You're Not the Boss of Me, Algorithm: Increased User Control and Positive Implicit Attitudes are Related to Greater Adherence to an Algorithmic Aid

Abstract

This study examined whether participants' adherence to an algorithmic aid was related to the degree of control they were provided at decision point and their attitudes toward new technologies and algorithms. It also tested the influence of control on participants' subjective reports of task demands whilst using the aid. 159 participants completed an online experiment centred on a simulated forecasting task, which required participants to predict the performance of school students on a standardised mathematics test. For each student, participants also received an algorithm-generated forecast of their score. Participants were randomly assigned to either the 'full control' (adjust forecast as much as they wish), 'moderate control' (adjust forecast by 30 per cent), or 'restricted control' (adjust forecast by 2 per cent) group. Participants then completed an assessment of subjective task load, a measure of their explicit attitudes towards new technologies, demographic and experience items (age, gender, and computer literacy), and a novel version of the Go/No-Go Association Task, which tested their implicit attitudes towards algorithms. The results revealed that participants who were provided with more control over the final forecast tended to deviate from it more greatly and reported lower levels of frustration. Further, participants showing more positive implicit attitudes towards algorithms were found to deviate less from the algorithm's forecasts, irrespective of the degree of control they were given. The findings allude to the importance of users' control and pre-existing attitudes in their acceptance of, and frustration in using a novel algorithmic aid, which may ultimately contribute to their intention to comply with/use them in the workplace. These findings can guide system developers and support workplaces implementing expert system technology.

Keywords: algorithm, algorithmic aid, expert system, user control, implicit attitudes, task demands, Go/No-Go Association Task

Highlights

- Gender differences existed for participants' explicit attitudes relating to new technology and perceptions of task demands when using an algorithmic aid.
- Participants tended to deviate from the algorithm's forecasts more greatly when given greater control at decision point.
- Participants experienced greater frustration using the algorithmic aid when they had no control at decision point compared to when they had full control.
- Participants with more positive implicit attitudes towards algorithms showed greater adherence to the algorithm's forecasts.

You're Not the Boss of Me, Algorithm: Increased User Control and Positive Implicit Attitudes are Related to Greater Adherence to an Algorithmic Aid

The onset of COVID-19 has seen a significant increase in our interaction with computers, from our use of communication technologies that facilitate information exchange among people, to others that enable collaboration between people and 'expert' systems. The latter interaction typically involves the system generating recommendations for the user, which it derives from the use of algorithms and associated processes (e.g., artificial intelligence (AI) and machine learning (ML) techniques). For instance, algorithmic aids have recently been developed to predict the path of bush fires (Zarghami & Dumrak, 2021), formulate recommendations for cancer management (Somashekhar, et al., 2018), and aid in the investigation of crimes (Schiliro, Beheshti, & Moustafa, 2021).

Such systems have regularly demonstrated superior accuracy to unaided human judgement and decision-making, and have particularly salient reliability advantages in tasks characterised by uncertainty (Dawes et al., 1989; Grove et al., 2000; Meehl, 1954). Despite this performance advantage, users' trust and acceptance of algorithmic aids will often decrease when they reveal themselves to be imperfect in such conditions; reducing the likelihood of user adherence/uptake (Dietvorst et al., 2015). A range of factors have been implicated in determining algorithm aversion (see Burton et al., 2019 and Mahmud, 2022 for recent reviews), however two notable factors are frequently cited as playing a significant role; the degree of system explainability and users' control in their interactions with systems.

The complex procedures involved in modern algorithms (e.g., ML) tend to result in limited understanding among users for how they operate (i.e., incomplete or inaccurate mental models). Indeed, to most users (and often system developers), a complex algorithm may represent a 'black box' of information processing (David et al., 2021; Innes & Morrison, 2021), with much of their inner workings being opaque. Further, there exists a relationship between complexity and transparency, whereby simple or more explainable systems may have limitations on performance that more complex, less explainable systems do not (Herm et al., 2022). As such, the reduced explainability associated with complex algorithms has implications for users' trust and acceptance.

The degree of explainability associated with algorithmic systems becomes more important in uncertain environments in which the reliability of the system will invariably be affected by less stable associations in the operational environment (i.e., 'low validity' environments; Kahneman & Klein, 2009). In a recent experimental study (David et al., 2021), researchers simulated an algorithmic forecasting system that provided users with financial advice. Participants either received advice from the algorithmic advisor, or a human, with algorithmic forecasts being accompanied by five types of Local and Global explanations, which differed in relation to the degree of detail regarding how the model came up with a specific prediction. They found that when the system made errors, the presence of local explanations (i.e., feature- or accuracy-based explanations of how the system worked) became more important in maintaining system adherence among users. This suggests that telling users about the information used by the system and its projected accuracy are important ways to maintain adherence to aids in uncertain conditions.

It is believed that the greater acceptance that stems from increased explainability can be partly attributed to the associated increase in feelings of control among users when using the system (Burton et al., 2019). In addition to the increased sense of control from explainability, users may achieve a more direct sense of control via their involvement in the judgement/decision process. This can range from providing input into data acquisition, to the capacity to modify the algorithm's recommendation at decision point.

In exploring how user control may affect algorithm users' adherence to algorithm-generated forecasts, Dievorst et al. (2018) asked participants to forecast students' scores on a standardised mathematics test. Participants were required to make 20 forecasts based on nine variables relating to 20 different students; race, socioeconomic status, desired occupation at age 30, predicted highest degree, region of country students were from, times taken the Preliminary Scholastic Assessment Test (PSAT), how many friends not going to college, favourite school subject, and Advanced Placement (AP) test completions. Real data were used to create algorithm-generated forecasts, which participants also accessed prior to making their own forecast. Participants were advised that, on average, the algorithm model was inaccurate by 17.5 per cent (i.e., the model was imperfect). Across a series of experiments, the results revealed that the participants were significantly more likely to use the imperfect algorithm if they were given the option to modify its forecasts, and consequently performed better having used it compared to those who did not. Further, this preference was no less salient in conditions where the degree of modification provided to participants was greatly restricted. This suggested that giving participants even a small degree of control had a positive effect in increasing appreciation for an algorithmic aid. Further, participants allowed control to modify the algorithm's outputs expressed greater satisfaction with it, and were more likely to believe that it was superior to unaided forecasting.

