintaining the conventional status quo, which traditionally frames humans as the conventional choice in human–algorithmic comparisons. Perhaps this bias can lead to a harsher evaluation of errors made by non-conventional algorithmic decision-makers compared to their human counterparts. To test this idea, we created scenarios in which the algorithm was described as the convention and found that when we framed the algorithm as the conventional option, there was a significant impact on people's judgments and choices. This significantly or even completely offsets the algorithm aversion effect usually observed in standard human–algorithm comparisons.

5.1. Alternate aversion

Our results support an alternate aversion account: The presence of a conventional option intensifies aversion toward the non-conventional alternative. In the context of human–algorithm comparisons, when humans are the default, an algorithm is not simply penalized because of inherent flaws; rather, it is its status as the 'alternate' option that drives harsher judgment. Conversely, when the algorithm is framed as conventional, human errors are judged more severely, and participants show a preference for the algorithm when making decisions about retention and future use. This finding suggests that the context—how conventional an option is perceived—plays a critical role in shaping responses to identical errors. Thus, rather than algorithm aversion being solely a function of algorithmic deficiencies, our work indicates that conventionality is a key contextual factor in these judgments.

The idea of alternate aversion helps explain inconsistencies in the literature on algorithm aversion and appreciation. Although earlier research documented robust algorithm aversion, more recent studies have shown cases of algorithm appreciation (Logg et al., 2019) where people prefer algorithmic advice over human input or are at least less averse to it (Bigman et al., 2023; Pálfi et al., 2022). Our findings suggest that these seemingly contradictory patterns may coexist depending on which option is established as the convention. When algorithms become the conventional option, the usual negative reaction to their errors is offset, potentially leading to a more favorable overall evaluation—even if errors are still recognized as severe.

5.2. Judgments versus choices

In our studies, there were two types of judgment—one about how severe the error was and the other about which agent should be retained and recommended for the future, in light of the error. Across

https://doi.org/10.1017/jdm.2025.8 Published online by Cambridge University Press

both studies, when the human is framed as the convention, the usual algorithm aversion appears in both judgments of mistake severity and recommendations of which system to use going forward. When the algorithm is presented as the convention, the algorithm's mistakes are not judged as less severe than the human's, but now recommendations about which to retain do tend to favor the algorithm. In short, the effect of the convention appears to be stronger on these future recommendations than on judgments of mistake severity, where algorithm aversion seemed to be more persistent.

One way to interpret this is that when the algorithm is the convention, people still judge algorithmic errors quite severely, but this bias was overcome or mitigated when it came to actionable decision-making. This finding seemingly contradicts research that has shown that once people see an algorithm make an error, they trust it less than a human agent in the future, even when the algorithm has better accuracy overall (Dietvorst et al., 2015). We do not think that it is a contradiction to earlier research. Instead, it reiterates the fact that in most human–algorithm comparisons, it is rarely considered whether people view the algorithm as the norm for that task, nor is the algorithm explicitly established as the conventional choice, as we did in these experiments. An argument can be made that this preference for the convention is rational, regardless of whether it leads to a bias for human or algorithmic error makers. We described convention based on three characteristics: familiarity and history with the agent, prevalence of the agent's use in the surrounding environment, and dependence of the system on the agent. Based on these factors, if an agent is the conventional method of doing things and it errs, penalizing it severely or discontinuing its use can be costly. Therefore, it is rational to not discard it right away. While people might feel a bias against the algorithm for all the reasons that have been historically proposed, they are likely to be more objective in their actual decisions. It is also noteworthy that while earlier research showed that superior accuracy was not enough for people to prefer algorithms over humans after mistakes, this work has shown that the algorithmic agent being the convention could be a more persuasive factor than objective performance. This is a potential area for future research, where the performance and conventionality of error makers could be further explored, particularly in how these factors interact to influence trust and decision-making. This duality in judgments and actionable decisions also suggests that people's aversion to algorithms after they make mistakes should not be considered an outright predictor of their actual decisions. If the algorithm is the conventional option for a task, people may choose to persist with it even after it has made mistakes. We can expect to see more of these trends as people's lives become increasingly intertwined with algorithms, and they become the conventional method for more tasks.

5.3. Prediction- versus detection-based differences

The distinction between prediction-based and detection-based errors might also have influenced participants' judgments to an extent. In Study 1, errors were related to predicting future outcomes (admissions decisions), which inherently involve uncertainty and complexity. People might be more lenient toward prediction-based errors due to the elements of inherent uncertainty in forecasting the future, especially in the case of human predictions (Ganzach and Krantz, 1991). For instance, participants could attribute such errors to better-than-expected performance by applicants during their college years resulting from unaccounted-for factors like personal growth or changes in motivation due to new experiences (Deci and Ryan, 2000; Kilgo et al., 2016). In contrast, detection errors might be viewed as more avoidable and thus judged more harshly. This context could have potentially contributed to why human analysts in the sound speaker scenario (Study 2) were judged more harshly relative to algorithms, compared to human admissions officers in the first study.

