Running Head: OVERCOMING ALGORITHM AVERSION

Overcoming Algorithm Aversion:

People Will Use Algorithms If They Can (Even Slightly) Modify Them

Berkeley J. Dietvorst Joseph P. Simmons Cade Massey

University of Pennsylvania

Corresponding Author:

Berkeley J. Dietvorst
The Wharton School
University of Pennsylvania
500 Jon M. Huntsman Hall
3730 Walnut Street
Philadelphia, PA 19104
diet@wharton.upenn.edu

Abstract

Although evidence-based algorithms consistently outperform human forecasters, people consistently fail to use them, especially after learning that they are imperfect. In this paper, we investigate how *algorithm aversion* might be overcome. In incentivized forecasting tasks, we find that people are considerably more likely to choose to use an algorithm, and thus perform better, when they can modify its forecasts. Importantly, this is true even when they are severely restricted in the modifications they can make. In fact, people's decision to use an algorithm is insensitive to the magnitude of the modifications they are able to make. Additionally, we find that giving people the freedom to modify an algorithm makes people feel more satisfied with the forecasting process, more tolerant of errors, more likely to believe that the algorithm is superior, and more likely to choose to use an algorithm to make subsequent forecasts. This research suggests that one may be able to overcome algorithm aversion by giving people just a slight amount of control over the algorithm's forecasts.

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, documented by decades of research, has been observed in many different domains, including forecasts of employee performance (see Highhouse, 2008), academic performance (Dawes, 1971; Dawes, 1979), prisoners' likelihood of recidivism (Thompson, 1952; Wormith & Goldstone, 1984), medical diagnoses (Adams et al., 1986; Beck et al., 2011; Dawes, Faust, & Meehl, 1989; Grove et al., 2000), demand for products (Schweitzer & Cachon, 2000), and so on (see Dawes, Faust, & Meehl, 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, and 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. Other studies show that people prefer to have humans integrate information (Diab, Pui, Yankelvich, & Highhouse, 2011; Eastwood, Snook, & Luther, 2012), and that they give more weight to forecasts made by experts than to forecasts made by algorithms (Onkal et al., 2009; Promberger & Baron, 2006). *Algorithm aversion* is especially pronounced when people have seen an algorithm err, even when they have seen that it errs less than humans do (Dietvorst, Simmons, & Massey, 2015).

Algorithm aversion represents a major challenge for any organization interested in making accurate forecasts and good decisions, and for organizations that would benefit from their customers using algorithms to make better choices. In this article, we offer an approach for overcoming algorithm aversion.

[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.

Overcoming Algorithm Aversion

Many scholars have theorized about why decision makers are reluctant to use algorithms that outperform human forecasters. One common theme is an intolerance of error. Einhorn (1986) proposed that algorithm aversion 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 even when people expected both humans and algorithms to make mistakes, and thus were resigned to the inevitability of error, they 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.

Fortunately, people's distaste for algorithms may be rooted in more than just an intolerance of error, but also in their beliefs about the qualities of human vs. algorithmic forecasts. Dietvorst et al. (2015) found that although people tend to think that algorithms are better than humans at avoiding obvious mistakes, appropriately weighing attributes, and consistently weighing information, they tend to think that humans are better than algorithms at learning from mistakes, getting better with practice, finding diamonds in the rough, and detecting exceptions to the rule. Indeed, people seem to believe that although algorithms are better than humans *on average*, the rigidity of algorithms means they may predictably misfire in any given instance.

This suggests that what people may find especially distasteful about using algorithms is the lack of flexibility, the inability to intervene when they suspect that the algorithm has it wrong. If this is true, then people may be more open to using an algorithm if they are allowed to slightly or occasionally alter its judgments. Although people's attempts to adjust algorithmic forecasts often make them worse (e.g. Carbone, Andersen, Corriveau, & Corson, 1983; Goodwin & Fildes, 1999; Hogarth & Makridakis, 1981; Lim & O'Connor, 1995; Willemain, 1991) the benefits associated with getting people to use the algorithm may outweigh the costs associated with making the algorithm's forecasts slightly worse. 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 the algorithm by only a tiny amount dramatically reduces algorithm aversion, then people's judgments will be much more reliant on the algorithm, and much more accurate as a result.

In this article, we explore whether people are more likely to use an algorithm for forecasting when they can restrictively modify its forecasts. In the first two studies, we find that giving people the ability to adjust an algorithm's forecasts decreases algorithm aversion and improves forecasts. Interestingly, in Study 3 we find that people's openness to using algorithms does not depend on how much they are allowed to adjust them; allowing people to adjust an algorithm's forecasts increases their likelihood of using the algorithm even if we severely restrict the amount by which they can adjust it. In Study 4, we explore the downstream consequences of allowing people to slightly modify an algorithm's forecasts. We find that allowing people to adjust an algorithm's forecasts increases their satisfaction with their forecasting process, prevents them from losing confidence in the algorithm after it errs, and increases their willingness to continue using the algorithm after receiving feedback. We also find that allowing people to adjust an algorithm's forecasts by a limited amount leads to better long-term performance than allowing them to adjust an algorithm's forecasts by an unlimited amount.

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 at http://opim.wharton.upenn.edu/bcs329/OvercomingAlgorithmAversion/.

Study 1

Methods

Overview. In Study 1, we asked participants to forecast students' scores on a standardized test from nine variables. All participants had the option of using a statistical model to make their forecasts, and we manipulated whether participants had the option to modify the model's forecasts. Participants 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. Compared to those who had to choose whether or not to use the model's forecasts exclusively or not at all, we expected participants who were restrictively able to modify the model's forecasts to be much more open to using the model, and to perform better as a result. We were also curious to learn whether some types of restrictive adjustments were better than others, and to see how much weight participants would give to the model's forecasts when they were free to use the model as much as they wanted.

