informs
http://pubsonline.informs.org/journal/mnsc/

MANAGEMENT SCIENCE
Vol. 64, No. 3, March 2018, pp. 1155–1170
ISSN 0025-1909 (print), ISSN 1526-5501 (online)

Overcoming Algorithm Aversion: People Will Use Imperfect Algorithms If They Can (Even Slightly) Modify Them

Berkeley J. Dietvorst,[a] **Joseph P. Simmons,**[b] **Cade Massey**[b]

[a] Booth School of Business, University of Chicago, Chicago, Illinois 60637; [b] The Wharton School, University of Pennsylvania, Philadelphia, Pennsylvania 19104

Contact: berkeley.dietvorst@chicagobooth.edu (BJD); jsimmo@wharton.upenn.edu (JPS); cadem@wharton.upenn.edu (CM)

Received: June 11, 2015
Revised: April 5, 2016; August 5, 2016
Accepted: August 15, 2016
Published Online in Articles in Advance: November 4, 2016

https://doi.org/10.1287/mnsc.2016.2643

Abstract. Although evidence-based algorithms consistently outperform human forecasters, people often fail to use them after learning that they are imperfect, a phenomenon known as *algorithm aversion*. In this paper, we present three studies investigating how to reduce algorithm aversion. In incentivized forecasting tasks, participants chose between using their own forecasts or those of an algorithm that was built by experts. Participants were considerably more likely to choose to use an imperfect algorithm when they could modify its forecasts, and they performed better as a result. Notably, the preference for modifiable algorithms held even when participants were severely restricted in the modifications they could make (Studies 1–3). In fact, our results suggest that participants' preference for modifiable algorithms was indicative of a desire for *some* control over the forecasting outcome, and not for a desire for *greater* control over the forecasting outcome, as participants' preference for modifiable algorithms was relatively insensitive to the magnitude of the modifications they were able to make (Study 2). Additionally, we found that giving participants the freedom to modify an imperfect algorithm made them feel more satisfied with the forecasting process, more likely to believe that the algorithm was superior, and more likely to choose to use an algorithm to make subsequent forecasts (Study 3). This research suggests that one can reduce algorithm aversion by giving people some control—even a slight amount—over an imperfect algorithm's forecast.

History: Accepted by Yuval Rottenstreich, judgment and decision making.
Supplemental Material: Data and the online supplementary materials are available at https://doi.org/10.1287/mnsc.2016.2643.

Keywords: decision making • decision aids • heuristics and biases • forecasting • confidence

Forecasts made by evidence-based algorithms are more accurate than forecasts made by humans.[1] This empirical regularity, supported by decades of research, has been observed in many different domains, including forecasts of employee performance (see Highhouse 2008), academic performance (Dawes 1971, 1979), prisoners' likelihood of recidivism (Thompson 1952, Wormith and Goldstone 1984), medical diagnoses (Adams et al. 1986, Beck et al. 2011, Dawes et al. 1989, Grove et al. 2000), demand for products (Schweitzer and Cachon 2000), and so on (for reviews, see Dawes et al. 1989, Grove et al. 2000, Meehl 1954). When choosing between the judgments of an evidence-based algorithm and a human, it is wise to opt for the algorithm.

Despite the preponderance of evidence demonstrating the superiority of algorithmic judgment, decision makers are often averse to using algorithms, opting instead for the less accurate judgments of humans. Fildes and Goodwin (2007) conducted a survey of 149 professional forecasters from a wide variety of domains (e.g., cosmetics, banking, manufacturing) and found that many professionals either did not use algorithms in their forecasting process or failed to give them sufficient weight. Sanders and Manrodt (2003) surveyed 240 firms and found that many did not use algorithms for forecasting, and that firms that did use algorithms made fewer forecasting errors. This failure to use forecasting algorithms extends beyond the corporate world. Vrieze and Grove (2009) surveyed 183 clinical psychologists and found that only 31% of them used algorithms when making clinical predictions.

Although many professional forecasters fail to use algorithms in practice, recent research has shown that people are not always averse to using algorithms to make predictions. Dietvorst et al. (2015) gave participants the choice of either exclusively using an algorithm's forecasts or exclusively using their own forecasts during an incentivized forecasting task, and they manipulated whether participants had experience with the algorithm prior to making this choice. They found that most participants chose to use the algorithm exclusively when they had no information about the algorithm's performance, suggesting that people are not always averse to exclusive reliance on algorithms. However, participants were much more likely to choose to use human rather than algorithmic forecasts once they had seen the algorithm perform and learned it was imperfect. Participants' failure to use the imperfect algorithm persisted even when they had explicitly seen the algorithm outperform the human's forecasts, and even when they recognized that the algorithm performed better than the human did on average. This suggests that people are reluctant to use superior algorithms that they know to be *imperfect*, a tendency that Dietvorst et al. called *algorithm aversion*.

Forecasters' reluctance to use superior but imperfect algorithms instead of inferior human forecasters represents a major challenge for any organization interested in making more accurate forecasts and better decisions, as well as for any organization that would benefit from persuading its customers to use algorithms. Because many real-world outcomes are far from perfectly predictable, even the best forecasting algorithms typically produce imperfect forecasts. As a result, reluctance to use an imperfect algorithm effectively results in a reluctance to use almost any algorithm after receiving performance feedback. In this paper, we offer an approach for overcoming people's aversion to using imperfect algorithms.

Overcoming Algorithm Aversion

Multiple scholars have theorized that people's reluctance to use algorithms for forecasting stems from an intolerance of inevitable error. Einhorn (1986) proposed that forecasters' intolerance of algorithms arises because although people believe that algorithms will necessarily err, they believe that humans are capable of perfection (also see Highhouse 2008). Moreover, Dietvorst et al. (2015) found that people were less tolerant of the algorithms' (smaller) mistakes than of the humans' (larger) mistakes. These findings do not invite optimism, as they suggest that people will avoid any algorithm that they recognize to be imperfect, even when it is less imperfect than its human counterpart.

If people's distaste for imperfect algorithms is in part driven by an intolerance of inevitable error, then people may be more open to using imperfect algorithms if they are given the opportunity to eliminate or reduce such errors.[2] Thus, people may be more willing to use an imperfect algorithm if they are given the ability to intervene when they suspect that the algorithm has it wrong.[3] Although people's attempts to adjust algorithmic forecasts often make them worse (e.g., Carbone et al. 1983, Goodwin and Fildes 1999, Hogarth and Makridakis 1981, Lim and O'Connor 1995, Willemain 1991), the benefits associated with getting people to use the algorithm may outweigh the costs associated with degrading the algorithm's performance. This is especially likely to be true if there is a limit on how much people can adjust the algorithm. If allowing people to adjust an imperfect algorithm by only a small amount dramatically increases their willingness to use it, then people's judgments will be much more reliant on the algorithm, and much more accurate as a result.

In three studies, we explore when and how forecasters choose to use imperfect algorithms. In so doing, we make five contributions to the literature on algorithm aversion. In Studies 1 and 2, we find that people will choose to use an imperfect algorithm's forecasts substantially more often when they can modify those forecasts, even if they are able to make only small adjustments to those forecasts. In Study 2, we find that people's choices are surprisingly insensitive to how much they are allowed to adjust an imperfect algorithm's forecasts. In Study 3, we find that people are similarly satisfied with adjusting an algorithm's forecasts in a constrained versus unconstrained manner, that forecasters who have the ability to adjust an algorithm's forecasts believe it performs better than those who do not, and that constraining the amount by which people can adjust an algorithm's forecasts leads to better performance in the long run (after feedback).

For each study, we report how we determined our sample size, all data exclusions (if any), all manipulations, and all measures. The exact materials and data from each study are available as online supplementary materials as well as at the following website: https://osf.io/5nz9c/.

Study 1

Methods

Overview. In Study 1, we asked participants to forecast students' scores on a standardized math test from nine variables. All participants had the option of using a statistical model to make their forecasts. Participants were informed that the model was imperfect, off by 17.5 percentiles on average. We manipulated whether or not participants had the option to modify the model's forecasts. Participants were assigned to one of four conditions. Specifically, they were assigned either to a condition in which they chose between using the model's forecasts exclusively or not at all, to one of two conditions in which they were restricted in how much or how frequently they could modify the model's forecasts if they chose to use them, or to a condition in which they received the model's forecasts and could use them as much as they wanted. We expected participants who were restrictively able to modify the model's forecasts, compared with those who had to choose whether or not to use the model's forecasts exclusively or not at all, to be much more open to using an imperfect algorithm and to perform better as a result. We were also curious to see how much weight participants would give to the model's forecasts when they were shown the model's forecasts before forming their own opinion, while being free to deviate from the model's forecasts by as much as they wanted.

Participants. This study was conducted in the Wharton behavioral lab. Participants received $10 for completing one hour of experiments, of which ours was a 20-minute portion. Participants could also earn up to a $5 bonus depending on their forecasting performance. We aimed to recruit over 300 participants for this study, so we ran it in two concurrent lab sessions (the Wharton lab has two separate locations) and collected as many participants as we could. The lab inadvertently allowed 19 participants to participate in the study twice. We dropped these participants' second set of responses from our data. Also, 4 participants exited the study before completing their forecasts, leaving us with a sample of 288 participants who completed their forecasts.[4] The final sample averaged 22 years of age and was 66% female.[5]

Procedures. This experiment was administered as an online survey. Participants began by giving consent and entering their lab identification number. Next, they learned that they would estimate the percentiles of 20 real high school seniors on a standardized math test. They also received a brief explanation of percentiles to ensure that they understood the task. Participants were ensured that the data described real high school students. Participants then read detailed descriptions of the nine variables that they would receive to make forecasts.[6] Figure 1 shows an example of the stimuli and variables.

Participants then learned that analysts had designed a statistical model to forecast students' percentiles. They (truthfully) learned that the model was based on data from thousands of high school seniors, that the model used the same variables that they would receive, that the model did not have any further information, and that it was "a sophisticated model, put together by thoughtful analysts." On the next page, participants learned that the model's estimates for each student were off by 17.5 percentiles on average (i.e., that the model was imperfect). Additionally, they were informed that the model may be off by more or less than 17.5 percentiles for the 20 students that they would be assessing.