5.4. Human outcomes

In Study 1, participants observed the admissions officer or the algorithm make an error in forecasting the future, which had the potential to directly and negatively affect human applicants when deployed for

https://doi.org/10.1017/jdm.2025.8 Published online by Cambridge University Press

admissions screening. People might be more opposed to algorithmic or AI-based judgments when these decisions directly impact human lives, as opposed to situations with objective, non-human outcomes (Castelo et al., 2019; Grove and Meehl, 1996). In Study 2, we used a different hypothetical scenario that did not have direct human outcomes—instead, an inanimate object (a sound speaker) was at the wrong end of the faulty decision-making. Some differences in results between the two studies, such as lower future recommendation rates for the algorithmic error maker in the admissions scenario compared to the sound speakers scenario, could perhaps be partly attributed to this factor. These results support the need for more contextual distinction in people's evaluations of and reactions to mistakes, such as who or what is affected by the errors.

5.5. Limitations and conclusion

Our research offers important insights into conventionality framing interventions and algorithm aversion, yet it is not without its limitations. People may have other assumptions and considerations beyond historic use, prevalence, and system dependence, which could influence their reactions observed in our studies. For instance, it is plausible that participants' lenient evaluations of status quo errors reflect an implicit assumption that changing the status quo incurs extra costs. Although we used system dependence as a proxy, we did not explicitly manipulate financial or logistical costs of changing the status quo. Even so, participants generally favored the conventional options for retention and future recommendations, with perhaps cost implications weighing in on those decisions. Meanwhile, in judgment-only measures (i.e., how bad the error is), algorithmic status quo errors remained harsher than human ones, perhaps because a bias against algorithms carries no direct cost in purely evaluative, non-committal judgments. Another such assumption could be acceptable accuracy. It is reasonable to assume that, in most cases, a conventional option must be working reasonably well to have longstanding prior use, and for people to perceive it as prevalent and recognize a cost associated with switching due to system dependence. Therefore, in such cases, what is the extent of acceptable accuracy—or error rate—that people are willing to accept in a given context to avoid overturning the status quo? This aspect was not manipulated in our current experiments within the same scenarios, where we used relatively high error rates as a proof of concept for our present theory. Along similar lines, participants may assume that the status quo or convention reflects the best-performing option, thereby interpreting any flaw in the conventional approach as only relatively problematic. For instance, if the widely adopted choice conventional is presumed superior performance, people may reason that a flawed non-conventional agent must be worse. While we did not systematically manipulate performance information in our studies, future research could address this by explicitly presenting or withholding performance data for both options to gage how much these assumptions drive convention preference.

We acknowledge that explicitly labeling one option as the industry's preferred choice might create an impression of a 'right answer'. However, this is central to our goal: to examine how the conventional status of the error maker shapes people's judgments. In real-world contexts, a well-established conventional option often carries the perception of being the default or 'correct' approach and replicating this dynamic was crucial for testing our hypotheses. Consequently, what might appear as experimenter demand is closer to the real-world dynamic we sought to capture. In future research, one could investigate naturally occurring status quo options (e.g., where algorithms are already the standard in certain industries) or incorporate more subtle cues of conventionality and measure demand characteristics directly.

The studies also relied heavily on controlled hypothetical scenarios which, while useful for isolating variables, may not fully capture the complexity of real-world decision-making environments. Future research could aim to replicate these findings in more naturalistic settings—such as actual interactions with an algorithmic agent (e.g., an AI chatbot) to establish its conventionality—to enhance ecological validity. To explore the effects of conventionality, we focused on specific tasks—college admissions and quality control for sound speakers—and varied the type of task—prediction based and detection based.

https://doi.org/10.1017/jdm.2025.8 Published online by Cambridge University Press

Further studies should explore a broader range and types of tasks to determine the generalizability of our findings.

An important consideration is the unique feature of AI's broad applicability, distinguishing it from technological advancements with more defined scopes, such as the previously mentioned printing press or calculators. The versatility of AI allows it to be integrated into a vast array of tasks—from navigation and writing assistance to data processing and robotics. However, this same breadth can also imply that familiarity with AI in one application does not necessarily translate to acceptance in another. For instance, someone accustomed to using AI for navigation may still resist adopting AI as a writing assistant or social companion. The broad scope of AI could both facilitate its widespread adoption and hinder it due to resistance in unfamiliar contexts. Future research could explore whether exposure to AI in multiple contexts reduces resistance or aversion to its adoption in new domains.