The increased satisfaction reported by participants in Dietvorst's (2018) study may flag the existence of other indirect benefits from increased user control, such as a reduction in perceived demands when using the system (e.g., mental and temporal), as well as their perceived degree of effort and frustration during interactions. Previous research has shown that sustained exposure to increased work demands can increase the potential for the onset of mental health concerns on workers (de Lange et al., 2003; Johnson et al., 2020; McClure, 2018). Indeed, the introduction of increased automation at work is predicted to introduce new psychosocial stressors (i.e., psychological harm that stems from changes in work design; Murray & Rostis, 2007) in the workplace, and an increased potential for mental health and wellbeing concerns (Johnson et al., 2020). Determining the potential role of user control as a protective factor relating to workers' mental health and wellbeing is a scarcely researched area and represents a critical line of enquiry. As such, the current study seeks to examine the potential association between user control when engaging algorithmic aids and their perceptions of subjective task demands.

Users' acceptance and perceptions of work demands may also be influenced by factors that precede the interaction with a specific algorithmic system. For instance, in a healthcare context, greater internal locus of control among individuals (i.e., a belief that they have control over events in their lives) has been associated with more negative attitudes about algorithm-aided care (Shaffer et al., 2012). Others have attributed a broad aversion to algorithms to a general distrust that stems from negative perceptions that people hold about algorithm-related technologies, and in some cases the professionals that use them (Prahl & van Swol, 2017; Promberger & Baron, 2006; Shaffer et al., 2012).

Workman (2005) concluded that people's attitudes towards algorithms are strongly associated with algorithm aversion, finding that more positive attitudes led to greater use of decision support systems. Further, general distrust prevails even when there is no clear indication that other human sources of judgement are superior to those generated from algorithms (Önkal et al., 2009). As such, some have speculated that these attitudes may stem from a persistent belief that algorithms should not be favoured over humans in making judgements or decisions (Kahneman, 2011; Prahl & Swol, 2021). Indeed, we note a strong degree of caution expressed among workers regarding the future of such technologies in the workplace, and in particular, their potential to supplant human workers (Abeliansky & Beulmann, 2021). Thus, it is reasonable to speculate that people may possess a pervasive prejudice against algorithmic aids, which results in a predilection to resist adhering to their recommendations (Arkes et al., 2007; Dietvorst et al., 2015; Sutherland et al., 2016). It is likely that such attitudes originate from algorithmic 'folk theories' (i.e., personal conceptions of how they work), which are frequently negative and/or erroneous (Ytre-Arne, & Moe, 2021).

Attitudes are personal predispositions to think or feel about something in a particular way and can be favourable or unfavourable in nature (Ajzen, 2001). They are linked to both stable and dynamic individual factors (Thatcher & Perrewe, 2002), and can differ in relation to the extent to which people are consciously aware of them (i.e., explicit or implicit in nature; Gawronski & Payne, 2010; Greenwalkd & Banaji, 1995). While some have explored attitudes that stem from the perceived usefulness of specific systems (Gregor & Benbasat, 1999; Taylor & Todd, 1995), no empirical research to date has examined the potential association between individuals' pre-existing, generalised attitudes (either explicit or implicit) towards new technology, or algorithms specifically, and their adherence to algorithmic aids.

Study Aims and Research Questions

The study employed a simulation of an algorithmic aid (used previously by Dietvorst et al., 2018), which was designed to assist users in the formulation of forecasts (predictions of students' mathematics test scores). Participants had no prior experience in forecasting mathematics test scores, which meant that differences in operational experience would not be a confounding influence in the design. A relatively high degree of explainability was present, with participants informed about the features the algorithm used in making its calculations.

We aimed to test whether participants' adherence to the algorithmic aid was related to the degree of control given to them during system use (i.e., the extent to which they could revise the forecasts generated by the system; full, moderate, or restricted control) and their attitudes towards new technologies (explicit) and algorithms (implicit). Further, we tested the relationship between variations of user control when using the algorithmic aid and participants' subjective reports of task demands, including their perceived mental demand, temporal demand, effort, and frustration level. In this research we aimed to investigate several exploratory research questions:

- Research Question 1 (RQ_1). Are participant characteristics (age, gender, computer literacy) associated with their perceptions of task demands when using an algorithmic aid?
- Research Question 2 (RQ_2). Are participant characteristics (age, gender, computer literacy) associated with their adherence to an algorithmic aid?

- Research Question 3 (RQ_3). What impact does variation in user control have on participants' adherence to an algorithmic aid?
- Research Question 4 (RQ_4). What impact does variation in user control have on participants' perceptions of task demands when using an algorithmic aid?
- Research Question 5 (RQ_5). Are participants' pre-existing (explicit) attitudes towards new technologies and/or (implicit) attitudes towards algorithms associated with their adherence to an algorithmic aid?
- Research Question 6 (RQ_6). Are participants' pre-existing (explicit) attitudes towards new technologies and/or (implicit) attitudes towards algorithms associated with their perceptions of task demands when using an algorithmic aid?

Method

Participants

Participants (N = 159) comprised a sample of adults (1 other, 71 male, and 87 female) residing predominantly in Canada (29.6%), Great Britain (27%), United States (23.3%), Australia (5.7%), and New Zealand (4.4%) (an additional 11 countries comprised the remaining 10% of the sample). Participants were recruited via the Prolific recruitment platform (https://www.prolific.co/) and were monetarily reimbursed for their time spent completing the experiment. Participants ranged in age from 18 to 62 years (M_{age} = 34.47, SD_{age} = 10.53). Approval for the conduct of the experiment was granted by Charles Sturt University's Human Ethics Research Committee (HREC Approval #H20231).