Participants. This study was conducted in our university's behavioral lab. Participants received $10 for completing one hour of experiments, of which ours was a 20-minute portion. Participants could earn up to a $5 bonus from our study 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 lab at our university has two separate locations) and collected as many participants as we could. The behavioral lab failed to stop 19 participants who had already taken the study from taking it again. 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. This sample averaged 22 years of age and was 66% female.

Procedures. This experiment was administered as an online survey. Participants began by giving consent and entering their lab identification number. Next, they learned about the experimental judgment task; 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.[2] Figure 1 shows an example of the stimuli and variables.

Figure 1.

Example of task stimuli used in Studies 1, 3, and 4.

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

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 5 percentiles of students' actual percentiles on average, and this bonus decreased by $1 for each additional 5 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."

[2] See the supplement for a more detailed description of this data and the statistical model.

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.[3]

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.[4] After making this choice, participants made 20 incentivized forecasts. The 20 students that participants judged were randomly drawn from a pool of 50 randomly selected high school seniors without replacement. The high school students were each 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

[3] Can't-change: "If you choose to use the statistical model's estimates, you will not be able to change the model's estimates." Adjust-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." 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." Use-freely: "For the 20 official estimates, you can choose to use the model's estimated percentiles as much as you would like to."

[4] 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.

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 5-point scales (1=*none*; 5=*a lot*), and answered two open-ended questions.[5] 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.[6] 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 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, $\chi^2(1, N = 145) = 24.19$, $p < .001$, and 76% of participants in the adjust-by-10 condition, $\chi^2(1, N = 146) = 28.40$, $p < .001$, chose to use the model.

Interestingly, participants who chose to use the model in the adjust-by-10 and change-10 conditions did not deviate from the model as much as they could have. Participants who chose to use the model in the adjust-by-10 condition provided forecasts that were 4.71 percentiles away from the model's forecasts

[5] We did not find interesting differences between conditions for the performance estimates and confidence measures in Studies 1-3. Thus, we report the results of these measures in the Online Supplement.

[6] Participants in the use-freely and can't-change conditions also learned how they performed compared to participants from the same condition in a previous study (Study S1 in the Supplement), reported their confidence in the model's forecasts and their own forecasts on 5-point scales, and reported their likelihood of using the model to complete this task in the future on 5-point scales. These questions were exploratory and we do not discuss them further.

on average, far less deviation than the 10 percentiles of adjustment that they were allowed. Participants who chose to use the model in the change-10 condition changed the model 8.54 times on average, and only 39% used all 10 of their changes.

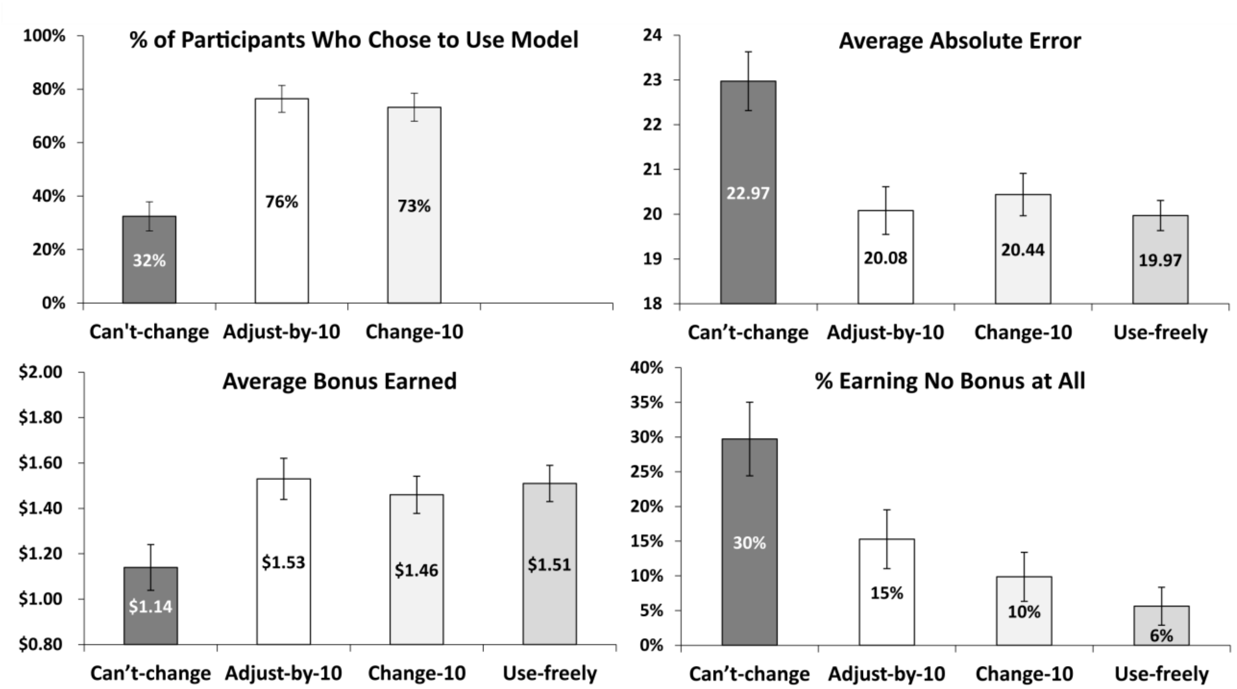
Also interesting is that there were differences between conditions in how much participants' forecasts deviated from the model. First, those who chose to use the model in the change-10 condition made larger adjustments to the model's forecasts than did those in the adjust-by-10 condition, altering them by 10.58 percentiles on average, $t(105) = -10.20$, $p < .001$. Although the adjust-by-10 and change-10 conditions performed similarly in Study 1, this result suggests that restricting the amount by which people can adjust from the model may be superior to restricting the number of unlimited adjustments they can make to the model. Second, the forecasts of those in the use-freely condition were in between, deviating more from the model (M = 8.18) than those in the adjust-by-10 condition, $t(124) = -6.17$, $p < .001$, but less than those in the change-10 condition, $t(121) = 3.28$, $p = .001$.[7,8] Although average deviations of 5-11 percentiles may sound like a lot, they are small compared to the average deviation of 18.66 among participants in the can't-change condition, who made forecasts without seeing the model's forecasts.