Next, participants learned about their incentives. Participants were paid a $5 bonus if their forecasts were within five percentiles of students' actual percentiles on average, and this bonus decreased by $1 for each additional five percentiles of average error in participants' forecasts (this payment rule is reproduced in Appendix A). Thus, participants whose forecasts were off by more than 25 percentiles received no bonus at all. Participants were required to type the following sentences to ensure that they understood the incentives: "During the official round, you will receive additional bonus money based on the accuracy of the official estimates. You can earn $0 to $5 depending on how close the official estimates are to the actual ranks."

Next, participants were assigned to one of four conditions. In the *can't-change* condition, participants learned that they would choose between exclusively using their own forecasts and exclusively using the model's forecasts. In the *adjust-by-*10 condition, participants learned that they would choose between exclusively using their own forecasts and using the model's forecasts, but that they could adjust all of the model's forecasts by up to 10 percentiles if they chose to use the model. In the *change*-10 condition, participants learned that they would choose between exclusively using their own forecasts and using the model's forecasts, but that they could adjust 10 of the model's 20 forecasts by any amount if they chose to use the model. Participants in the *use-freely* condition learned that they would receive the model's forecasts and could use them as much as they wanted when making their 20 forecasts. Participants were required to type a sentence that described their condition to ensure that they understood the procedures.[7]

Finally, participants in the can't-change, adjust-by-10, and change-10 conditions decided whether or not to use the statistical model's forecasts.[8] After

Figure 1. Example of Task Stimuli Used in Studies 1–3

Race	White, non-Hispanic
Socioeconomic status (first = lowest, fifth = highest)	Fifth quintile (highest)
Desired occupation at age 30	Healthcare Practitioners and Technical Occupations
Predicted highest degree	Complete Bachelor's degree
Region of country	South
Times taken PSAT	Twice
How many friends are not going to college	None of them
Favorite school subject	Social studies/history/government/civics
Taken any AP test	No

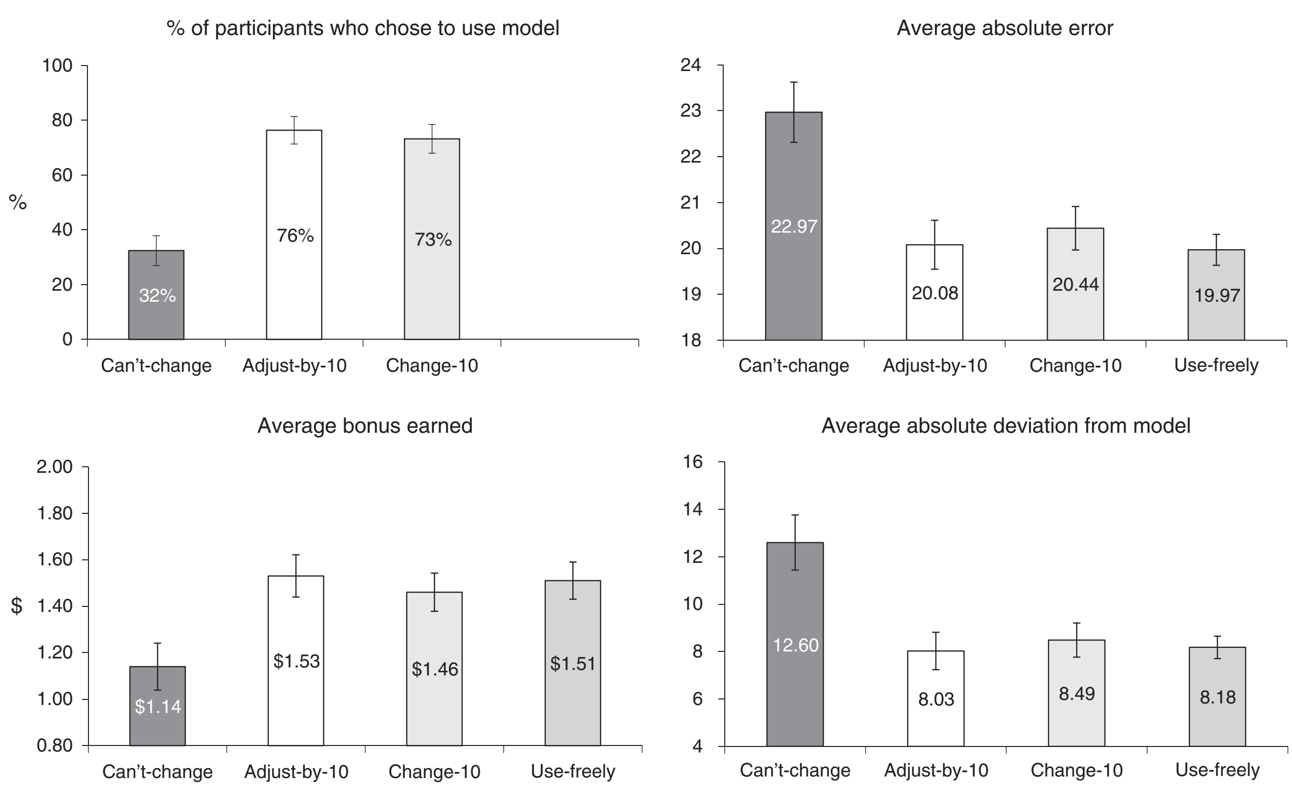
making this choice, participants made 20 incentivized forecasts. The 20 students that participants judged were randomly drawn, without replacement, from a pool of 50 randomly selected high school seniors. Each high school student was presented on an individual page of the survey. Participants in the use-freely condition saw the information describing a student (see Figure 1), saw the model's forecast for that student, and entered their forecast for that student. Participants who chose not to use the model in the can't-change, adjust-by-10, and change-10 conditions made their forecasts without seeing the model's forecasts. Participants in these conditions who chose to use the model entered their own forecasts anyway. In the can't-change conditions, their own forecasts did not determine their payment; in the adjust-by-10 condition, these forecasts were used to determine their payment and were required to be within 10 percentiles of the model's forecasts; and in the change-10 condition, these forecasts were used to determine their payment but could not differ from the model for more than 10 of the forecasts.

After completing the forecasts, participants estimated their own average error and the model's average error, reported their confidence in the model's forecasts and their own forecasts on five-point scales (1 = none; 5 = a lot), and answered two open-ended questions.[9] The first open-ended question asked participants in the can't-change, adjust-by-10, and change-10 conditions to report why they chose to have their bonus determined by the model's forecasts or their own forecast, depending on which they had chosen; participants in the use-freely condition reported how much they had used the model's forecasts. The second question asked all participants to report their thoughts and feelings about the statistical model. After completing these questions, participants learned their bonus and reported it to a lab manager.[10] Finally, participants reported their age, gender, and highest-completed level of education.

Results

Choosing to Use the Model. As predicted, participants in the adjust-by-10 and change-10 conditions, who were restrictively able to modify the model's forecasts, chose to use the model's imperfect forecasts much more often than participants in the can't-change condition, who could not modify the model's forecasts (see Figure 2). Whereas only 32% of participants in the can't-change condition chose to use the model's forecasts, 73% of participants in the change-10 condition ($\chi^2(1, N = 145) = 24.19$, $p < 0.001$) and 76% of participants in the adjust-by-10 condition ($\chi^2(1, N = 146) = 28.40$, $p < 0.001$) chose to use the model. (See Study S2 in the online supplementary materials for a replication of this result using a different forecasting task.)

Figure 2. Study 1: Participants Who Could Restrictively Modify the Model's Forecasts Were More Likely to Choose to Use the Model, and They Performed Better as a Result



Note. Errors bars indicate ±1 standard error.

As a result of their infrequent use of the model, participants in the can't-change condition provided forecasts that were farther from the model's than participants in the adjust-by-10 ($t(144) = 3.24$, $p = 0.002$), change-10 ($t(143) = 2.98$, $p = 0.003$), and use-freely ($t(143) = 3.47$, $p < 0.001$) conditions (see Figure 2). Participants who chose to use the model in the adjust-by-10 and change-10 conditions deviated from the model less than participants in the use-freely condition. Whereas participants in the use-freely condition adjusted the model's forecasts by 8.18 percentiles on average, participants who chose to use the model in the adjust-by-10 and change-10 conditions adjusted the model's forecasts by an average of 4.71 percentiles ($t(124) = -6.17$, $p < 0.001$) and 5.29 percentiles, ($t(121) = -4.79$, $p < 0.001$), respectively. While these comparisons are imperfect because of potential selection concerns, they suggest that restricting the amount by which people can adjust a model's forecasts does result in forecasts that are closer to the model's. We cleanly test this hypothesis in Study 3.

We also found evidence that participants in the use-freely condition voluntarily integrated the model's forecasts into their own forecasts. Participants in the use-freely condition provided forecasts that deviated from the model's forecasts less than half as much ($M = 8.18$) as participants in the can't-change condition ($M = 18.66$) ($t(143) = 12.52$, $p < 0.001$), who all made their own forecasts without seeing the model's (regardless of their choice). This suggests that providing forecasters with a model's forecast before they have a chance to form their own opinion may lead them to anchor on the model's forecast and, as a result, rely more heavily on the algorithm.

Forecasting Performance. As shown in Figure 2, participants who had the option to adjust the model's forecasts outperformed those who did not. Participants' forecasts in the can't-change condition were less accurate, and earned them smaller bonuses, than the forecasts of participants in the adjust-by-10, change-10, and use-freely conditions.[11]

Figure 3 displays the distribution of participants' performance by condition. Three things are apparent from the figure. First, although the model's estimates were far from perfect, reliance on the model was strongly associated with better performance. Indeed, failing to choose to use the model was much more likely to result in very large average errors (and bonuses of $0). Second, participants in the can't-change condition performed worse precisely because they were less likely to use the model, and not because their forecasting ability was worse. Third, exposing all participants in the use-freely condition to the model's forecasts seems to have prevented them from making very large errors, as no participant erred by more than 28 percentiles on average.