Materials and Procedure

An online experiment was employed. The experiment utilised a modified version of the forecasting task and associated algorithm from Dietvorst et al. (2018) (adapted for an Australian population). The task required participants to forecast school students' scores on a standardised mathematics test based on the student's profile, which comprised six variables (socioeconomic status, desired occupation at age 30, predicted highest degree, completed university preparation courses, how many friends not going to university, and favourite school subject). Participants were presented with 20 different high school student profiles and were asked to use their profile information to predict their percentage score on a standardised mathematics test. Additionally, participants received an algorithm-generated forecast of each student's score. Participants were informed that the forecast had been made on the same student profile information. They were told that the algorithm's forecast was intended to help guide their own forecasts; however, they would have an opportunity to modify their forecasts by sliding the score bar to the left to lower the score or right to raise the score. Two sets of 10 scenarios were used; one in which the system's accuracy deviated from the true, real-world outcome on average by 28.2% (high error-rate), and one that deviated by only 8.35% (low error-rate). This was designed to test whether participants' degree of deviation from the system's forecast was a function of the perceived accuracy of the algorithm.

Participants were randomly allocated into either the 'full control', 'moderate control', or 'restricted control' condition. Those in the 'full control' condition were given guidance that they could adjust the forecast as much as they wished. Participants in the 'moderate control' condition were able to adjust the forecast by approximately 30 per cent, while those in the 'restricted control' condition could adjust them by only 2 per cent. To control for potential learning effects in the current design, participants did not receive explicit feedback on their forecast accuracy. To ensure that participants were motivated to be as accurate as possible in formulating their forecasts, they were notified of a small bonus payment should their forecast performance be in the top 5% of participants. However, unbeknown to them, in the interest of fairness all participants received the additional payment at the conclusion of the study.

Following completion of the forecasting task, participants completed an assessment of several components of subjective task load, providing ratings on four separate (21-point) subscales from the NASA Task Load Index (NASA-TLX; Hart, 1988, 2006); perceived mental demand, temporal demand, effort, and frustration. They also completed a measure of their explicit attitudes towards new technologies (Technology Readiness Index (TRI 2.0); Parasuraman & Colby, 2015). These measures have been shown to have stable psychometric properties (Braarud, 2021; Parasuraman & Colby, 2015). Next, they completed demographic items (age, gender, and computer literacy) on the QuestionPro (2021) platform. Finally, participants were directed to the Inquisit platform (Inquisit 6, 2021) where they completed a novel version of the Go/No-Go Association Task (GNAT; Nosek & Banaji, 2001), which was designed to measure their implicit attitudes towards algorithms.

The GNAT measures an implicit attitude by assessing the strength of association between a target category (e.g., Algorithm) and two ends of an attribute dimension (e.g., Positive and Negative). The strength of association is assessed by the degree to which items related to the target category (e.g., Artificial Intelligence, Machine Learning, Computers, etc.) or attribute (e.g., Positive) can be discriminated from distracter items that are not related to those concepts (Nosek & Banaji, 2001). The first condition requires that participants simultaneously identify stimuli that represent the target category (Algorithm) and attribute (Positive), while the second condition comprises the same target category (Algorithm) and the alternative attribute (Negative). The extent to which algorithms are associated with positive versus negative attributes is reflected in the relative ease of discriminating algorithm-related items with the positive versus the negative attributes. In this study, we hoped that algorithms would be generally regarded positively because of their associated benefits. If so, participants' accuracy in discriminating algorithm-related items and positive items from distracters ought to be higher than their accuracy in discriminating algorithm-related items and negative items from distracters. The difference in accuracy between these conditions is taken as a measure of implicit attitude toward algorithms.

The GNAT requires the same response 'go' (the participant presses the space bar) to items that are related to the target category (e.g., Algorithm) and/or a particular evaluative attribute (e.g., Positive). This 'go' response represents the signal. The 'no-go' response (the participant does not press any key) is required when the items presented do not belong to the target category or attribute (i.e., the noise). Participants with greater accuracy in the positive compared to negative condition are thought to possess relatively positive implicit attitudes towards algorithms, while those with greater accuracy on the negative condition are thought to possess relatively negative implicit attitudes.

The study design contained limited disclosure, in that participants were advised that the study focused on how individuals make predictions based on limited objective information but did not detail the intention to investigate how the role of user control or attitudes may impact compliance with algorithmic outputs. Upon completion of the study, participants were provided a written debrief statement that detailed these specific aims and experimental manipulations.

Design

The study employed an experimental design whereby participants were randomly assigned to one of three conditions when engaging the algorithmic aid in the forecasting task; 'full control' (adjust the algorithm aid's forecast as much as they wish), 'moderate control' (adjust by 30 per cent), or 'restricted control' (adjust by 2 per cent) group. The impacts of this manipulation were tested in relation to two outcome variables; participants' adherence to the algorithmic aid (RQ_3) and their perceptions of task demands when using the aid (RQ_4). Additionally, tests of association were conducted between: participant characteristics (age, gender, computer literacy) and their perceptions of task demands (RQ_1) and adherence to an algorithmic aid (RQ_2); participants' pre-existing (explicit) attitudes towards new technologies and (implicit) attitudes towards algorithms, and their adherence to the algorithmic aid (RQ_5), and their perceptions of task demands when using the aid (RQ_6).

Results

Data Reduction and Analytical Strategy

All scale item responses (including the four NASA-TLX subscales and TRI-2) were reverse-scored where appropriate and reduced to a mean score for each participant (including mean scores for each of the four NASA-TLX subscales). For the GNAT, participants' correct (i.e., hit) and incorrect (i.e., miss) responses were tallied for the positive and negative conditions. A *d*-prime score was calculated for each participant in both the positive and negative conditions. This was done by standardising tallied scores and subtracting the transformed 'misses' from the transformed 'hits'; providing an indication of relative accuracy in classifying concepts. Those with higher scores in the positive compared to negative condition are thought to possess relatively positive attitudes towards new technologies, while those with higher scores on the negative compared to positive condition are thought to possess relatively negative attitudes. An overall difference score was calculated by subtracting the *d*-prime 'negative' from the *d*-prime 'positive'. The higher the resultant difference score, the more positive the implicit attitude towards algorithms.