[7] These t-tests compare the average adjustment across all 20 trials in the adjust-by-10 and use-freely conditions to the average adjustment made on the 10 changeable trials in the change-10 condition. If participants in the change-10 condition altered fewer than 10 trials, then we coded the remaining changeable trials as having adjustments of zero. For example, if a participant in the change-10 condition altered 5 trials by an average of 10 percentiles, then her average adjustment was 5.00, because on 5 of the changeable trials she adjusted zero percentiles away from the model. Alternatively, we could have restricted our comparison to trials on which participants' forecasts actually deviated from the model. This analysis reveals a similar result: the largest adjustment in the change-10 condition ($M = 12.52$), the smallest adjustment in the adjust-by-10 condition ($M = 5.16$), with the use-freely condition in between ($M = 9.10$), all p's < .001.

[8] The fact that those in the use-freely condition deviated less from the model than those in those in the change-10 condition is the result of a selection effect: We compared the trials that participants in the change-10 condition *selected* to alter to *all* of the trials in the use-freely condition. Because those who chose to the use the model in the change-10 condition could only alter the 10 forecasts that they believed to be most likely to need altering, it is more appropriate to compare the change-10's adjustments on these trials to the use-freely condition's 10 most extreme adjustments. When we do this, the use-freely condition adjusted *more* ($M = 12.82$) than the change-10 condition ($M = 10.58$), $t(121) = -2.27$, $p = .025$.

Figure 2

Study 1: Participants who could restrictively modify the model's forecasts were more likely to choose to use the model, and performed better as a result.



Note: Errors bars indicate ±1 standard error.

Forecasting performance. As shown in Figure 2, participants who had the option to adjust the model's forecasts outperformed those who did not. The forecasts of participants 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.[9]

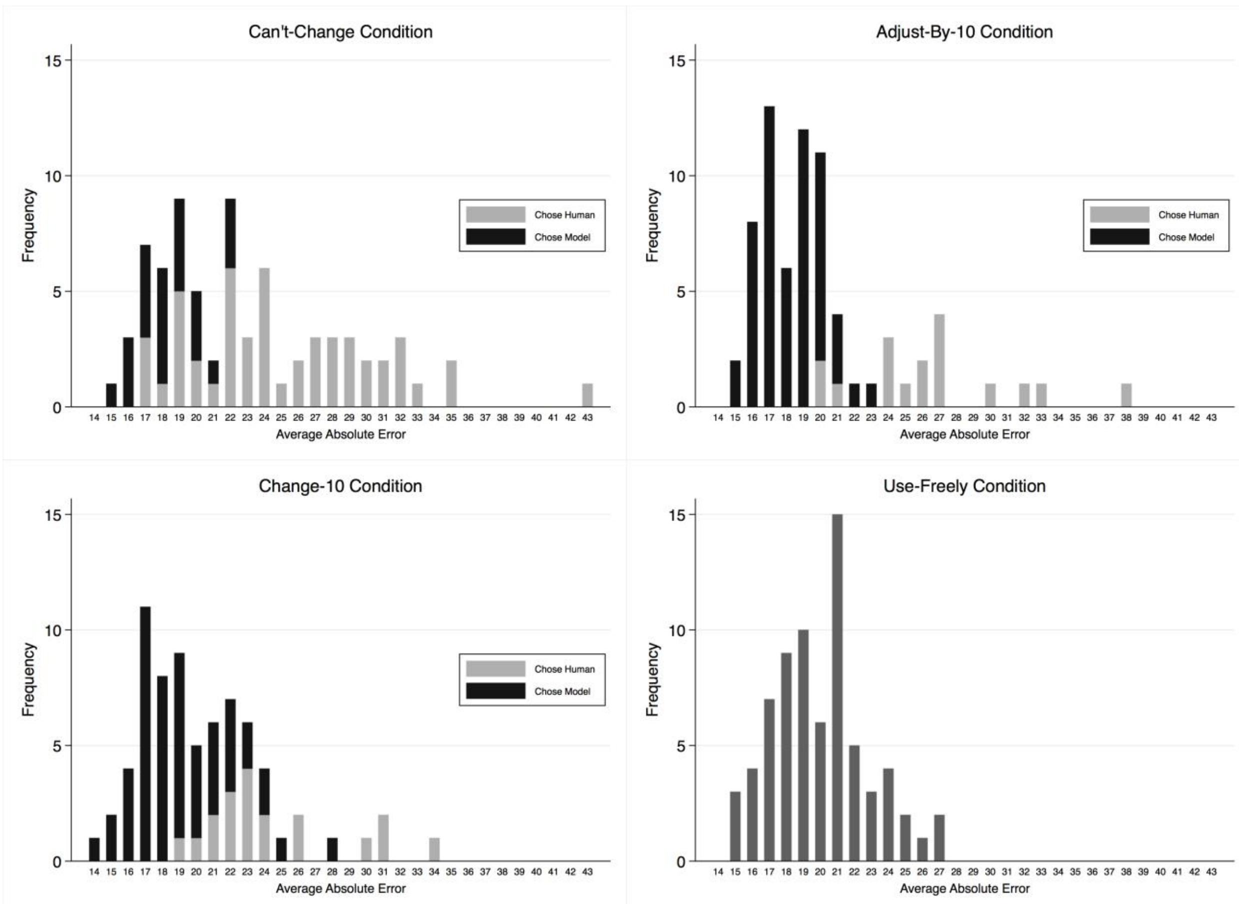
Figure 3 displays the distribution of participants' performance by condition. Three things are apparent from the figure. First, reliance on the model was strongly associated with better performance. Indeed, failing to choose to use the model was much more likely to result in very large average errors (and bonuses of $0). Second, participants in the can't-change condition performed worse precisely because