Discussion. In sum, participants who could restrictively modify the model's imperfect forecasts were more likely to choose to use the model's forecasts than those who could not. As a result, they performed better and earned more money. Additionally, participants who could use the model's forecasts freely seemed to anchor on the model's forecasts, which improved their performance by reducing their tendency to make large errors.

Study 2

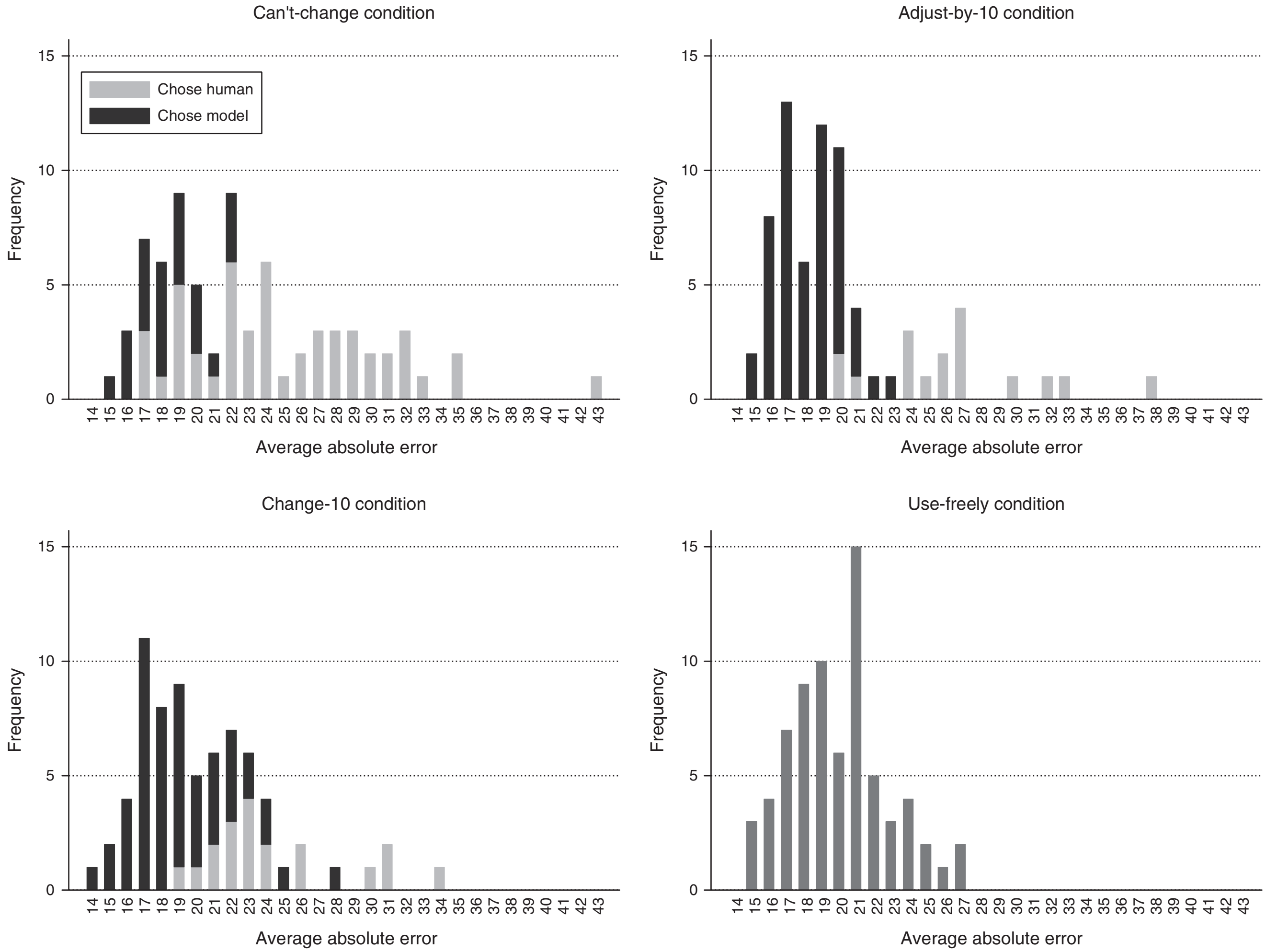
Methods

Overview. Study 1 (and Study S2 in the online supplementary materials) showed that people were more likely to choose to use an imperfect algorithm if they were given the option to restrictively adjust its forecasts. In Study 2, we explored people's sensitivity to the restriction on their adjustments. Would further restricting the amount by which people can adjust their forecasts diminish their willingness to use the algorithm's imperfect forecasts, or would people be willing to commit to using an imperfect algorithm as long as they are given even a modicum of control over its forecasts?

To answer this question, we asked participants to engage in the same student forecasting task as in Study 1, and we randomly assigned them to one of four experimental conditions: a can't-change condition that was unable to modify the algorithm's forecasts or one of three conditions in which they could adjust the model's forecasts by either 10, 5, or 2 percentiles. If participants' use of the model depends on *how much* control they have over its imperfect forecasts, then they should be more likely to choose to use the model when they can adjust it by a larger amount (10 percentiles) than by a smaller amount (2 percentiles). However, if participants simply need to have *some* control over the model's imperfect forecasts in order to choose it, then they should be equally likely to choose to use the model no matter whether they can adjust the model by 10, 5, or even 2 percentiles.

Participants. Participants, recruited through Amazon Mechanical Turk (MTurk), earned $1 for completing the study and could earn up to an additional $0.50 depending on their forecasting performance. We decided in advance to recruit 800 participants (200 per condition). Participants began the study by answering a question designed to check whether they were carefully reading instructions. We prevented the 107 participants who failed this check from participating. Additionally, 78 participants quit the survey before being assigned to a condition, and 53 additional participants quit before completing their forecasts. We replaced these participants, and our final sample consisted of 816 participants who completed their forecasts. The final sample averaged 34 years of age and was 48% female.

Figure 3. Study 1: The Distribution of Participants' Average Absolute Errors by Condition and Whether or Not They Chose to Use the Model's Forecasts



Procedure. This study used the same forecasting task as in Study 1: participants predicted the percentiles of high school students on a standardized math test. The procedure was the same as Study 1's except for six changes. First, the four experimental conditions were different. Participants were randomly assigned to either a can't-change condition, an adjust-by-10 condition, an adjust-by-5 condition, or an adjust-by-2 condition. In the can't-change condition, participants who chose to use the model could not modify its forecasts, whereas in the adjust-by-X conditions, participants who chose to the use the model could adjust it by X percentiles. For example, in the adjust-by-2 condition, participants who decided to use the model's forecasts could adjust its forecasts by up to two percentiles. Second, we recruited participants from Amazon Mechanical Turk instead of the laboratory. Third, as previously mentioned, we added a reading check to the beginning of the survey to identify and remove participants who were not reading instructions. Fourth, we used a different payment rule. Participants were paid a $0.50 bonus if their official forecasts were within five percentiles of students' actual percentiles. This bonus decreased by $0.10 for each additional five percentiles of error in participants' forecasts (this payment rule is reproduced in Appendix B). As a result, participants whose forecasts were off by more than 25 percentiles received no bonus. Fifth, at the end of the survey we asked participants to recall the model's average error. Sixth, we did not include the exploratory questions described in Endnote 8.

Results

Choosing to Use the Model. Consistent with the results of Study 1, participants who had the option to adjust the model's imperfect forecasts chose to use the model more often than participants who could not modify its forecasts (see Figure 4). Whereas only 47% of participants in the can't-change condition chose to use the model's forecasts, 70% of participants in the adjust-by-X conditions chose to the model ($\chi^2(1, N = 834) = 36.46$, $p < 0.001$). Additionally, and somewhat surprisingly, we found that participants' decision to use the model in the adjust-by-X conditions did not depend on how much they were able to adjust the model: 71%, 71%, and 68% chose to the model in the adjust-by-10, adjust-by-5, and adjust-by-2 conditions, respectively. These three conditions did not differ significantly ($\chi^2(2, N = 623) = 0.42$, $p = 0.809$). Although we cannot reject the possibility that participants may have been slightly sensitive to the amount by which they could adjust the model, we can conclude that their willingness to use the model was not *detectably* altered by imposing an 80% reduction of the amount by which they could adjust. (See Study S3 in the online supplementary materials for a replication of this insensitivity using the change-X forecasting process.)