The quantitative variables included: participants' age, gender (coded as female, male, other (list), and other prefer not to say), mental demand, temporal demand, effort, frustration, explicit attitudes towards new technology, implicit attitudes towards new technology, average deviation from algorithm forecasts with relatively low and high error-rates (in %), and average deviation from all algorithm forecasts (in %). We tested the quantitative data for normality using Shapiro-Wilk tests, which revealed that all variables except explicit attitudes were not normally distributed. As such, non-parametric Spearman's rank correlational analyses were employed for tests of association. Cohen's (1988) convention for interpreting effect sizes was used to guide interpretation (Small .10, Medium .30, Large .50). For tests of difference across conditions (e.g., differences based on user control), analyses of variance (ANOVA) were employed as they are considered more powerful than non-parametric alternatives and robust to violations of normality. Any individual outliers (i.e., participant data points greater than three standard deviations from the mean) were transformed to the next most extreme (non-outlier) value (Hills, 2011). There were no missing data and all 159 cases were retained for analysis. For comparisons of deviation and task load, Levene's test revealed unequal variances. However, ANOVA was expected to be robust for deviations from this assumption as groups were roughly equal and the largest variance was not more than four to five times the smallest variance (Field, 2013).

User characteristics

A series of Spearman's rank correlational analyses were computed to test the relationship between users' characteristics (i.e., age, gender, and computer literacy), their perception of task demands when using the algorithm (i.e., mental demand, temporal demand, effort, and frustration) (RQ_1), their attitudes towards new technologies (explicit) and algorithms (implicit), as well as the average amount that they deviated from the algorithm's forecast (RQ_2). The relative influence of the degree of control given to users in formulating their final forecast (which was assigned randomly by the research team) was controlled for in the correlation. Gender was reduced to a dichotomous variable (excluding the participant who identified as 'other' as the small sample size precluded meaningful analysis).

The results revealed an association between participants' gender and their computer literacy $r(155) = .352$, $p < .001$ (medium effect), whereby male participants tended to rate themselves as possessing a higher degree of computer literacy than their female counterparts. Both variables were also associated with explicit attitudes towards technology, with male participants ($r(155) = .214$, $p = .007$; small effect) and those with higher computer literacy ($r(155) = .326$, $p < .001$; medium effect) both reporting more positive attitudes towards technology on the TRI-2 scale. Finally, participants' gender also had a small association with both mental ($r(155) = .227$, $p = .007$; small effect) and temporal demand ($r(155) = .198$, $p = .013$; small effect), with male participants tending to report a greater level of demand during forecasting, irrespective of their assigned control condition.

Algorithm accuracy

Descriptive analyses revealed that across all conditions participants deviated from the algorithm's forecast by 8.5% on average (N.B., a comparison across different user control conditions is provided below in the 'User Control' section). As previously noted, two sets of decision scenarios were used; one in which the algorithm's accuracy deviated from the true, real-world outcome on average by 28.2%, and one that deviated by only 8.35%. A Wilcoxon Signed-Rank test compared users' mean deviations from both sets of forecasts and revealed no significant difference in participants' average deviation, $z(N = 159) = .901$, $p = .368$. Critically, this meant that participants' degree of deviation from the algorithm's forecasts was not driven by a perceived difference in the algorithm's accuracy. This result also justified the use of a total deviation average for the remaining analyses.

User Control

To test the impact of variation in user control on participants' adherence to an algorithmic aid (RQ_3), a one-way ANOVA was performed. Users' average deviation from the algorithm's forecasts was compared across the three conditions of control, revealing a statistically significant difference, $F(2, 156) = 136.024$, $p < .001$, $\eta^2 = .64$ (large effect). Games-Howell post hoc comparisons revealed that significant differences existed between all levels of control, with the greatest observed difference being between the full control (M = 13.26, SD = 5.35) and the restricted control conditions (M = 1.19, SD = .36), $t(156) = 16.01$, $p < .001$, Cohen's d = 3.18 (very large effect).

Next, ANOVAs were used to compare participants' perception of task demands when using the system (i.e., mental demand, temporal demand, effort, and frustration) across the three conditions of control (RQ_4), revealing a statistically significant difference in perceived frustration only $F(2, 156) = 3.361$, $p = .037$, $\eta^2 = .041$ (small effect). Games-Howell post hoc comparisons revealed that a significant difference was between the full and restricted control conditions, $t(156) = 2.51$, $p = .042$, Cohen's d = 0.48 (medium effect); participants in the restricted control group (M = 17.26, SD = 22.39) reported being significantly more frustrated than those in the full control condition (M = 8.32, SD = 13.40). Participants' average deviation from the algorithm's forecasts and their perceived frustration levels as a function of the relative control provided in each condition, are displayed in Figure 1.

****

Figure 1 About Here

****

User Attitudes

Spearman's rank correlational analyses were used to test for associations between participants' explicit and implicit attitudes toward technology, their adherence to the algorithmic aid (RQ_5), and their perceptions of task demands (RQ_6), including mental demand, temporal demand, perceived effort, and frustration. The relative influence of the degree of control given to users in formulating their forecasts (which was assigned randomly by the research team) was controlled for in the analysis to determine the relative variance in outcomes explained by attitudes alone. That is, it was important to determine the unique contribution of users' attitudes beyond the influence of the degree of control given to them. The results revealed a negative association between participants' deviation from the algorithm's forecasts and their implicit attitudes only (i.e., GNAT performance) ($r(156) = -.150$, $p = .03$; small effect), whereby participants possessing more negative implicit attitudes towards algorithms tended to deviate from the algorithm's forecasts to a significantly greater degree than those with more positive attitudes (RQ_5). No association was found between attitude variables and task demands (RQ_6).

Discussion

Overview

We aimed to test whether participants' adherence to the algorithmic aid was related to the degree of control they were provided at decision point (i.e., full, moderate, or restricted control) and their attitudes towards technologies (explicit) and algorithms (implicit). Further, we tested the relationship between variations of user control when using the algorithmic aid and participants' subjective reports of work demands, including their perceived mental demand, temporal demand, effort, and frustration level.