[9] Participants in the can't-change condition made larger errors on average than participants in the adjust-by-10, $t(144) = 3.40$, $p < .001$, change-10, $t(143) = 3.09$, $p = .002$, and use-freely, $t(143) = 4.01$, $p < .001$, conditions. This translated into participants in the can't-change condition earning smaller bonuses than participants in the adjust-by-10, $t(144) = -2.90$, $p = .004$, change-10, $t(143) = -2.53$, $p = .013$, and use-freely, $t(143) = -2.88$, $p = .005$, conditions.

they were less likely to use the model, and not because their forecasting ability was worse. Third, participants' use of the model in the use-freely condition seems to have prevented them from making very large errors, as no participant erred by more than 28 percentiles on average.

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.



Discussion. In sum, participants who could restrictively modify the model's 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 also seemed to anchor on the model's forecasts, which improved their performance by reducing their chances of making large errors.

Study 2

Study 2 was a replication of Study 1 with a different forecasting task and a different participant population.

Methods

Participants. We ran Study 2 with participants from Amazon Mechanical Turk (MTurk). Participants earned $1 for completing the study and could earn up to an additional $0.60 for good forecasting performance. We decided in advance to recruit 1000 participants (250 per condition). Participants began the study by answering a question designed to check whether they were carefully reading instructions. We prevented the 223 participants who failed this check from participating and 297 additional participants quit the survey before completing their forecasts. We replaced these participants, and our final sample consisted of 1,040 participants who completed their forecasts. This sample averaged 33 years of age and was 53% female.

Procedure. The procedure was the same as Study 1's except for five changes. First, we used a different forecasting task. Participants forecasted the rank (1-50) of individual U.S. states in terms of their number of departing airline passengers in 2011. Participants received the following information to make forecasts about each state: the state's name, number of major airports (as defined by the Bureau of Transportation), 2010 census population rank (1 to 50), total number of counties rank (1 to 50), 2008 median household income rank (1 to 50), and 2009 domestic travel expenditure rank (1 to 50).[10] Figure 4 shows an example of the stimuli that participants saw during the forecasting task. The 20 states that participants judged were randomly drawn without replacement from a pool of all 50 states. The model's forecasts were off by 4.3 ranks on average, and participants were told this.

[10] See the supplement for a more detailed description of this data and the statistical model.

Figure 4.

Example of task stimuli used in Study 2.

State	Colorado
Number of Major Airports	1
Census Population Rank - 2010	22
Number of Counties Rank	24
Median Household Income Rank - 2008	13
Domestic Travel Expenditure Rank - 2009	16

Second, as previously mentioned, we added a reading check to the beginning of the survey to identify and remove participants who were not reading instructions. Third, because the range of possible forecasts was 1-50 instead of 1-100, we replaced the adjust-by-10 condition with an adjust-by-5 condition. Fourth, we used a different payment rule. Participants were paid $0.60 if their forecasts were within 1 rank of states' actual ranks on average; this bonus decreased by $0.10 for each additional unit of error in participants' forecasts (this payment rule is reproduced in Appendix B). As a result, participants whose forecasts were off by more than 6 ranks received no bonus. Fifth, at the end of the survey we asked participants to recall the model's average error.

Results

Choosing to use the model. As in Study 1, giving participants the option to restrictively adjust the model's forecasts increased their likelihood of choosing to use the model (see Figure 5). Whereas only 47% of participants in the can't-change condition chose to use the model's forecasts, 77% of participants in the change-10 condition, $\chi^2(1, N = 542) = 49.37$, $p < .001$, and 75% of participants in the adjust-by-5 condition, $\chi^2(1, N = 530) = 44.33$, $p < .001$, chose to use the model.

Also consistent with Study 1, participants who chose to use the model in the adjust-by-5 and change-10 conditions did not deviate from the model as much as they could have. Though they were allowed to adjust by 5 ranks, participants who chose to use the model in the adjust-by-5 condition provided forecasts that were only 1.83 ranks away from the model's ranks on average. Participants who chose to use the model in the change-10 condition changed the model 7.44 times on average, and only 31% used all 10 of their changes.