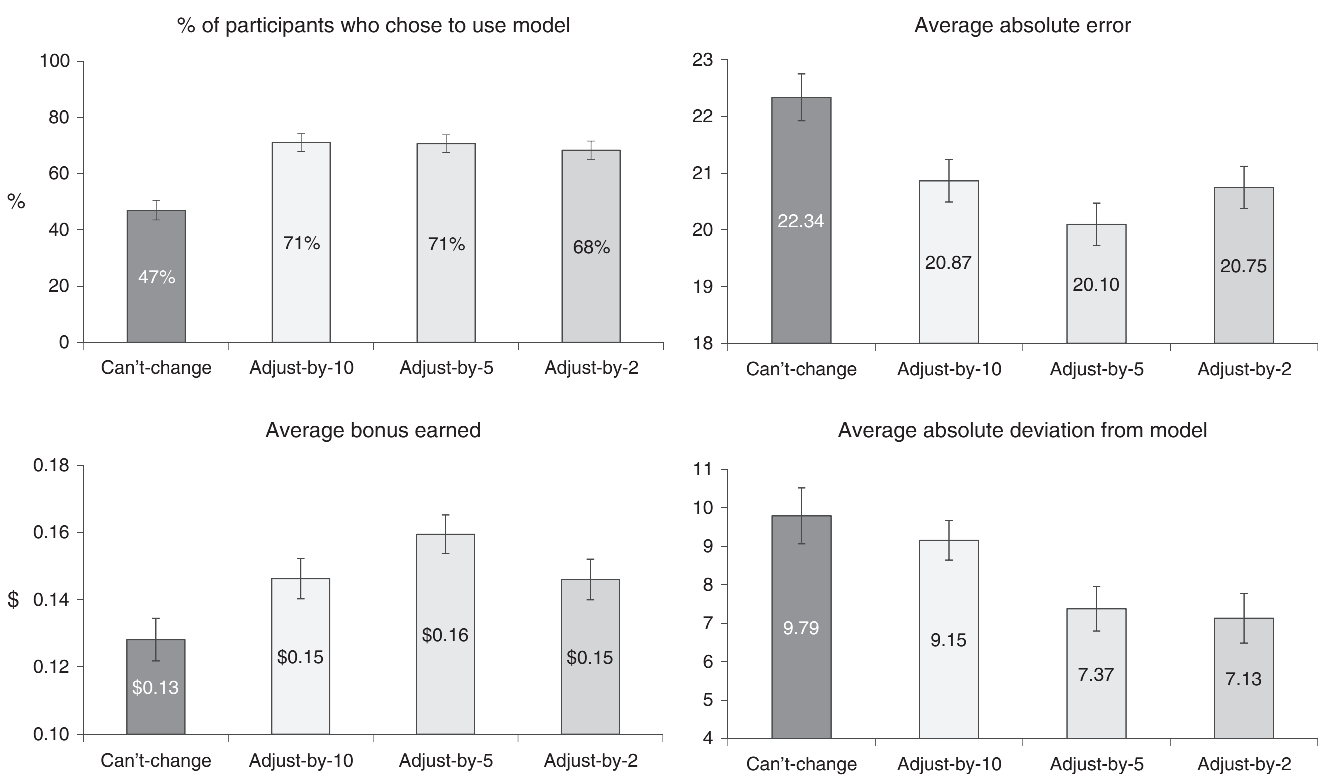
Compared with participants in the can't-change condition, participants in the adjust-by-10 condition deviated from the model directionally less ($t(407) = 0.71$, $p = 0.475$), and participants in the adjust-by-5 ($t(409) = 2.61$, $p = 0.010$) and adjust-by-2 ($t(406) = 2.74$, $p = 0.006$) conditions deviated significantly less (see Figure 4). Participants who chose to the use the model in the adjust-by-X conditions did not deviate from the model as much as they could have, regardless of whether they were in the adjust-by-10 ($M = 5.00$), adjust-by-5 ($M = 2.61$), or adjust-by-2 ($M = 1.33$) condition. Given the desire of participants in the adjust-by-10 condition to adjust by five percentiles on average, it is surprising (to us) that those in the adjust-by-5 and adjust-by-2 conditions did not adjust by close to the maximum amount.

Forecasting Performance. As in Study 1, participants who were given the option to adjust the model's imperfect forecasts performed better than those who were not (see Figure 4). Participants in the can't-change condition made significantly larger errors than participants in each of the adjust-by-X conditions and earned smaller bonuses as a result.[12]

Figure 5 displays the distribution of participants' performance by condition. We again see that reliance on the model was strongly associated with better performance, even though its forecasts were far from perfect. Also, participants in the can't-change condition performed worse precisely because they were less likely to use the model, and not because their forecasting ability was worse.

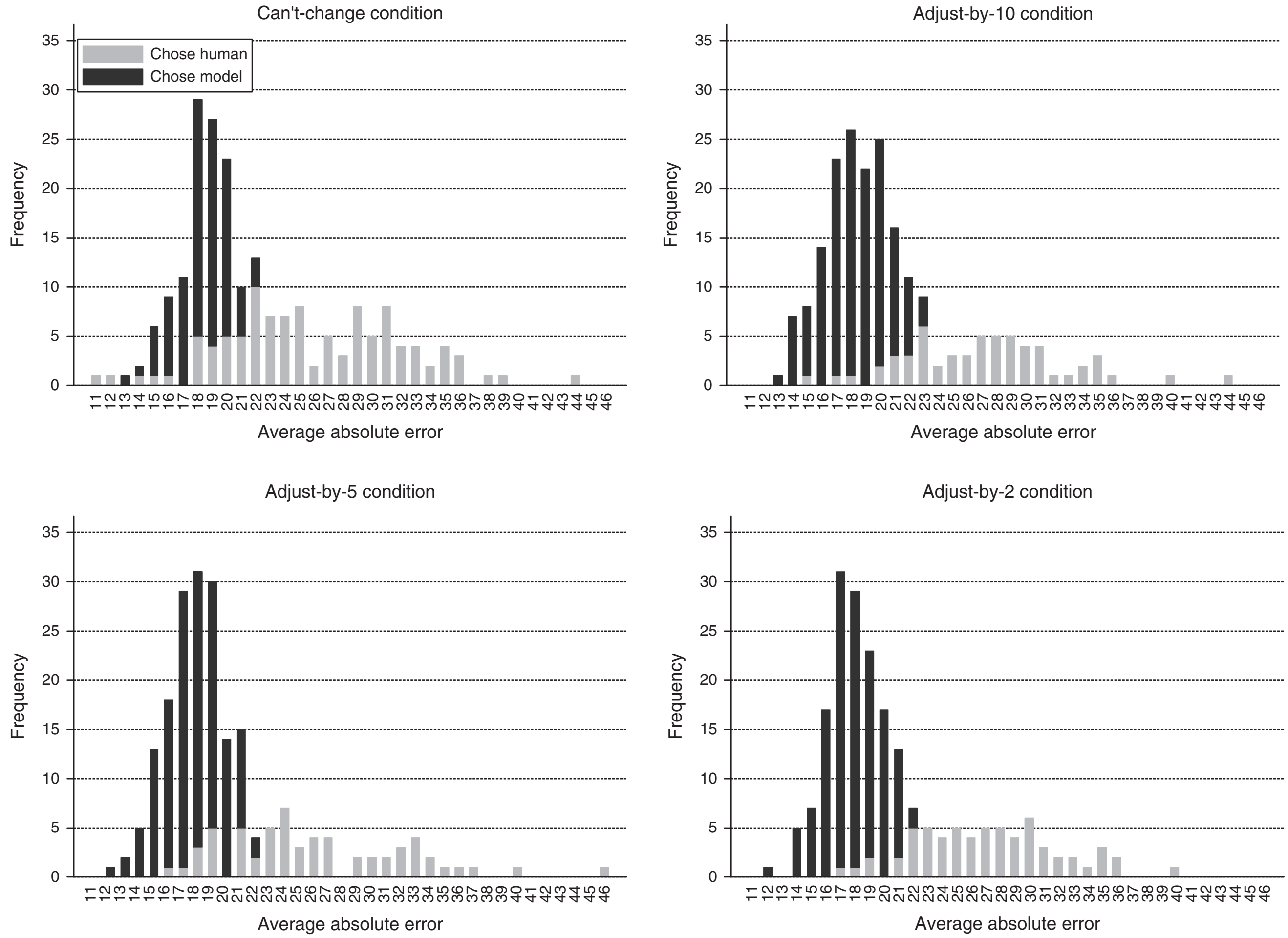
Discussion. In Study 2, participants were once again more likely to choose to use an imperfect algorithm's forecasts if they could modify those forecasts. Moreover, they were relatively insensitive to the amount by which they could adjust the model's forecasts. This finding suggests that, while it is beneficial to give people control over an imperfect algorithm's forecasts, the amount of control needed to obtain those benefits is actually quite small.

Figure 4. Study 2: Participants Who Could Restrictively Modify the Model's Forecasts Were More Likely to Choose to Use the Model, and They Performed Better as a Result



Note. Errors bars indicate ±1 standard error.

Figure 5. Study 2: The Distribution of Participants' Average Absolute Errors by Condition and Whether or Not They Chose to Use the Model's Forecasts



Study 3

In Studies 1 and 2, we found that people were much more likely to choose to use an imperfect algorithm if they were allowed to adjust its forecasts by even a small amount (see also Studies S2 and S3 in the online supplementary materials). However, whereas in each of these studies the decision to use the algorithm was made before participants experienced what it was like to use it and before receiving performance feedback, overcoming algorithm aversion over the long term requires a willingness to use the algorithm even after using it and making errors. This is no small feat, as prior research shows that people punish algorithms more than humans for making the same mistake, rendering them especially reluctant to choose to use algorithms after seeing them err (Dietvorst et al. 2015).

In Study 3, we investigated how people's experience with a forecasting process in which they either can or cannot modify an imperfect algorithm's forecasts affects their judgments and subsequent decisions. Using the same forecasting task as Studies 1 and 2, we conducted this experiment in two stages. In the first stage of 10 forecasts, participants were randomly assigned to adhere to one of three forecasting methods. In the *model-only* condition, participants were forced to use only the model's estimates for each forecast. In the *adjust-by-*10 condition, participants could adjust the model's forecasts by up to 10 percentiles. In the *use-freely* condition, participants were given the model's forecasts and could adjust them as much as they wanted. After completing a round of forecasts, participants were asked to indicate their satisfaction with, and confidence in, the forecasting process they just used. Then participants learned their performance for their first round of forecasts.

Next, participants answered questions about three forecasting processes and then chose which of them to use for a second round of 10 forecasts. All participants chose among using the model exclusively (model-only), using their own forecasts exclusively (human-only), and either adjusting the model's forecasts in a constrained (adjust-by-10) or unconstrained (use-freely) manner. Participants in the use-freely condition and half of participants in the model-only condition had the option of adjusting the model in an unconstrained manner (use-freely). Participants in the adjust-by-10 condition and the other half of participants in the model-only condition had the option of adjusting the model in a constrained manner (adjust-by-10).

This design allowed us to answer four open questions. First, we examined how experience with different forecasting processes translates into satisfaction with, and confidence in, those processes. After one round of forecasts, are forecasters more satisfied with, and confident in, a process in which they (1) use an imperfect algorithm's forecasts exclusively, (2) adjust an imperfect algorithm's forecasts in a constrained manner, or (3) adjust an imperfect algorithm's forecasts in an unconstrained manner? Studies 1 and 2 suggest that people do prefer adjusting an imperfect algorithm's forecasts over using them exclusively; moreover, these studies also suggest that people may not object to being partially constrained to the imperfect algorithm.

Second, we examined whether people are willing to continue adjusting an imperfect algorithm's forecasts in a constrained or unconstrained manner after getting performance feedback. If allowing people to adjust an imperfect algorithm's forecasts is an effective long-term prescription for algorithm aversion, people will have to be willing to stick with this forecasting process after learning that it produces errors. It is not at all obvious that this is the case, as past research (Dietvorst et al. 2015) has found that people fail to use an algorithm exclusively after learning that it produces imperfect (albeit superior) forecasts.