Our findings indicate that participant characteristics (age, gender, computer literacy) are associated with their perceptions of task demands (RQ_1), but not with their adherence to an algorithmic aid (RQ_2). Specifically, males showed a greater level of perceived demand (mental and temporal) during the decision-making process, irrespective of the level of control they had over the algorithm's forecast at decision point. Additionally, male participants tended to rate themselves as possessing a higher degree of computer literacy than their female counterparts and show more positive attitudes towards technology. Finally, those participants reporting higher levels of computer literacy also reported more positive (explicit) attitudes towards new technology.

We found that variations in user control impacted participants' adherence to an algorithmic aid (RQ_3) and their perceptions of task demands when using the aid (RQ_4). Specifically, participants tended to deviate from the algorithm's recommendation more greatly when they were given more control over the final outcome. They also typically experienced greater levels of frustration when they had very little freedom to modify the algorithm's forecast. Finally, our findings revealed that while participants' pre-existing (explicit) attitudes towards new technologies were not associated with their adherence to an algorithmic aid (RQ_5) or their perceptions of task demands (RQ_6), their implicit attitudes towards algorithms were related to adherence (RQ_5). Indeed, participants with more positive implicit attitudes towards algorithms were less prone to deviate from the algorithm's recommendations at decision point. The theoretical and practical applications are discussed below.

User characteristics

While studies examining gender differences in computer literacy remain sparse (Gebhardt et al., 2019), there are compelling data for gender-based differences in computer technology performance, particularly in student populations. For instance, Fraillon et al. (2014) reported that the performance of female students was substantially higher than that of male students in 12 out of the 14 countries they examined. However, such performance differences are rarely shown to be aligned to frequency of use (Punter et al., 2017), interest, or confidence (Meelissen & Drent, 2008), all of which tend to be greater among male computer users. Similarly, the current finding that males showed more positive attitudes than females towards computers has precedent in the evidence base (OECD, 2011).

Critical to our current investigation, neither gender nor computer literacy were associated with participants' relative deviation from algorithm-generated forecasts. Interestingly though, male participants did tend to report greater levels of mental and temporal demand during forecasting. Such gender-based differences may stem from other associated differences in areas such as problem-solving, motivation, and preferred decision process, the latter of which would seem particularly relevant in this study. Mann et al. (1989) posits that decision 'style' may affect the way decisions are formulated, the goals that are set, the information that is collected, and the methods used to select strategies. Indeed, empirical work has established differences in users' preference for information search strategies when using decision support interfaces, which can significantly impact users' perceptions of task demands (Morrison et al., 2010; Perry et al., 2013). Importantly, in these studies, differences in users' preference were not associated with differences in decision quality, meaning that the incorporation of user preferences into decision aid design could improve user experience without jeopardizing the integrity of the system. In the current study, as the output from a simulated algorithmic aid appeared to act as an anchor point from which participants forecasts were formulated (discussed more below), participants' adjustment from the anchor may be impacted if starting from a decision point that does not align with their preferred strategies. Finally, decision aids have historically been constructed largely upon the assumption of an androgynous user (Powell & Johnson, 1995) and there has been a dearth of investigation examining individual differences and their role in human-machine compatibility and acceptance, particularly for emerging 'smart' machines. As such, user gender (along with other potentially key differences not explored here; e.g., training, experience, context-specific issues, and organisational influences) represent important lines of inquiry in human-centered algorithmic aid design.

Algorithm Accuracy

Participants in the experiment showed a willingness to deviate from the algorithm's forecasts across all conditions of user control. As participants were not provided feedback regarding the algorithm's accuracy, nor were there differences in response across relatively high- and low-error algorithm forecasts, participants' degree of deviation was not a function of the accuracy of the algorithm, or their perceptions of its accuracy. This suggests that they were operating under a high degree of uncertainty. In such situations where decision-makers are faced with a high degree of task complexity, incomplete or missing information, and limited experience, they will regularly increase their reliance on mental heuristics (Tversky & Kahneman, 1974). In particular, it would appear that participants in the current study employed the anchoring and adjustment heuristic (Dale, 2015; Tversky & Kahneman, 1974), whereby their judgements were 'anchored' to a reference point that influenced their subsequent forecast. Reliance on such anchoring will presumably encourage adherence and a lower degree of deviation from the anchor, however, participants in our study still demonstrated a willingness to adjust their judgements.

As algorithmic aids are likely to be used most frequently by relatively less experienced workers who are less certain of the parameters of the job and concomitant moderators of performance, it will be important to consider how the degree of user uncertainty may impact algorithm acceptance. One strategy shown to improve adherence and appreciation of algorithm-generated outputs in uncertain environments has been to increase the degree of granularity associated with the relative explainability of the system (David et al., 2021); this was kept constant in the current study with all participants provided featurebased explanations of how the system worked. However, acceptance can be further promoted via accuracy-based feedback, which involves presenting users with information regarding the historical accuracy of the system. For instance, Saragih and Morrison (2021) found that users were more likely to accept algorithm-generated advice (deviating less from the algorithm) when they were informed of a superior (expert) level of performance by the system. Similarly, David et al. (2021) highlighted benefits from accuracy-based feedback, particularly in maintaining users' trust after system error.

User Control, Algorithm Deviation, and Task Demands

Participants given full control to adjust the algorithm's forecast were found to deviate from it more substantially than those given a moderate or restricted degree of control. This finding has significant implication to algorithmic aid use, as it suggests that, under conditions of uncertainty, users will exploit their control in ways that may adversely impact decision quality. This would suggest that a degree of user restriction may be required to promote greater adherence in such conditions. However, the findings also imply that such restrictions may come at a cost to worker mental health and wellbeing, with those workers in the most restricted condition experiencing significantly greater frustration compared to those fully in control.

While no difference in frustration was found between moderate and restricted degrees of control, deviation under such conditions did differ; with users showing greater deviation with increased control. Although Dietvorst et al. (2018) found that the capacity to modify a forecast increased the participants' satisfaction with the forecasting process, as was the case here, they also found that individuals who could only restrictively adjust the forecast were not significantly less satisfied than those who had greater levels of control over adjustments. Their findings suggest that giving an operator a level of control in the forecasting process when interacting with algorithmic aids, even if that level of control is minimal, increases their satisfaction in using such systems. In our study, such increased satisfaction (inferred from decreased frustration), could only materialise when participants were given full control over the final outcome.