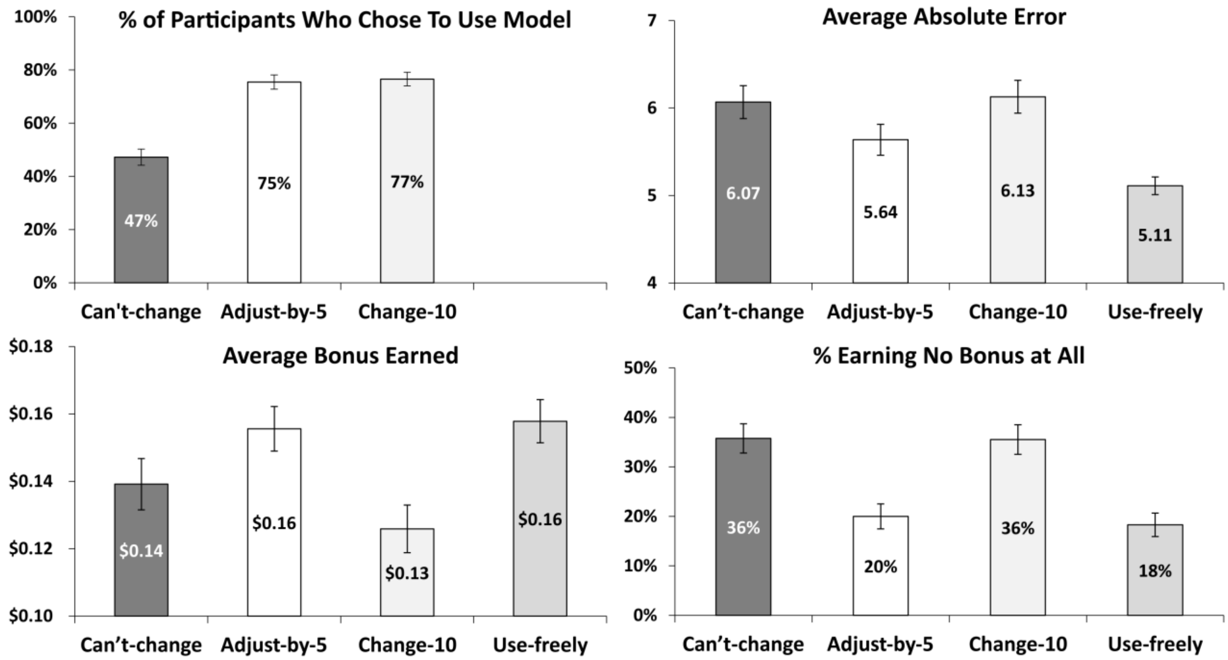
There were again differences between conditions in how much participants' forecasts deviated from the model when they did choose to use it. First, participants in the change-10 condition made larger adjustments to the model's forecasts (*M* = 4.17) than did those in the adjust-by-5 condition (*M* = 1.83), t(388) = -9.21, *p* <.001. As shown in the next section, the performance of those in the change-10 condition suffered as a result. Second, the forecasts of those in the use-freely condition were again in between, deviating more from the model (*M* = 2.64) than those in the adjust-by-5 condition, *t*(457) = -4.81, *p* < .001, but less than those in the change-10 condition, *t*(465) = 5.86, *p* < .001.[11,12] The deviations of 2-4 ranks exhibited by these conditions were small in comparison to those made by participants in the can't-change condition, who made forecasts without seeing the model's forecasts. (M = 7.91).

[11] As in Study 1, these t-tests compare the average adjustment across all 20 trials in the adjust-by-5 and use-freely conditions to the average adjustment made on the 10 changeable trials in the change-10 condition. If we instead restrict our comparison to trials on which participants' forecasts actually deviated from the model, we get a similar result: the largest adjustment in the change-10 condition (*M* = 5.37), the smallest adjustment in the adjust-by-10 condition (*M* = 2.38), with the use-freely condition in between (*M* = 3.40), all *p*'s < .001.

[12] As in Study 1, this difference between the use-freely condition and the change-10 condition is the result of a selection effect. When we compare the change-10's adjustments on their 10 alterable trials to the use-freely condition's 10 most extreme adjustments, the use-freely condition adjusted non-significantly *more* (*M* = 4.44) than the change-10 condition (*M* = 4.17), *t*(465) = -0.81, *p* = .420.

Figure 5

Study 2: Participants who could restrictively modify the model's forecasts were more likely to choose to use the model, and the adjust-by-5 condition performed better as a result.



Note: Errors bars indicate ±1 standard error.

Forecasting performance. As shown in Figure 5, participants who were restrictively able to adjust the model's forecasts, or who were able to freely use the model, outperformed those who could not adjust the model's forecasts. Unexpectedly, participants in the change-10 condition, who were able to adjust up to 10 forecasts as much as they liked, performed about the same as those in the can't-change condition, and worse than those in the adjust-by-5 and use-freely conditions.[13] This had two causes. First, those in the change-10 condition performed even worse (M = 9.02) than the can't-change condition (M = 7.70) when they chose not to use the model, $t(196) = -2.30$, $p = .023$; this may reflect a selection effect, such that those who opted not to use the model in the change-10 (and adjust-by-5) conditions may have been