Third, we examined whether people's perceptions of the accuracy of the model's forecasts relative to their own differs between the use-freely, adjust-by-10, and model-only forecasting processes. Participants who have had the ability to adjust the model's forecasts may actually hold the model in higher regard, perhaps because their increased satisfaction with the forecasting process bleeds into their feelings about the model (e.g., Finucane et al. 2000).

Fourth, we examined whether adjusting an imperfect algorithm's forecasts in a constrained versus an unconstrained manner in round 1 leads to better forecasting performance in round 2.

Methods

Participants. MTurk participants earned $1 for completing the study and could earn up to an additional $1 depending on their forecasting performance. We decided in advance to recruit 800 participants (200 per condition). Participants began the study by answering a question designed to check whether they were carefully reading instructions. We prevented the 206 participants who failed this check from participating. Additionally, 154 participants quit the survey before being assigned to a condition, and 54 additional participants quit before completing their forecasts. We replaced these participants and had a sample of 818 participants who completed their forecasts. The final sample averaged 33 years of age and was 49% female.

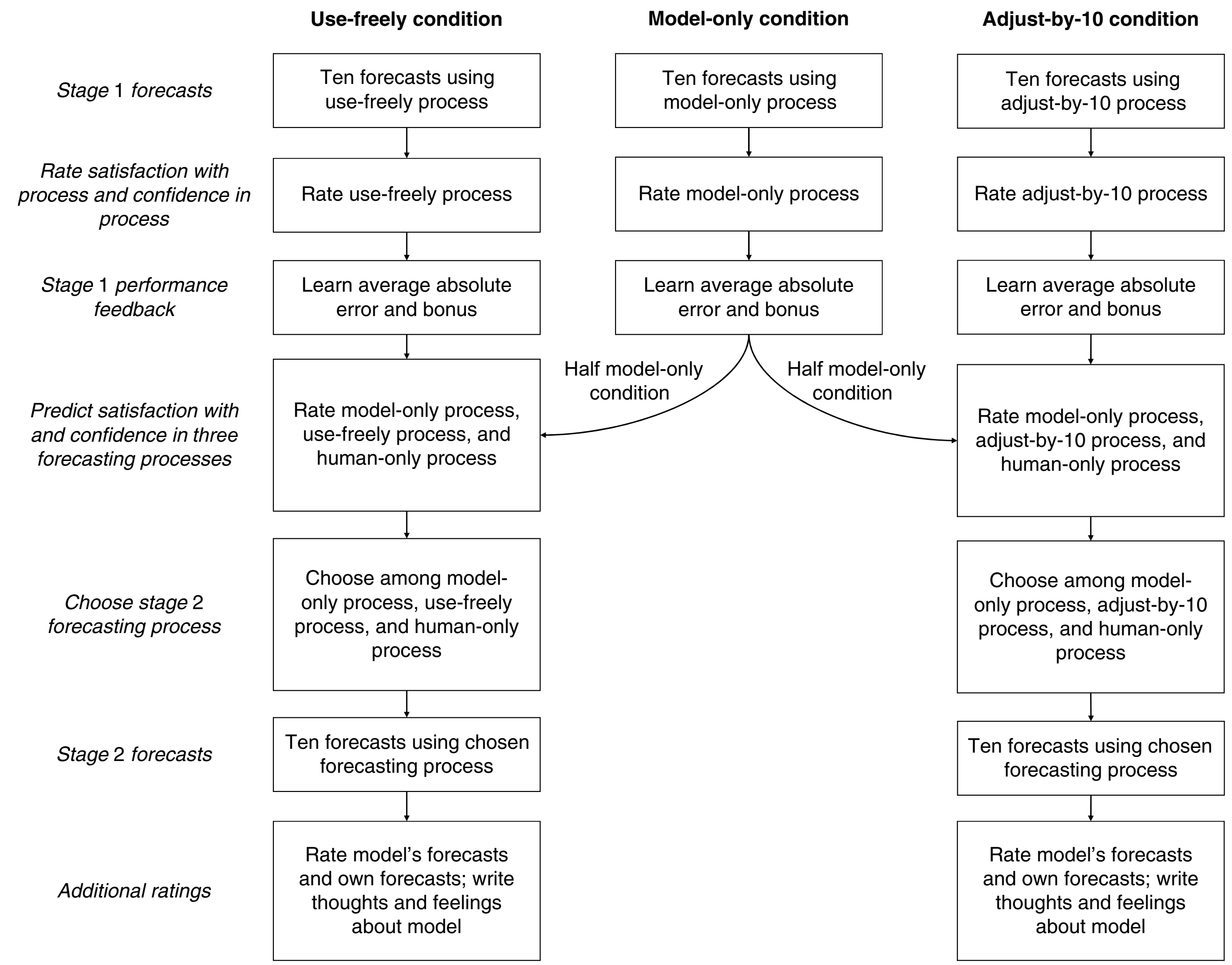
Procedure. This study was administered as an online survey. Participants began the survey by indicating their informed consent and entering their Mechanical Turk identification number. They then answered a question designed to ensure that they were reading the instructions. Only those who answered this question correctly proceeded to the remainder of the survey, which introduced participants to the forecasting task (predicting students' performance on a standardized test), introduced participants to the statistical model, and informed participants that the model was off by 17.5 percentiles on average. This part of the survey was identical to Studies 1 and 2.

Figure 6 shows the rest of the procedure of Study 3. After reading about the forecasting task, participants were told that they would make 10 forecasts and that their performance would be incentivized. Just as in Study 2, they learned that they would be paid a $0.50 bonus if their official forecasts were within five percentiles of students' actual percentiles on average and that this bonus decreased by $0.10 for each additional five percentiles of average error. Participants were then assigned to one of three conditions. One-half of the participants were assigned to the model-only condition, in which they were forced to use the model's forecasts without being able to adjust them. One-quarter of the participants were assigned to the use-freely condition, in which they received the model's forecasts and could adjust them as much as they wanted. And the remaining one-quarter of participants were assigned to the adjust-by-10 condition, in which they received the model's forecasts and could adjust them up to 10 percentiles. Participants were not given the option to only use their own forecasts as they were in the adjust-by-10 conditions used in Studies 1 and 2. Participants were required to type two sentences describing their condition's forecasting procedure to ensure that they understood the instructions.[13]

Next, participants completed their first set of 10 forecasts.[14] After participants completed these forecasts, they were reminded of the forecasting process that they had used and asked to rate how satisfied they were with that process on a five-point scale (1 = very dissatisfied; 5 = very satisfied) and how much confidence they had that the process performed well (1 = none; 5 = a lot). On the next page, participants learned how much their first 10 forecasts had erred on average and how much money they had earned.

Next, participants were presented with three forecasting processes (human-only, model-only, and either adjust-by-10 or use-freely), asked about their

Figure 6. Study 3 Procedure



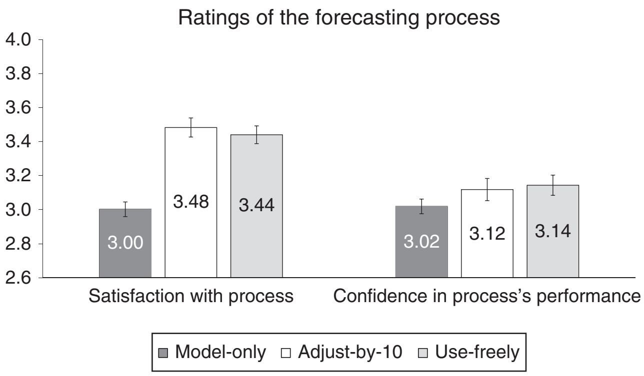
satisfaction with (1 = very dissatisfied; 5 = very satisfied) and confidence in (1 = none; 5 = a lot) these three forecasting processes,[15] and then asked to choose among these three forecasting processes for a second set of 10 forecasts with the same incentives as the first set. Participants in the use-freely condition and half of participants in the model-only condition rated and chose among the use-freely, model-only, and human-only forecasting processes. Participants in the adjust-by-10 condition and the other half of participants in the model-only condition rated and chose among the adjust-by-10, model-only, and human-only forecasting processes.

After completing the second set of 10 forecasts, participants estimated their own average error and the model's average error, reported their confidence in the model's forecasts and their own forecasts on five-point scales (1 = none; 5 = a lot), and reported their thoughts and feelings about the statistical model. Finally, participants reported their age, gender, and highest level of education.

Results

Confidence in and Satisfaction with Forecasting Process. After participants were randomly assigned to a forecasting process and made their first set of 10 incentivized forecasts, they rated their satisfaction with and confidence in their assigned forecasting process. Participants who were assigned to the model-only process in round 1 were less satisfied with their forecasting process than participants who were assigned to the adjust-by-10 ($t(614) = -6.59$, $p < 0.001$) and use-freely ($t(620) = -6.17$, $p < 0.001$) processes (see Figure 7). Also, participants who were assigned to the model-only process in round 1 were directionally less confident in their forecasting process than participants who were assigned to the adjust-by-10 process ($t(614) = -1.29$, $p = 0.196$) and marginally less confident in their forecasting process than participants who were assigned to the use-freely process ($t(620) = -1.68$, $p = 0.093$). Thus, allowing participants to modify the model's forecasts increased their satisfaction with their forecasting process.

Figure 7. Study 3: Participants Who Could Modify the Model's Forecasts Were More Satisfied and Confident



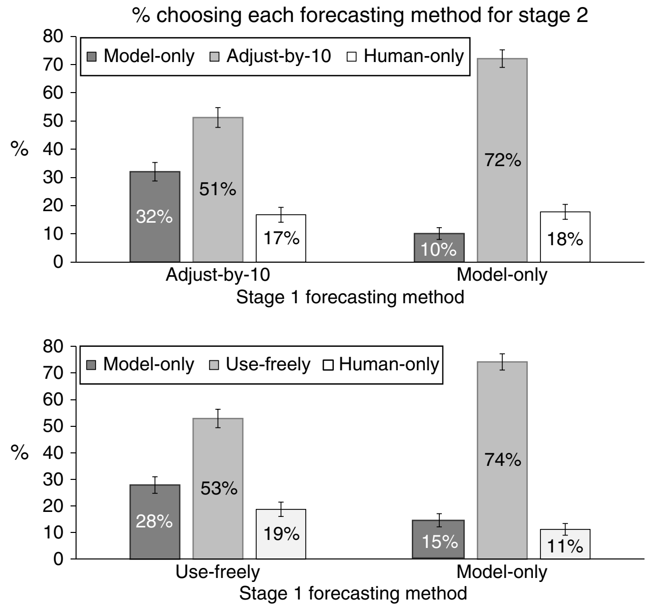
Note. Errors bars indicate ±1 standard error.