The findings regarding deviation and frustration need to be considered concurrently as the issues are entangled. Increasing control may have differential effects on individual and system outcomes; on one hand reduced frustration and presumably better long-term outcomes at an individual worker level, while on the other hand, an increased potential for unreliability and error as a result of non-compliance with the technology. Further, any reduction in human input into algorithmic aids may have broad and significant implications to decision quality.

For instance, humans are still regularly considered to be the "gold standard" for assessing data that cannot easily be interpreted and trained for machine learning models, such as creativity (Baer & McKool 2014; Cheng & Bernstein, 2015). Humans are particularly adept at providing subjective judgements of data that are difficult to measure objectively through statistical techniques (Cheng &Bernstein 2015; Einhorn, 1974). Additionally, human experts have highly organised domain-specific knowledge that enables them to recognise and interpret non-routine data using 'skilled intuition' (Kahneman & Klein, 2009), which may represent difficult to predict outliers in statistical models (Blattberg & Hoch, 1990). Therefore, we recognise the value of augmenting algorithms with human skills and knowledge, and so do not argue for the exclusion of humans from all workplace systems. Instead, we posit that system designers should provide the user a balanced degree of control in system operation; that is, the maximum extent of user control at which there are no significant deteriorations in system performance. Further, we suggest that non-control related impediments to user satisfaction (e.g., interface compatibility issues discussed above, and attitude-related issues described below) should be identified and minimised.

User Attitudes

Our findings revealed that some of the variance in participants' tendency to deviate from the algorithm's forecast could be attributed to their pre-existing implicit attitudes towards algorithms. In this instance we found no role for participants' conscious (explicit) attitudes towards new technologies in their interactions or perceptions of the algorithm. This may have significant implications to the likely effectiveness of different interventions addressing users' attitudes.

The attempt to quantify implicit attitudes and relate them to instances of behavioural interactions with algorithmic aids is the first of its kind in the extant literature and represents a significant contribution to our understanding of human-algorithm compatibility. The finding that implicit attitudes are related to algorithmic adherence outside of the impacts of varied degrees of user control, underlines an alternative avenue for increasing users' adherence to algorithmic aid outputs. This is particularly important as we predict that most systems will require some degree of restriction on user inputs and adjustments to maintain the reliability and integrity of the system. It may also be possible to improve individual level workplace demands indirectly by improving attitudes towards the technology.

Critically, we must note that the measurement and modification of implicit attitudes has a long and controversial history in the social sciences (see Meissner et al., 2019 for a recent review), and the techniques used here require further validation in this context prior to their implementation in practice. One attractive avenue for investigation is the use of implicit bias training among prospective algorithm users. Such training aims to expose workers to their implicit biases (in this case, towards algorithms), and provide strategies to adjust their existing (non-conscious and automatic) patterns of thinking. Creating awareness of workers' implicit biases is key, as there is often very little realisation of their existence, which may partly explain why we failed to reveal an association between workers' non-conscious, implicit attitudes toward algorithms, and their conscious, explicit attitudes toward new technologies. Recent meta-analytic evidence for implicit bias interventions (Lai, 2014) revealed that counterstereotype training was the most effective method for reducing or eliminating implicit bias. This technique involves identifying and reinforcing counterstereotypes, which might involve workers imagining positive instances of human-algorithm interactions, increasing knowledge in relation to their development and inner workings, as well as acquiring firsthand experiences with effective algorithmic aids in their own workplace.

Conclusion

This study comprised a novel exploration of numerous user- and system-related factors that may contribute to users' acceptance of algorithmic aids. Primarily, our findings allude to the importance of users' control over the algorithm's outputs in their acceptance of, and frustration in using a novel forecasting algorithm. Further, they suggest that users' acceptance of such systems may be partly influenced by their existing implicit attitudes towards algorithms, which may ultimately contribute to their intention to comply with/use them in the workplace. The study employed an experimental design that is largely consistent with the extant literature on algorithm aversion. This work should now be extended to incorporate alternative populations, methods and paradigms (e.g., the observation of interactions in natural settings a la' the Naturalistic Decision Making movement; Klein, 2008), which will enable the triangulation and validation of the current procedures and findings, extending their generalisability to algorithmic aids implemented in complex and uncertain work settings.

References

Abeliansky, A. and Beulmann, M. (2021). Are They Coming for Us? Industrial Robots and the Mental Health of Workers. *SSRN*. http://dx.doi.org/10.2139/ssrn.3438287

Ajzen, I. (2001). Nature and operation of attitudes. *Annual Review of Psychology, 52*, 27–58. https://doi.org/10.1146/annurev.psych.52.1.27

Arkes H.R. (2001). Overconfidence in Judgmental Forecasting. In: Armstrong J.S. (eds) Principles of Forecasting. International Series in Operations Research & Management Science, vol 30. Springer, Boston, MA. https://doi.org/10.1007/978-0-306-47630-3_22

Baer, J., & McKool, S.S. (2014). The Gold Standard for Assessing Creativity. International *Journal of Quality Assurance in Engineering and Technology Education, 3*, 81-93. https://doi.org/10.4018/ijqaete.2014010104

Blattberg, R. C., & Hoch, S. J. (1990). Database Models and Managerial Intuition: 50% Model + 50% Manager. *Management Science, 36*(8), 887-899. https://doi.org/10.1287/mnsc.36.8.887

Braarud, P. Ø. (2021). Investigating the validity of subjective workload rating (NASA TLX) and subjective situation awareness rating (SART) for cognitively complex human–machine work. *International Journal of Industrial Ergonomics, 86*(14), 103233.