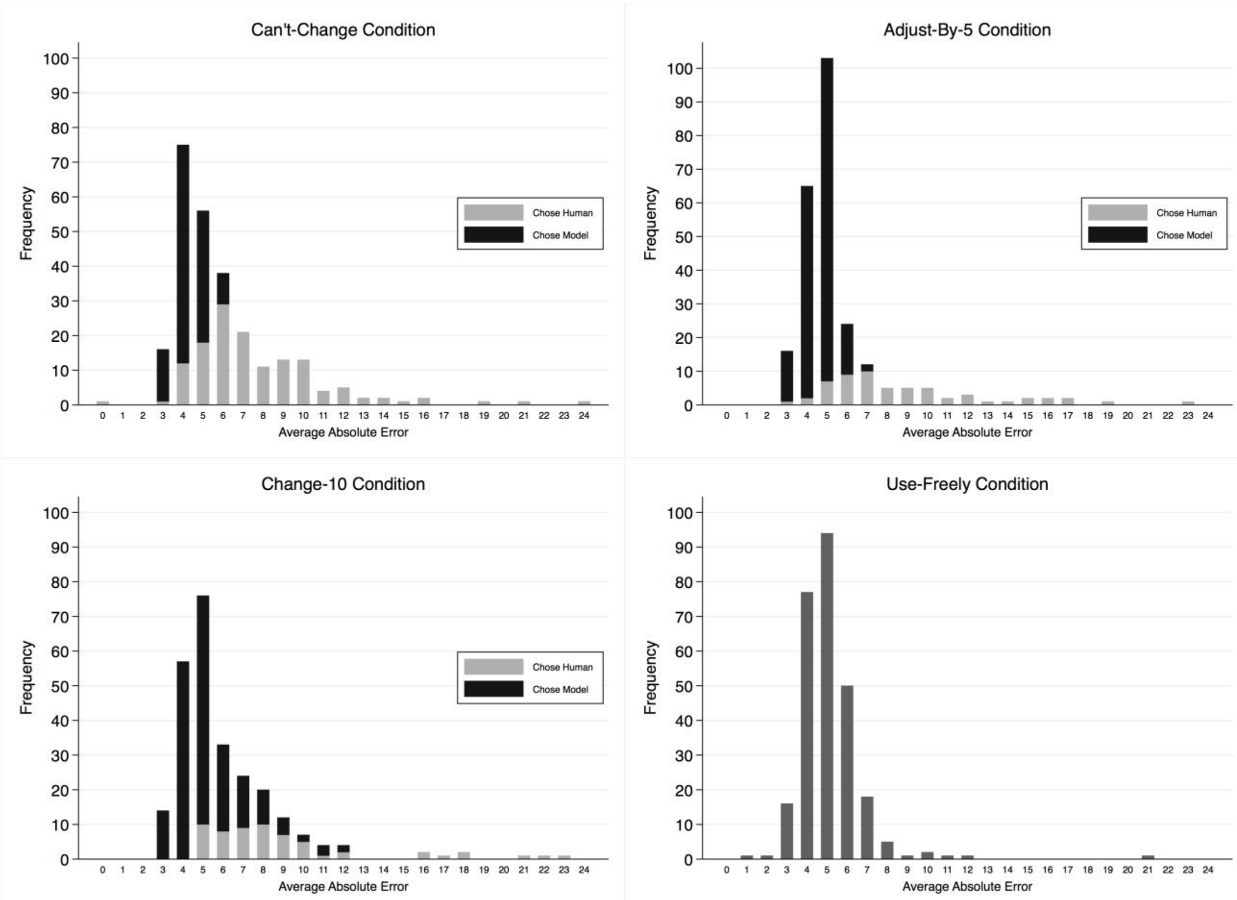
[13] Participants in the can't-change condition made marginally larger errors on average than participants in the adjust-by-5 condition, $t(511) = 1.67$, $p = .097$, and significantly larger errors than those in the use-freely condition, $t(529) = 4.52$, $p < .001$. Participants in the change-10 condition performed similarly to those in the can't change condition, $t(520) = -0.23$, $p = .818$, and worse than those in the adjust-by-5, $t(507) = -1.90$, $p = .058$, and use-freely conditions, $t(525) = 4.82$, $p < .001$. As a result, participants in the can't-change condition earned marginally smaller bonuses than participants in the adjust-by-5, $t(511) = -1.62$, $p = .105$, and use-freely, $t(529) = -1.88$, $p = .061$, conditions. And participants in the change-10 condition earned smaller bonuses than participants in the adjust-by-5, $t(507) = 3.06$, $p = .002$, and use-freely, $t(525) = -3.35$, $p < .001$, conditions.

more careless forecasters than those who did. Second, because even some participants who did use the model in the change-10 condition deviated considerably from the model on their changeable trials, even their performance was mediocre (see Figure 6).

Figure 6 displays the distribution of participants' performance by condition. We again see that reliance on the model was strongly associated with better performance, and that 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. We also again see that participants' use of the model in the use-freely condition seems to have prevented them from making very large errors, possibly because all participants in the use-freely condition were exposed to the model's forecasts and anchored on them to some degree.

Figure 6

Study 2: The distribution of participants' average absolute errors by condition and whether or not they chose to use the model's forecasts.



Discussion. This study replicated the choice results of Study 1; giving participants the option to adjust the model, either restrictively or freely, increased their use of the model. However, deviations from the model were larger, and performance worse, when allowing participants to freely alter 10 of the forecasts than when allowing them to adjust each forecast by a limited amount. This result does not seem to be anomalous as it is consistent with the results of Study S2 (in the online supplemental materials). Participants in the change-10 condition of Study S2 made large adjustments to the model's forecasts and did not perform better than those in the can't-change condition, even though they chose to use the model's forecasts significantly more often than participants in the can't-change condition. Since the adjust-by-5 and change-10 processes were equally attractive options for participants in this study and the change-10 process does not consistently improve performance, this suggests that limiting the amount by which people can change the model may be better than limiting the number of unlimited adjustments they can make. We thus focus the remainder of our investigation on conditions that restrict the amount of adjustment, rather than the number of trials on which they can make large adjustments.