Participants' satisfaction with and confidence in their forecasting process did not differ by whether they adjusted the algorithm's forecasts in a constrained or unconstrained manner. Participants in the adjust-by-10 condition were about equally as satisfied with ($t(410) = 0.56$, $p = 0.578$) and confident in ($t(410) = -0.29$, $p = 0.774$) their assigned forecasting process as were participants in the use-freely condition, even though they had less freedom to modify the algorithm's forecasts. This again suggests that forecasters may not object to being constrained to an imperfect algorithm as long as they can modify its forecasts.

Choice of Second Forecasting Process. When it came time to choose a forecasting process for the second round of incentivized forecasts, most participants chose to use the model's imperfect forecasts (see Figure 8). In fact, participants in each condition chose to use the model in some manner (either exclusively or partially) over 80% of the time. Participants chose to combine their own judgment with the algorithm's forecast (51%–74%) most often, instead of using their own forecasts exclusively (11%–19%) or using the algorithm's forecasts exclusively (10%–32%). This again suggests that adjusting an imperfect algorithm's forecasts is a palatable forecasting process. Also, participants in the adjust-by-10 condition, who were constrained to the imperfect algorithm, chose to adjust the model's forecasts by 10 at virtually the same rate (53%) as participants in the use-freely condition, who were unconstrained, chose to adjust the model's forecasts freely (51%) ($\chi^2(1, N = 410) = 0.15$, $p = 0.699$). This suggests that constraining forecasters to an imperfect algorithm's forecasts does not increase their likelihood of abandoning the algorithm or their forecasting process.

There is one substantial difference between the choices of participants who could adjust the model's forecasts in stage 1 and those who could not. Participants who could (restrictively or freely) modify the model's forecasts in stage 1 were much more likely to choose the "model-only" option (30%) than

Figure 8. Study 3: Participants in Each Condition Chose to Modify the Model's Forecasts Most Often, Instead of Only Using Their Own Forecasts or Only Using the Model's Forecasts



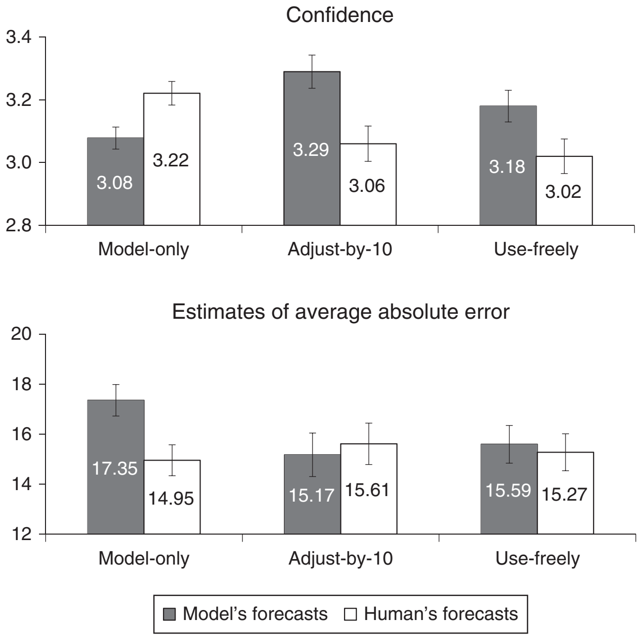
Notes. Errors bars indicate ±1 standard error. Participants in the top panel chose among the model-only, adjust-by-10, and human-only forecasting process for their stage 2 forecasts. Participants in the bottom panel chose among the model-only, use-freely, and human-only forecasting process for their stage 2 forecasts.

participants who could not modify the model's forecasts in stage 1 (12%) ($\chi^2(1, N = 823) = 38.45$, $p < 0.001$). This suggests that participants who were able to modify the model's forecasts may have held the model in higher regard (relative to themselves) compared with participants who were not able to modify the model's forecasts.[16]

Perceptions of Model and Self. Participants' confidence ratings and performance estimates suggest that allowing people to modify an imperfect algorithm's forecasts may improve their perceptions of the algorithm relative to themselves (see Figure 9). Compared with participants who could not modify the model's forecasts during the first set of forecasts, participants who could modify the model's forecasts had more confidence in the model's forecasts relative to their own ($t(816) = -5.86$, $p < 0.001$) and estimated that the model's average absolute error was better relative to their own ($t(817) = 3.92$, $p < 0.001$).[17] These results suggest that people may hold algorithms in higher regard relative to themselves if they previously had the ability to modify the algorithm's forecasts, and they are consistent with participants' increased selection of the model-only forecasting process in the adjust-by-10 and use-freely conditions.

Forecasting Performance. As shown in Figure 10, participants in the adjust-by-10 condition deviated from

Figure 9. Study 3: Participants Who Could Modify the Model's Forecasts Were More Confident in the Model's Forecasts and Thought That the Model Performed Better Relative To Themselves



Note. Error bars indicate ±1 standard error.

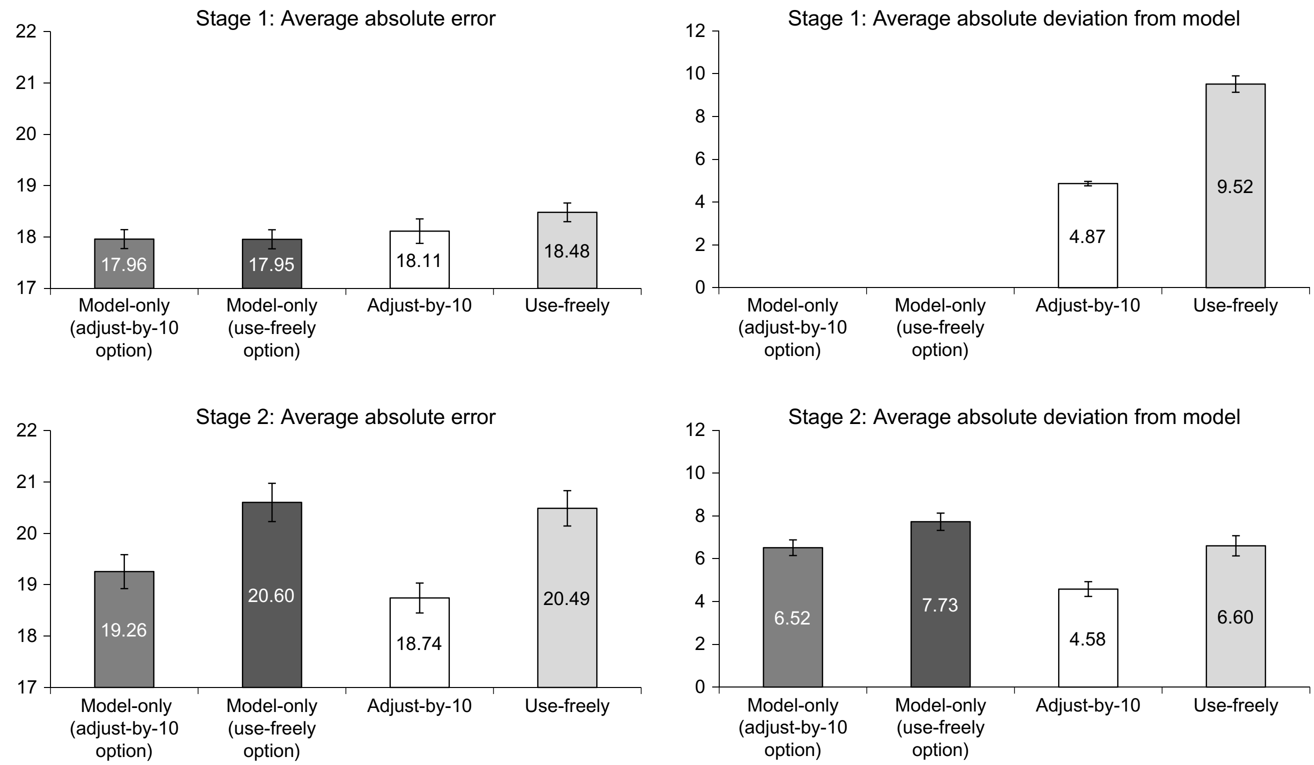
the model's forecasts far less than participants in the use-freely condition during the stage 1 forecasts ($t(410) = -11.61$, $p < 0.001$). Although this only translated into directionally better average absolute errors in stage 1 of this study, ($t(410) = -1.22$, $p = 0.224$), we are confident that giving more weight to a model's forecasts would significantly increase forecasting performance in a larger sample of forecasts. Participants in the model-only condition, who could not adjust the model's forecasts, had average absolute errors that were directionally better than participants in the adjust-by-10 condition ($t(616) = -0.63$, $p = 0.531$) and significantly better than participants in the use-freely condition ($t(622) = -2.32$, $p = 0.021$). As a result of these performance differences, participants in the use-freely condition earned significantly smaller bonuses ($M =$\$0.18) than participants in the model-only condition ($M =$ \$0.19) ($t(622) = 3.03$, $p = 0.003$) and directionally smaller bonuses than participants in the adjust-by-10 condition ($M =$ \$0.19) ($t(410) = 1.27$, $p = 0.207$). Thus, the more participants were required to use the model in the first forecasting stage, the better they performed.