Our studies, using a simple conventionality intervention, showed that a bias against algorithms is significantly offset when the algorithmic agent is framed as the conventional choice. These findings suggest that algorithm aversion is highly context dependent rather than a universal bias and that longstanding human cognitive biases—such as the preference for the status quo—play a key role. This could in turn help explain and predict human interactions with algorithms across different contexts. Put simply, algorithm aversion might not be as pervasive or complex as it initially appears, and understanding these dynamics can improve our theories of human–algorithm and human–AI interactions.

Data availability statement. Data and materials are available on the project's Open Science Framework page (https://osf.io/29sj7/).

Acknowledgements. The data and ideas presented in the article were previously disseminated at the Society of Judgment and Decision-Making Annual Meeting in 2024.

Funding statement. This article was funded by the Natural Sciences and Engineering Research Council of Canada grants awarded to Jonathan A. Fugelsang and Derek J. Koehler.

Competing interests. We have no known conflicts of interests to disclose.

References

Adams, D., (1999, August 29). A hitchhiker's guide to the internet. *Sunday Times* [London, England]. https://link-gale-com.proxy.lib.uwaterloo.ca/apps/doc/A57986354/AONE?u=uniwater&sid=bookmark-AONE&xid=0417cd62

Baron, J., & Ritov, I. (2004). Omission bias, individual differences, and normality. *Organizational Behavior and Human Decision Processes*, *94*(2), 74–85. https://doi.org/10.1016/j.obhdp.2004.03.003

Bigman, Y. E., & Gray, K. (2018). People are averse to machines making moral decisions. *Cognition*, *181*, 21–34. https://doi.org/10.1016/j.cognition.2018.08.003

Bigman, Y. E., Wilson, D., Arnestad, M. N., Waytz, A., & Gray, K. (2023). Algorithmic discrimination causes less moral outrage than human discrimination. *Journal of Experimental Psychology: General*, *152*(1), 4–27. https://doi.org/10.1037/xge0001250

Bonezzi, A., & Ostinelli, M. (2021). Can algorithms legitimize discrimination? *Journal of Experimental Psychology: Applied*, *27*(2), 447–459. https://doi.org/10.1037/xap0000294

Bonezzi, A., Ostinelli, M., & Melzner, J. (2022). The human black-box: The illusion of understanding human better than algorithmic decision-making. *Journal of Experimental Psychology: General*, *151*(9), 2250–2258. https://doi.org/10.1037/xge0001181

Brzezicki, M. A., Bridger, N. E., Kobetić, M. D., Ostrowski, M., Grabowski, W., Gill, S. S., & Neumann, S. (2020). Artificial intelligence outperforms human students in conducting neurosurgical audits. *Clinical Neurology and Neurosurgery*, *192*, 105732. https://doi.org/10.1016/j.clineuro.2020.105732

Carabantes, M. (2020). Black-box artificial intelligence: An epistemological and critical analysis. *AI & SOCIETY*, *35*(2), 309–317. https://doi.org/10.1007/s00146-019-00888-w

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

Dawes, R. M. (1979). The robust beauty of improper linear models in decision making. *American Psychologist*, *34*(7), 571–582. https://doi.org/10.1037/0003-066X.34.7.571

Dawes, R. M., Faust, D., & Meehl, P. E. (1989). Clinical versus actuarial judgment. *Science*, *243*(4899), 1668–1674. https://doi.org/10.1126/science.2648573

https://doi.org/10.1017/jdm.2025.8 Published online by Cambridge University Press

Deci, E. L., & Ryan, R. M. (2000). The "what" and "why" of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry*, *11*(4), 227–268. https://doi.org/10.1207/S15327965PLI1104_01

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), 114–126. https://doi.org/10.1037/xge0000033

Dinner, I., Johnson, E. J., Goldstein, D. G., & Liu, K. (2011). Partitioning default effects: Why people choose not to choose. *Journal of Experimental Psychology: Applied*, *17*(4), 332–341. https://doi.org/10.1037/a0024354

Eastwood, J., Snook, B., & Luther, K. (2012). What people want from their professionals: Attitudes toward Decision-making strategies. *Journal of Behavioral Decision Making*, *25*(5), 458–468. https://doi.org/10.1002/bdm.741

Einhorn, H. J. (1986). Accepting error to make less error. *Journal of Personality Assessment*, *50*(3), 387–395. https://doi.org/10.1207/s15327752jpa5003_8