Study 3

Methods

Overview. Studies 1 and 2 showed that people were more likely to choose to use an algorithm if they were given the option to restrictively adjust its forecasts. In Study 3, 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 model, or would people be willing to commit to using a model 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 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 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. MTurk participants 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 and 131 additional participants quit the survey before completing their forecasts. We replaced these participants, and our final sample consisted of 816 participants who completed their forecasts. This sample averaged 34 years of age and was 48% female.

Procedure. This study used the forecasting task from Study 1, in which participants predicted the percentiles of high school students on a standardized math test. The procedure was the same as Study 1's except for five 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 2 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 C). 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.

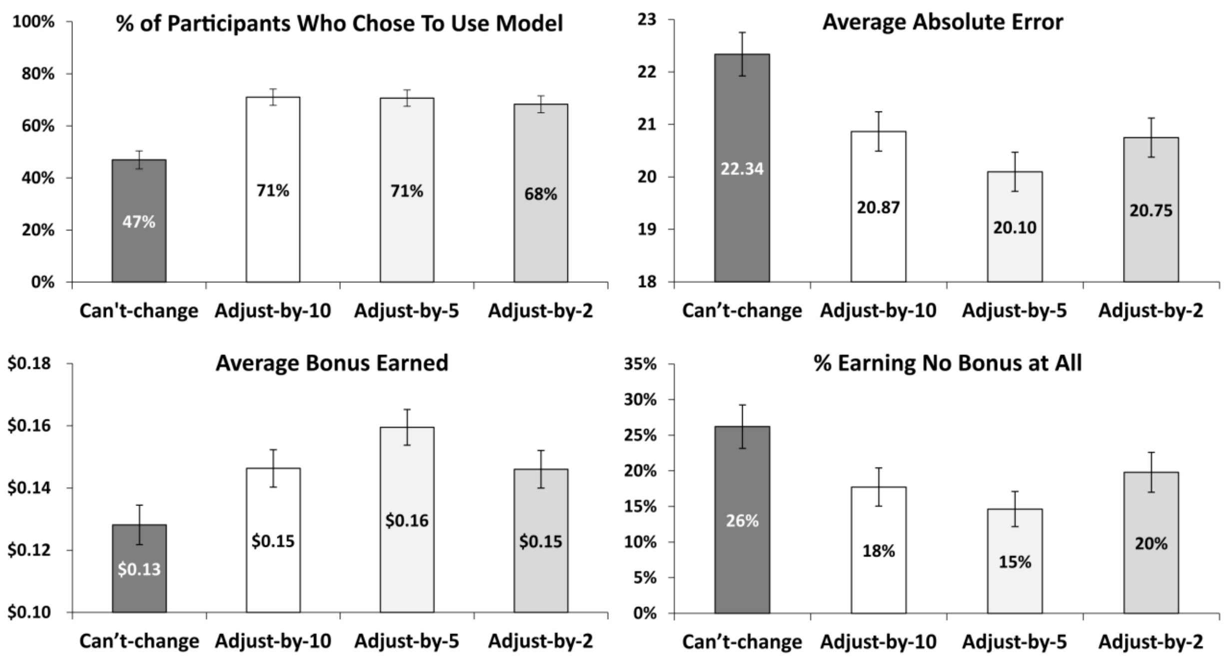
Results

Choosing to use the model. Consistent with the results of Studies 1 and 2, participants who had the option to adjust the model's forecasts chose to use the model more often than participants who could not modify its forecasts (see Figure 7). 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, $\chi^2(1, N = 834) = 36.46$, $p < .001$. 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. These three conditions did not differ significantly, $\chi^2(2, N = 623) = 0.42$, $p = .809$. 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 a fivefold restriction on the amount by which they could adjust. (See Study S2 in the supplement for a replication of this insensitivity using the change-X forecasting process).

Whether they were in the adjust-by-10 ($M = 5.00$), adjust-by-5 ($M = 2.61$), or adjust-by-2 condition ($M = 1.33$), participants who chose to the use the model did not deviate from the model as much as they could have. This is again surprising. Given the desire of participants in the adjust-by-10 condition to adjust by 5 percentiles on average, it is surprising that those in the adjust-by-5 and adjust-by-2 conditions did not adjust by close to the maximum amount. It seems that giving people the option to restrictively adjust the model's forecasts results in forecasts that are even closer to the model than they are required to be.

Figure 7

Study 3: Participants who could restrictively modify the model's forecasts were more likely to choose to use the model, and performed better as a result.



Note: Errors bars indicate ±1 standard error.