Brougham, D., & Haar, J. (2018). Smart Technology, Artificial Intelligence, Robotics, and Algorithms (STARA): Employees' perceptions of our future workplace. *Journal of Management & Organization 24*, 239–257. https://doi.org/10.1017/jmo.2016.55

Burton, J.W., Stein, M., & Jensen, T.B. (2019). A systematic review of algorithm aversion in augmented decision making. *Journal of Behavioral Decision Making, 33*(2), 1-20. https://dx.doi.org/10.1002/bdm.2155

Cheng, J. & Bernstein, M. S. (2015). Flock: Hybrid Crowd-Machine Learning Classifiers. In Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing (CSCW '15). Association for Computing Machinery, New York, NY, USA, 600–611. https://doi.org/10.1145/2675133.2675214

Cohen, J. (1988). *Statistical Power Analysis for the Behavioural Sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.

Dale, S. (2015). Heuristics and biases: The science of decision-making. *Business Information Review, 32*(2), 93–99. https://doi.org/10.1177/0266382115592536

David, D. B., Resheff, Y. S., & Tron, T. (2021). Explainable AI and Adoption of Financial Algorithmic Advisors: An Experimental Study. In *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society (AIES '21).* Association for Computing Machinery, New York, NY, USA, 390–400. https://doi.org/10.1145/3461702.3462565

Dawes, R. M., Faust, D., & Meehl, P. E. (1989). *Clinical Versus Actuarial Judgment. Science, 243*(4899), 1668-1674. https://doi.org/doi:10.1126/science.2648573

de Lange, A. H., Taris, T. W., Kompier, M. A., Houtman, I. L., & Bongers, P. M. (2003). "The very best of the millennium": longitudinal research and the demand-control-(support) model. *Journal of Occupational Health Psychology, 8*(4), 282–305. https://doi.org/10.1037/1076-8998.8.4.282

Dievorst, 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://dx.doi.org/10.1037/xge0000033

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2018). Overcoming algorithm aversion: People will use imperfect algorithms if they can (even slightly) modify them. *Management Science, 64*(3), 1155–1170. https://doi.org/10.1287/mnsc.2016.2643

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

Field, A. (2013). *Discovering statistics using SPSS* (4th ed.). Sage publications.

Fraillon, J., Ainley, J., Schulz, W., Friedman, T., & Gebhardt, E. (2014). Preparing for Life in a Digital Age: the IEA International Computer and Information Literacy Study 2013 International Report. *Springer Open*. https://research.acer.edu.au/ict_literacy/8

Gawronski, B., & Payne, B. K. (Eds.). (2010). *Handbook of implicit social cognition: Measurement, theory, and applications*. The Guilford Press.

Gebhardt, E., Thomson, S., Ainley, J., & Hillman, K. (2019). Gender Differences in Computer and Information Literacy: An In-depth Analysis of Data from ICILS. *Springer Open*. https://doi.org/10.1007/978-3-030-26203-7

Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review, 102*(1), 4–27. https://doi.org/10.1037/0033-295X.102.1.4

Gregor, S., & Benbasat, I. (1999). Explanations from Intelligent Systems: Theoretical Foundations and Implications for Practice. *MIS Quarterly, 23*(4), 497–530. https://doi.org/10.2307/249487

Grove, W. M., Zald, D. H., Lebow, B. S., Snitz, B. E., & Nelson, C. (2000). Clinical versus mechanical prediction: a meta-analysis. *Psychological Assessment*, *12*(1), 19-30.

Hart, S. G. (2006). NASA-Task Load Index (NASA-TLX); 20 Years Later. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 50*(9), 904-908. https://doi.org/10.1177/154193120605000909

Hart, S. G., & Staveland, L. E. (1988). *Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research*. In P. A. Hancock & N. Meshkati (Eds.), Advances in Psychology (Vol. 52, pp. 139-183). North-Holland.

Herm, L., Heinrich, K., Wanner, J., & Janiesch, C. (2022). Stop ordering machine learning algorithms by their explainability! A user-centered investigation of performance and explainability. *International Journal of Information Management.* https://doi.org/10.1016/j.ijinfomgt.2022.102538

Hills, A. (2011). *Foolproof guide to statistics using IBM SPSS*. Pearson Australia.

Innes, J. M., & Morrison, B. W. (2021). *Artificial intelligence and psychology*. In A. Elliott (Ed.), The Routledge social science handbook of AI (pp. 30-57). (Routledge International Handbooks). Routledge.

Inquisit 6 [Computer software]. (2021). Retrieved from https://www.millisecond.com

Johnson, A., Dey, S., Nguyen, H., Groth, M., Joyce, S., Tan, L., Glozier, N., & Harvey, S. B. (2020). A review and agenda for examining how technology-driven changes at work will impact workplace mental health and employee well-being. *Australian Journal of Management, 45*(3), 402–424. https://doi.org/10.1177/0312896220922292

Kahneman, D. (2011). Thinking, fast and slow. Farrar, Straus and Giroux.

Kahneman, D., & Klein, G. (2009). Conditions for Intuitive Expertise: A Failure to Disagree. *American Psychologist, 64*(6), 515–526. https://doi.org/10.1037/a0016755

Klein, G. (2008). Naturalistic Decision Making. *Human Factors, 50*(3), 456–460. https://doi.org/10.1518/001872008X288385

Lai, C. (2014). Reducing implicit racial references: I. A comparative investigation of 17 interventions. *Journal of Experimental Psychology, 143*(4), 1765–1785. https://doi.org/10.1037/a0036260

Mahmud, H., Islam, A. K. M. N., Ahmed, S. I., & Smolander, K. (2022). What influences algorithmic decision-making? A systematic literature review on algorithm aversion. *Technological Forecasting and Social Change*, *175*, 121390. https://doi.org/https://doi.org/10.1016/j.techfore.2021.121390

Mann, R., Watson, H., Cheney, P. and Gallagher, C. (1989). Accommodating Cognitive Style through DSS Hardware and Software in Sprague, R. and Watson, H. (Eds), Decision Support Systems: Putting Theory into Practice, Prentice Hall, pp. 118-130.

McClure, P. K. (2018). "You're Fired," Says the Robot: The Rise of Automation in the Workplace, Technophobes, and Fears of Unemployment. *Social Science Computer Review, 36*(2), 139–156. https://doi.org/10.1177/0894439317698637

Meehl, P. E. (1954). *Clinical versus statistical prediction.* Minneapolis, MN: University of Minnesota Press.