More important is how participants fared in stage 2, when they could choose to either completely use the model, completely use their own forecasts, or adjust the model's forecasts. Participants in the adjust-by-10 condition deviated from the model's forecasts less than participants in the use-freely condition ($t(403) = -3.45$, $p < 0.001$). Also, participants in the model-only condition who had the option to use the adjust-by-10 process deviated from the model's forecasts less than participants in the model-only condition who had the option to use the use-freely process ($t(411) = -2.23$, $p = 0.027$). As described below, these differences in deviation from the model were driven by participants who chose to use the adjust-by-10 and use-freely forecasting processes.

Participants who chose to use the model-only process did not deviate at all from the model's forecasts, and the degree to which participants who chose the human-only process deviated from the model did not differ by condition ($F(3,128) = 1.95$, $p = 0.125$). However, participants in the adjust-by-10 condition who chose to use the adjust-by-10 process deviated from the model substantially less ($M = 4.65$) than participants in the use-freely condition who chose to use the use-freely process ($M = 6.84$) ($t(208) = -4.21$, $p < 0.001$). Additionally, participants in the model-only condition who chose to use the adjust-by-10 process deviated from the model substantially less ($M = 5.31$) than participants in the model-only condition who chose to use the use-freely process ($M = 8.26$) ($t(300) = -7.23$, $p < 0.001$). Thus, participants who were given the option to restrictively modify the model's forecasts (i.e., adjust-by-10) deviated from the model much less than participants who were given the option to modify the model without restriction (i.e., use-freely). As shown in the next paragraph, this difference in deviation from the model did translate into significant performance differences.

In stage 2, participants who used the adjust-by-10 process performed better than those who used the use-freely process. Participants who chose the adjust-by-10 process had lower average absolute errors ($M = 17.90$) than participants who chose the use-freely process ($M = 20.25$) ($t(510) = -6.28$, $p < 0.001$), lower average absolute errors than participants who chose to use their own forecasts ($M = 24.50$) ($t(382) = 14.01$, $p < 0.001$), and similar average absolute errors to participants who chose to use the model's forecasts ($M = 18.20$) ($t(424) = 0.97$, $p = 0.331$). Participants who used the adjust-by-10 process outperformed those who used the use-freely process specifically because they provided forecasts that were closer to the model's ($M = 5.04$) compared with participants who used the use-freely process ($M = 7.67$) ($t(510) = -8.10$, $p < 0.001$).[18] As a result of these differences between the adjust-by-10 and use-freely forecasting processes, participants who had the option to use the adjust-by-10 process in stage 2 (i.e., the adjust-by-10 condition, half of the model-only condition) had lower average absolute errors than, and earned more money than, participants who had the option to instead use the use-freely process (i.e., the use-freely condition, the other half of the model-only condition).[19]

Figure 10. Study 3: Participants Who Had the Option to Adjust the Model Restrictively in the Second Stage of Forecasts Performed Better and Earned More Money



Note. Errors bars indicate ±1 standard error.

Discussion. Taken together, these results inform multiple open questions regarding people's use of imperfect algorithms. First, they highlight the substantial benefits of letting people modify an imperfect algorithm's forecasts. It increases their satisfaction with the process, their confidence in and perceptions of the model relative to themselves, and their use of the model on subsequent forecasts. Second, the results show that people who are able to modify an imperfect algorithm's forecasts by a limited amount will not necessarily be less satisfied than if they can modify it by an unlimited amount, and that people will elect to modify an algorithm's forecasts in a constrained manner even after using this process and seeing it err. Third, the results show that restricting people's adjustments to the model, rather than allowing them to use it freely, prevents them from making forecasts that deviate greatly from the model and thus improves their forecasting performance.

General Discussion

Our studies show that people will use imperfect algorithms to make incentivized forecasts so long as they can slightly modify them. Although people often fail to use imperfect algorithms exclusively, they will commit to using imperfect algorithms in a constrained manner. Furthermore, we found evidence that people are relatively insensitive to the amount by which they can modify the imperfect algorithm's forecasts when making this decision. We also found that allowing people to adjust an algorithm's forecasts has additional benefits. Participants who were able to modify an imperfect algorithm's forecasts reported higher satisfaction with their forecasting process and thought that the algorithm performed better relative to themselves compared with participants who could not modify the algorithm's forecasts. Additionally, we found that people are not less satisfied modifying an algorithm's forecasts in a constrained manner versus an unconstrained manner. Finally, we found that restricting the amount by which people can modify an algorithm's forecasts leads them to deviate from the algorithm less and thus to perform better.

These findings have many important implications for managers trying to increase employees' and customers' use of algorithms. First, framing the decision of whether or not to use an algorithm as an all-or-nothing decision is likely to be counterproductive. People are unlikely to commit to using an algorithm's forecasts exclusively after getting performance feedback or learning that it is imperfect. Furthermore, forcing employees into a regime in which they have to use an imperfect algorithm's forecasts exclusively may lead them to become dissatisfied or push for a change. However, asking people to commit to an algorithm's forecasts that they can modify by a limited amount seems much more palatable. People will be much more likely to choose to use an imperfect algorithm if they can modify its forecasts, and employees will not necessarily be dissatisfied if they are partially constrained to an imperfect algorithm's forecasts. Finally, the results of Study 3 suggest that constraining employees to an imperfect algorithm not only will be acceptable to them but will also increase their forecasting performance by keeping their forecasts closer to the algorithm's. If for some reason having employees make constrained adjustments to an algorithm's forecasts is not possible, Study 1 shows that having employees make unconstrained adjustments to an algorithm's forecasts can also substantially improve their forecasting performance.

Our techniques for increasing people's choice of algorithms are likely to work for people's choice of other decision aids as well. For example, people often give less weight to other people's advice than they should when left to their own devices (see Bonaccio and Dalal 2006); however, it is possible that people will commit to using another person's advice in a constrained manner. To test this possibility, we ran a study with four between-subjects conditions that differed in whether (1) participants had the option to use the imperfect forecasts of another person (human conditions) or a statistical algorithm (algorithm conditions),[20] and (2) participants would be unable to adjust the forecasts of that entity (can't-change conditions) or able to adjust the forecasts of that entity by up to five percentiles (adjust-by-5 conditions). This resulted in a 2 (entity: human versus algorithm) × 2 (adjustment: can't-change versus adjust-by-5) design (see Study S5 in the online supplementary materials for a detailed description of the methods and results). We found that giving participants the ability to modify the other entity's forecasts increased participants' choice of the human's forecasts (21%–41%) to a similar degree that it increased their choice of the algorithm's forecasts (46%–69%; $z(N = 2{,}014) = -0.01$, $p = 0.996$). Thus, in general, people may be more willing to precommit to using information in a constrained manner when they have the ability to modify it.

Limitations and Future Directions

The studies in this paper leave some questions unanswered. First, results may differ with different forecasting tasks. For example, when forecasters have important information that an algorithm does not have, allowing them to make large adjustments to the algorithm's forecasts may sometimes increase accuracy (e.g., Fildes et al. 2009, Lawrence et al. 2006). In these cases, constraining forecasters tightly to an imperfect algorithm's forecasts may not improve forecasting performance.

Second, there could be conditions under which the effects we found would be diminished or eliminated. For example, people may not be willing to use algorithms that are far more imperfect than the algorithms that we employed. Additionally, although we did find that participants were insensitive to the amount that they could adjust the model's forecasts, we only gave participants the option to adjust the model by 2–10 percentiles. It is possible that more participants would have chosen to use the model if they could adjust it to a greater degree (e.g., 20 percentiles) or that fewer participants would have chosen to use the model if they could adjust it to a smaller degree (e.g., one percentile). Third, while we did use two different populations and tasks in our studies (including those in the online supplementary materials), it is possible that the effects we found are dependent on some characteristics of those tasks or populations. Future work could investigate the effects shown in this paper with different populations of participants, with different algorithms that are more or less accurate than those used in our studies, and in different forecasting domains. Research with a population of professional forecasters would be especially informative.

In conclusion, we found that letting people adjust an imperfect algorithm's forecasts increases their likelihood of using it and their confidence in it. We also found that people are insensitive to the amount by which they can adjust the algorithm's forecasts and that restricting the amount that people can adjust an algorithm's forecasts leads to better performance. Participants in our studies did often worsen the algorithm's forecasts when given the ability to adjust them. However, we may have to accept this error so that, overall, people make less error.

Acknowledgments

The authors thank the Wharton Behavioral Laboratory and the Wharton Risk Center Ackoff Doctoral Student Fellowship for financial support.

Appendix A. Payment Rule for Study 1

Participants in Study 1 were paid as follows:
$5—within 5 percentiles of a student's actual percentiles on average
$4—within 10 percentiles of a student's actual percentiles on average
$3—within 15 percentiles of a student's actual percentiles on average
$2—within 20 percentiles of a student's actual percentiles on average
$1—within 25 percentiles of a student's actual percentiles on average.

Appendix B. Payment Rule for Studies 2 and 3

Participants in Studies 2 and 3 were paid as follows:
$0.50—within 5 percentiles of a student's actual percentiles on average
$0.40—within 10 percentiles of a student's actual percentiles on average

\$0.30—within 15 percentiles of a student's actual percentiles on average
\$0.20—within 20 percentiles of a student's actual percentiles on average
\$0.10—within 25 percentiles of a student's actual percentiles on average

*Participants in Study 3 were paid separately for each of two rounds of 10 forecasts.

Endnotes

[1] In this paper, the term "algorithm" describes any evidence-based forecasting formula, including statistical models, decision rules, and all other mechanical procedures used for forecasting.

[2] This logic applies not only to reliance on algorithmic forecasts but to reliance on any forecasting process that people believe to be inevitably imperfect. For example, in Study S5 in the online supplementary materials, we find that people not only prefer using an algorithm's forecasts when they can modify them but also prefer using the forecasts of a (nonexpert) past participant when they can modify those forecasts. We believe this occurred because participants in our study believed that both entities—the algorithm and the past participant—were inevitably going to err.

[3] For the studies in this paper, we define "using an algorithm" as the amount of positive weight given to the algorithm's forecast. By this definition, algorithm use results in forecasts that are closer to the algorithm's forecasts than they would have otherwise been. Note that algorithm use is a not a binary variable, as greater reliance on the algorithm results in forecasts that are even closer to it.

[4] Attrition did not significantly differ across conditions in any of the studies in this paper. See the online supplementary materials for a full breakdown of attrition across studies.