Ganzach, Y., & Krantz, D. H. (1991). The psychology of moderate prediction: II. Leniency and uncertainty. *Organizational Behavior and Human Decision Processes*, *48*(2), 169–192. https://doi.org/10.1016/0749-5978(91)90011-H

Grove, W. M., & Meehl, P. E. (1996). Comparative efficiency of informal (subjective, impressionistic) and formal (mechanical, algorithmic) prediction procedures: The clinical–statistical controversy. *Psychology, Public Policy, and Law*, *2*(2), 293–323. https://doi.org/10.1037/1076-8971.2.2.293

Highhouse, S. (2008). Stubborn Reliance on Intuition and Subjectivity in Employee Selection. *Industrial and Organizational Psychology*, *1*(3), 333–342. https://doi.org/10.1111/j.1754-9434.2008.00058.x

Johnson, E. J., & Goldstein, D. (2003). Do defaults save lives? *Science*, *302*(5649), 1338–1339. https://doi.org/10.1126/science.1091721

Johnson, E. J., Shu, S. B., Dellaert, B. G. C., Fox, C., Goldstein, D. G., Häubl, G., Larrick, R. P., Payne, J. W., Peters, E., Schkade, D., Wansink, B., & Weber, E. U. (2012). Beyond nudges: Tools of a choice architecture. *Marketing Letters*, *23*(2), 487–504. https://doi.org/10.1007/s11002-012-9186-1

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

Kahneman, D., Knetsch, J. L., & Thaler, R. H. (1991). Anomalies: The endowment effect, loss aversion, and status quo bias. *Journal of Economic Perspectives*, *5*(1), 193–206. https://doi.org/10.1257/jep.5.1.193

Kilgo, C. A., Mollet, A. L., & Pascarella, E. T. (2016). The estimated effects of college student involvement on psychological well-being. *Journal of College Student Development*, *57*(8), 1043–1049. https://doi.org/10.1353/csd.2016.0098

Logg, J. M. (2022). The psychology of big data: developing a "theory of machine" to examine perceptions of algorithms. In *The psychology of technology*. 349–378). American Psychological Association. https://doi.org/10.1037/0000290-011

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

Meehl, P. E. (1954). *Clinical versus statistical prediction: A theoretical analysis and review of the literature*. Minneapolis, MN: University of Minnesota Press.

Önkal, D., Goodwin, P., Thomson, M., Gönül, S., & Pollock, A. (2009). The relative influence of advice from human experts and statistical methods on forecast adjustments. *Journal of Behavioral Decision Making*, *22*(4), 390–409. https://doi.org/10.1002/bdm.637

Pálfi, B., Arora, K., & Kostopoulou, O. (2022). Algorithm-based advice taking and clinical judgement: Impact of advice distance and algorithm information. *Cognitive Research: Principles and Implications*, *7*(1), 70. https://doi.org/10.1186/s41235-022-00421-6

Promberger, M., & Baron, J. (2006). Do patients trust computers? *Journal of Behavioral Decision Making*, *19*(5), 455–468. https://doi.org/10.1002/bdm.542

Rainie, L., & Anderson, J. (2017), *Code-dependent: Pros and cons of the algorithm age* Pew Research Center: Internet, Science & Tech https://www.pewresearch.org/internet/2017/02/08/code-dependent-pros-and-cons-of-the-algorithm-age/

Ritov, I., & Baron, J. (1992). Status-quo and omission biases. *Journal of Risk and Uncertainty*, *5*(1). 49–61 https://doi.org/10.1007/BF00208786

Samuelson, W., & Zeckhauser, R. (1988). Status quo bias in decision making. *Journal of Risk and Uncertainty*, *1*(1), 7–59. https://doi.org/10.1007/BF00055564

Shaffer, V. A., Probst, C. A., Merkle, E. C., Arkes, H. R., & Medow, M. A. (2013). Why do patients derogate physicians who use a computer-based diagnostic support system? *Medical Decision Making*, *33*(1), 108–118. https://doi.org/10.1177/0272989X12453501

Thaler, R. (1980). Toward a positive theory of consumer choice. *Journal of Economic Behavior & Organization*, *1*(1), 39–60. https://doi.org/10.1016/0167-2681(80)90051-7

Tversky, A., & Kahneman, D. (1991). Loss aversion in riskless choice: A reference-dependent model. *The Quarterly Journal of Economics*, *106*(4), 1039–1061. https://doi.org/10.2307/2937956

Cite this article: Tariq, H., Fugelsang, J. A., and Koehler, D. J. (2025). Using conventional framing to offset bias against algorithmic errors. *Judgment and Decision Making*, e22. https://doi.org/10.1017/jdm.2025.8