Meelissen, M. R. M., & Drent, M. (2008). Gender differences in computer attitudes: does the school matter? *Computers in Human Behavior, 24*(3), 969-985. https://doi.org/10.1016/j.chb.2007.03.001

Meissner F., Grigutsch, L. A., Koranyi N., Müller, F., & Rothermund, K. (2019). Predicting Behavior with Implicit Measures: Disillusioning Findings, Reasonable Explanations, and Sophisticated Solutions. *Frontiers in Psychology, 10*. https://doi.org/10.3389/fpsyg.2019.02483

Morrison, B. W., Wiggins, M. W., & Porter, G. (2010). User preference for a control-based reduced processing decision support interface. *International Journal of Human-Computer Interaction, 26*(4), 297-316. https://doi.org/10.1080/10447310903575465

Murray, W.C., & Rostis, A. (2007). "Who's Running the Machine?" A Theoretical Exploration of Work Stress and Burnout of Technologically Tethered Workers. *Journal of Individual Employment Rights, 12*, 249-263. https://doi.org/10.2190/IE.12.3.F

Nosek, B. A., & Banaji, M. R. (2001). The Go/No-go Association Task. *Social Cognition, 19*(6), 625–666. https://doi.org/10.1521/soco.19.6.625.20886

OECD (2011), Education at a Glance 2011: OECD Indicators, OECD Publishing. http://dx.doi.org/10.1787/eag-2011-en

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

Parasuraman, A., & Colby, C. L. (2015). An Updated and Streamlined Technology Readiness Index: TRI 2.0. *Journal of Service Research, 18*(1), 59–74. https://doi.org/10.1177/1094670514539730

Perry, N. C., Wiggins, M. W., Childs, M., & Fogarty, G. (2013). The Application of Reduced-Processing Decision Support Systems to Facilitate the Acquisition of Decision-Making Skills. *Human Factors, 55*(3), 535–544. https://doi.org/10.1177/0018720812467367

Powell, P. L., & Johnson, J.E.V. (1995). Gender and DSS design: the research implications. *Decision Support Systems, 14*(1), 27-58. https://doi.org/10.1016/0167-9236(94)00014-J

Prahl, A., & Van Swol, L. (2021). Out With the Humans, in With the Machines?: Investigating the Behavioral and Psychological Effects of Replacing Human Advisors With a Machine. *Human-Machine Communication, 2*, 209-234. https://doi.org/10.30658/hmc.2.11

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

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

Punter, R., Meelissen, M., & Glas, C. (2017). Gender differences in computer and information literacy: An exploration of the performances of girls and boys in ICILS 2013. *European Educational Research Journal, 16*(6), 762–780. https://doi/10.1177/1474904116672468

QuestionPro. [online] Available at: <https://www.questionpro.com/blog/research-design/> [Accessed 2021].

Saragih, M., & Morrison, B. W. (2021). The Effect of past Algorithmic Performance and Decision Significance on Algorithmic Advice Acceptance. *International Journal of Human–Computer Interaction, 38*(13), 1228-1237. https://doi.org/10.1080/10447318.2021.1990518

Schiliro, F., Beheshti, A., & Moustafa, N. (2021). A novel cognitive computing technique using convolutional networks for automating the criminal investigation process in policing. In K. Arai, S. Kapoor, & R. Bhatia (Eds.), Intelligent Systems and Applications: Proceedings of the 2020 Intelligent Systems Conference (IntelliSys) Volume 1 (Vol. 1, pp. 528-539). (Advances in Intelligent Systems and Computing; Vol. 1250 AISC). Springer, Springer Nature.

Shaffer, V. A., Probst, C. A., Merkle, E. C., Arkes, H. R., & Medow, M. A. (2012). 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

Somashekhar, S. P., Sepúlveda, M. J., Puglielli, S., Norden, A. D., Shortliffe, E. H., Rohit Kumar, C., Rauthan, A., Arun Kumar, N., Patil, P., Rhee, K., & Ramya, Y. (2018). Watson for Oncology and breast cancer treatment recommendations: agreement with an expert multidisciplinary tumor board. *Annals of Oncology: Official Journal of the European Society for Medical Oncology, 29*(2), 418–423. https://doi.org/10.1093/annonc/mdx781

Sutherland, S. C., Harteveld, C., & Young, M. E. (2016). Effects of the Advisor and Environment on Requesting and Complying With Automated Advice. *ACM Transactions on Interactive Intelligent Systems, 6*(4), 1-36. https://doi.org/10.1145/2905370

Taylor, S., & Todd, P. A. (1995). Understanding Information Technology Usage: A Test of Competing Models. *Information Systems Research, 6*(2), 144–176. http://www.jstor.org/stable/23011007

Thatcher, J. B., & Perrewé, P. L. (2002). An Empirical Examination of Individual Traits as Antecedents to Computer Anxiety and Computer Self-Efficacy. *MIS Quarterly, 26*(4), 381–396. https://doi.org/10.2307/4132314

Tversky, A., & Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases. *Science, 185*(4157), 1124–1131. https://doi.org/10.1126/science.185.4157.1124

Workman, M. (2005). Expert decision support system use, disuse, and misuse: a study using the theory of planned behavior. *Computers in Human Behavior, 21*(2), 211-231. https://doi.org/10.1016/j.chb.2004.03.011

Ytre-Arne, B., & Moe, H. (2021). Folk theories of algorithms: Understanding digital irritation. *Media, Culture & Society, 43*(5), 807–824. https://doi.org/10.1177/0163443720972314

Zarghami, S.A., & Dumrak, J. (2021). Implications of artificial intelligence for bushfire management. *Australian Journal of Emergency Management, 36*(1), 84-91. https://doi.org/10.47389/36.1.84

List of Figures (with legend)

Figure 1.

Participants' mean (error bars show standard errors) deviation from the algorithm's forecasts (%) and their perceived frustration levels across the three user control conditions (i.e., full, moderate, and restricted).

Acknowledgments

This project was resourced by the Centre for Work Health and Safety, which is funded by the New South Wales Government of Australia.