[5] For Study 1, these demographics exclude eight participants who did not report their gender and age. The demographics that we report in Studies 2 and 3 exclude seven and eight participants, respectively, for the same reason.

[6] See the online supplementary materials for a more detailed description of these data and the statistical model.

[7] For can't-change, "If you choose to use the statistical model's estimates, you will not be able to change the model's estimates." For adjust-by-10, "If you choose to use the statistical model's estimates, you will be able adjust the model's estimate for each student by up to 10 percentiles." For change-10, "If you choose to use the statistical model's estimates, you will be able to overrule 10 of the model's estimates and use your own estimates instead." For use-freely, "For the 20 official estimates, you can choose to use the model's estimated percentiles as much as you would like to."

[8] The first option was, "Use only the statistical model's estimated percentiles to determine my bonus" for the can't-change condition; "Use the statistical model's estimated percentiles to determine my bonus, adjusting them up to 10 percentiles if need be" for the adjust-by-10 condition; and "Use the statistical model's estimated percentiles to determine my bonus, overruling up to 10 of them if need be" for the change-10 condition. The second option was, "Use only my estimated percentiles to determine my bonus" for all three conditions.

[9] We did not find interesting differences between conditions on the performance estimates and confidence measures in Studies 1 and 2. Thus, we report the results of these measures in the online supplementary materials.

[10] Participants in the use-freely and can't-change conditions also learned how they performed compared with participants from the same condition in a previous study (Study S1 in the online supplementary materials), reported their confidence in the model's forecasts and their own forecasts on five-point scales, and reported their likelihood of using the model to complete this task in the future on five-point scales. These questions were exploratory. We did not include them in any other study, and we do not discuss them further.

[11] Participants in the can't-change condition made larger errors on average than participants in the adjust-by-10 ($t(144) = 3.40$, $p < 0.001$), change-10 ($t(143) = 3.09$, $p = 0.002$), and use-freely ($t(143) = 4.01$, $p < 0.001$) conditions. This translated into participants in the can't-change condition earning smaller bonuses than participants in the adjust-by-10 ($t(144) = -2.90$, $p = 0.004$), change-10 ($t(143) = -2.53$, $p = 0.013$), and use-freely ($t(143) = -2.88$, $p = 0.005$) conditions.

[12] Participants in the can't-change condition made significantly larger errors on average than participants in the adjust-by-10 ($t(407) = 2.64$, $p = 0.009$), adjust-by-5 ($t(409) = 4.02$, $p < 0.001$), and adjust-by-2 ($t(406) = 2.85$, $p = 0.005$) conditions. As a result, participants in the can't-change condition earned significantly smaller bonuses than participants in the adjust-by-10 ($t(407) = -2.08$, $p = 0.039$), adjust-by-5 ($t(409) = -3.67$, $p < 0.001$), and adjust-by-2 ($t(406) = -2.04$, $p = 0.042$) conditions.

[13] For model-only, "For the following 10 estimates, you will use the model's estimates. You will not be able to change the model's estimates." For use-freely, "For the following 10 estimates, you can use the model's estimates as much as you would like to. You will see the model's estimate and you can use it to form your estimate." For adjust-by-10, "For the following 10 estimates, you will use the model's estimates. You will be able adjust the model's estimate for each student by up to 10 percentiles."

[14] Unlike in Studies 1 and 2, participants who could not change the model's forecasts did not make their own forecasts. Instead, they simply viewed the model's forecast for each student.

[15] We do not report the results of these ratings in the paper as it is not clear if the differences among conditions are due to performance feedback or due to participants being exposed to two additional forecasting processes while making their ratings. The results of these measures are presented in Figure S3 in the online supplementary materials.

[16] Participants in the model-only condition did not make their own forecasts and so did not receive explicit feedback about their performance. This may have contributed to their relative preference for their own judgment.

[17] These analyses were conducted with ordinary least squares regressions that included participants' confidence ratings or average absolute error (AAE) estimates as the dependent variable, with two observations per participant (confidence in model and confidence in human, or model AAE estimate and human AAE estimate). The regressions included a dummy indicating whether each observation described the model or the participant, a dummy indicating whether or not the participant could adjust the model's forecasts during the first round, and an interaction between both dummies. Standard errors were clustered by participant. The t-tests reported correspond to the coefficient of the interaction term in each regression.

[18] We conducted a mediation analysis, where the dependent variable was participants' average absolute error, the mediator was participant's average absolute deviation from the model's forecasts, and the independent variable was whether participants used the adjust-by-10 process or the use-freely process. We included all participants who used the adjust-by-10 process or use-freely process for stage 2. We then used Preacher and Hayes's (2008) bootstrapping procedure to obtain unbiased 95% confidence intervals around the mediated effects. Average absolute deviation from the model's forecasts significantly mediated the effect of having the adjust-by-10 versus the use-freely option on average absolute error (−2.19, −0.85).

[19] Participants in the adjust-by-10 condition had lower average absolute errors than ($t(403) = -3.86$, $p < 0.001$), and earned more money than ($t(403) = 3.59$, $p < 0.001$), participants in the use-freely condition. They also had lower average absolute errors than ($t(404) = 3.92$,

$p < 0.001$), and earned more money than ($t(404) = -3.69$, $p < 0.001$), participants in the model-only condition who had the use-freely option. Participants in the model-only condition who had the adjust-by-10 option had lower average absolute errors than ($t(410) = -2.59$, $p = 0.010$), and earned more money than ($t(410) = 2.99$, $p = 0.003$), participants in the use-freely condition. They also had lower average absolute errors than ($t(411) = -2.71$, $p = 0.007$), and earned more money than ($t(411) = 3.10$, $p = 0.002$), participants in the model-only condition who had the use-freely option.

[20] All participants were told that the algorithm's or human's forecasts were off by 17.5 percentiles on average.

References

Adams ID, Chan M, Clifford PC, Cooke WM, Dallos V, De Dombal FT, McIntyre N (1986) Computer aided diagnosis of acute abdominal pain: A multicentre study. *British Medical J.* 293(6550):800–804.

Beck AH, Sangoi AR, Leung S, Marinelli RJ, Nielsen TO, van de Vijver MJ, Koller D (2011) Systematic analysis of breast cancer morphology uncovers stromal features associated with survival. *Sci. Translational Medicine* 3(108):108ra113.

Bonaccio S, Dalal RS (2006) Advice taking and decision-making: An integrative literature review, and implications for the organizational sciences. *Organ. Behav. Human Decision Processes* 101(2): 127–151.

Carbone R, Andersen A, Corriveau Y, Corson PP (1983) Comparing for different time series methods the value of technical expertise individualized analysis, and judgmental adjustment. *Management Sci.* 29(5):559–566.

Dawes RM (1971) A case study of graduate admissions: Application of three principles of human decision making. *Amer. Psychologist* 26(2):180–188.

Dawes RM (1979) The robust beauty of improper linear models in decision making. *Amer. Psychologist* 34(7):571–582.

Dawes RM, Faust D, Meehl PE (1989) Clinical versus actuarial judgment. *Science* 243(4899):1668–1674.

Dietvorst BJ, Simmons JP, Massey C (2015) Algorithm aversion: People erroneously avoid algorithms after seeing them err. *J. Experiment. Psych.: General* 144(1):114–126.

Einhorn HJ (1986) Accepting error to make less error. *J. Personality Assessment* 50(3):387–395.

Fildes R, Goodwin P (2007) Against your better judgment? How organizations can improve their use of management judgment in forecasting. *Interfaces* 37(6):570–576.

Fildes R, Goodwin P, Lawrence M, Nikolopoulos K (2009) Effective forecasting and judgmental adjustments: An empirical evaluation and strategies for improvement in supply-chain planning. *Internat. J. Forecasting* 25(1):3–23.

Finucane ML, Alhakami A, Slovic P, Johnson SM (2000) The affect heuristic in judgments of risks and benefits. *J. Behav. Decision Making* 13(1):1–17.

Goodwin P, Fildes R (1999) Judgmental forecasts of time series affected by special events: Does providing a statistical forecast improve accuracy? *J. Behav. Decision Making* 12(1):37–53.

Grove WM, Zald DH, Lebow BS, Snitz BE, Nelson C (2000) Clinical versus mechanical prediction: A meta-analysis. *Psych. Assessment* 12(1):19–30.

Highhouse S (2008) Stubborn reliance on intuition and subjectivity in employee selection. *Indust. Organ. Psych.* 1(3):333–342.

Hogarth RM, Makridakis S (1981) Forecasting and planning: An evaluation. *Management Sci.* 27(2):115–138.

Lawrence M, Goodwin P, O'Connor M, Önkal D (2006) Judgmental forecasting: A review of progress over the last 25 years. *Internat. J. Forecasting* 22(3):493–518.

Lim JS, O'Connor M (1995) Judgemental adjustment of initial forecasts: Its effectiveness and biases. *J. Behav. Decision Making* 8(3):149–168.

Meehl PE (1954) *Clinical versus Statistical Prediction: A Theoretical Analysis and Review of the Literature* (University of Minnesota Press, Minneapolis).

Preacher KJ, Hayes AF (2008) Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behav. Res. Methods* 40(3):879–891.

Sanders NR, Manrodt KB (2003) The efficacy of using judgmental versus quantitative forecasting methods in practice. *Omega* 31(6):511–522.

Schweitzer ME, Cachon GP (2000) Decision bias in the newsvendor problem with a known demand distribution: Experimental evidence. *Management Sci.* 46(3):404–420.

Thompson RE (1952) A validation of the Glueck Social Prediction Scale for proneness to delinquency. *J. Criminal Law: Criminology Police Sci.* 43(4):451–470.

Vrieze SI, Grove WM (2009) Survey on the use of clinical and mechanical prediction methods in clinical psychology. *Professional Psych.: Res. Practice* 40(5):525–531.

Willemain TR (1991) The effect of graphical adjustment on forecast accuracy. *Internat. J. Forecasting* 7(2):151–154.

Wormith JS, Goldstone CS (1984) The clinical and statistical prediction of recidivism. *Criminal Justice Behav.* 11(1):3–34.