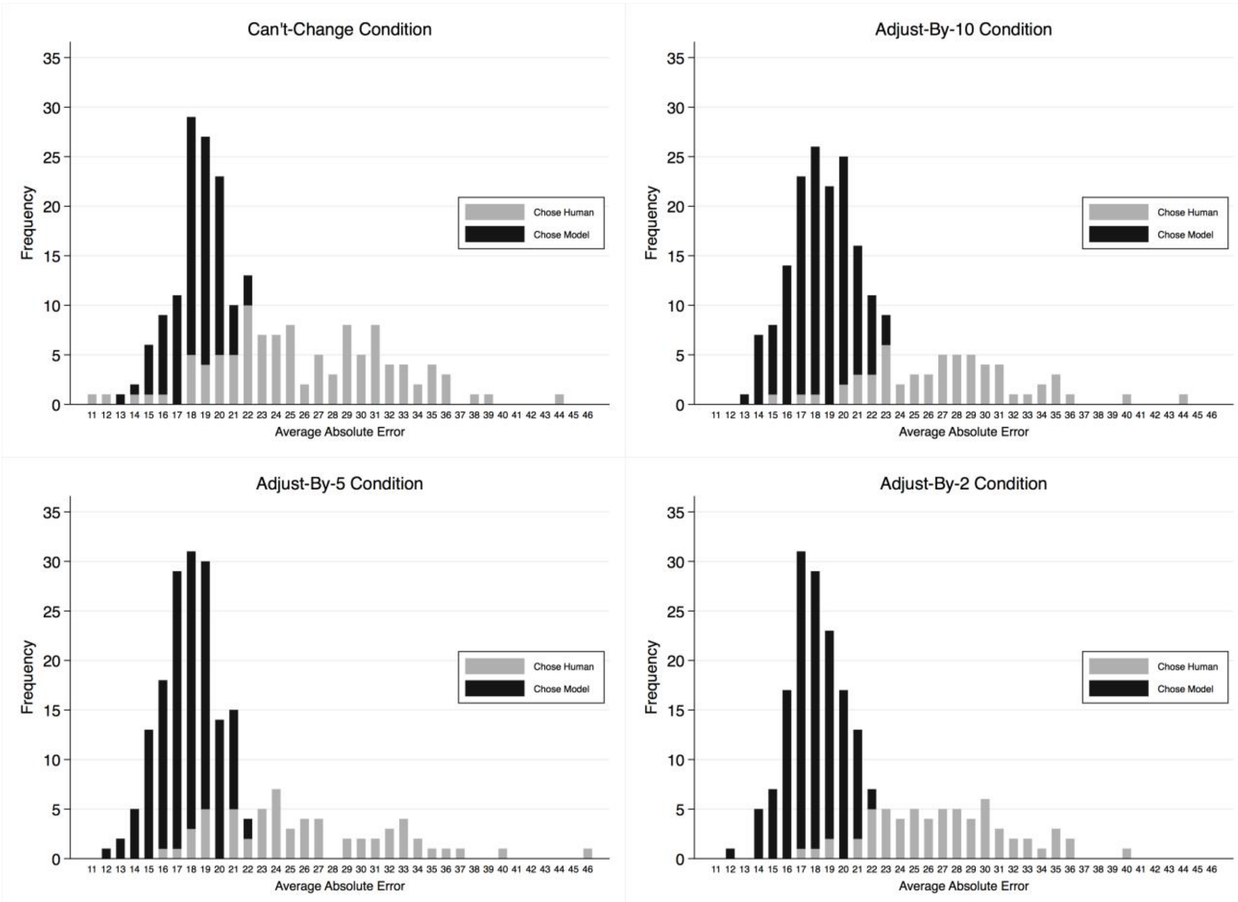
Forecasting performance. As in Studies 1 and 2, participants who were given the option to adjust the model's forecasts performed better than those who were not (see Figure 7). Participants in the can't-change condition made significantly larger errors than participants in each of the adjust-by-X conditions.[14]

Figure 8 displays the distribution of participants' performance by condition. We again see that reliance on the model was strongly associated with better performance, and that 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.

[14] Participants in the can't-change condition made significantly larger errors on average than participants in the adjust-by-10, $t(407) = 2.64$, $p = .009$, adjust-by-5, $t(409) = 4.02$, $p < .001$, and adjust-by-2, $t(406) = 2.85$, $p = .005$, conditions. As a result, participants in the can't-change condition earned significantly smaller bonuses than participants in the adjust-by-10, $t(407) = -2.08$, $p = .039$, adjust-by-5, $t(409) = -3.67$, $p < .001$, and adjust-by-2, $t(406) = -2.04$, $p = .042$, conditions.

Figure 8

Study 3: The distribution of participants' average absolute errors by condition and whether or not they chose to use the model's forecasts.



Discussion. In Study 3, participants were once again more likely to choose to use an 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 *some* control over an algorithm's forecasts, giving them additional control may not further reduce algorithm aversion.

Study 4

In Studies 1-3, we found that people were much more likely to choose to use an algorithm if they were allowed to adjust its forecasts by even a small amount. 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, overcoming algorithm aversion over the long term requires a willingness to use the algorithm even after using it. 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 4, we investigated how experience with different ways of using an algorithm affects people's subsequent decisions to use it. Using the same forecasting task as Studies 1 and 3, 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 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. They were then asked to indicate their satisfaction with, and confidence in, three forecasting processes that they would choose between for a second round of 10 forecasts. Half of the model-only participants, and all of the use-freely participants, chose among using the model for all forecasts, using themselves for all forecasts (without seeing the model's forecasts), and using the model freely. The other half of the model-only participants, and all of the adjust-by-10 participants, chose among using the model for all forecasts, using themselves for all forecasts (without seeing the model's forecasts), and using the model a restrictive amount (adjusting it by up to 10 percentiles).

This design allowed us to answer four open questions: (1) Are people more satisfied with a forecasting process that allows them to modify a model's forecasts than with one that does not? (2) Are people more forgiving of forecasting errors when they were able to modify the model's forecasts than when they were not? (3) Does giving people the opportunity to modify a model's forecasts make them think the model is better? (4) Are people more likely to choose to use the model when they were previously able to modify its forecasts than when they were not?

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 and 208 additional participants quit the survey before completing their forecasts. We replaced these participants, and had a sample of 818 participants who completed their forecasts. This 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 ID Number. They then completed 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) and the statistical model. This part of the survey was identical to Studies 1 and 3.

Figure 9 shows the rest of the procedure of Study 4. After reading about the forecasting task, participants were told that would make 10 forecasts and that their performance would be incentivized. Just as in Study 3, 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 required to type two sentences describing their condition's forecasting procedure to ensure that they understood the instructions.[15]

Next, participants completed their first set of 10 forecasts.[16] 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 5-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.

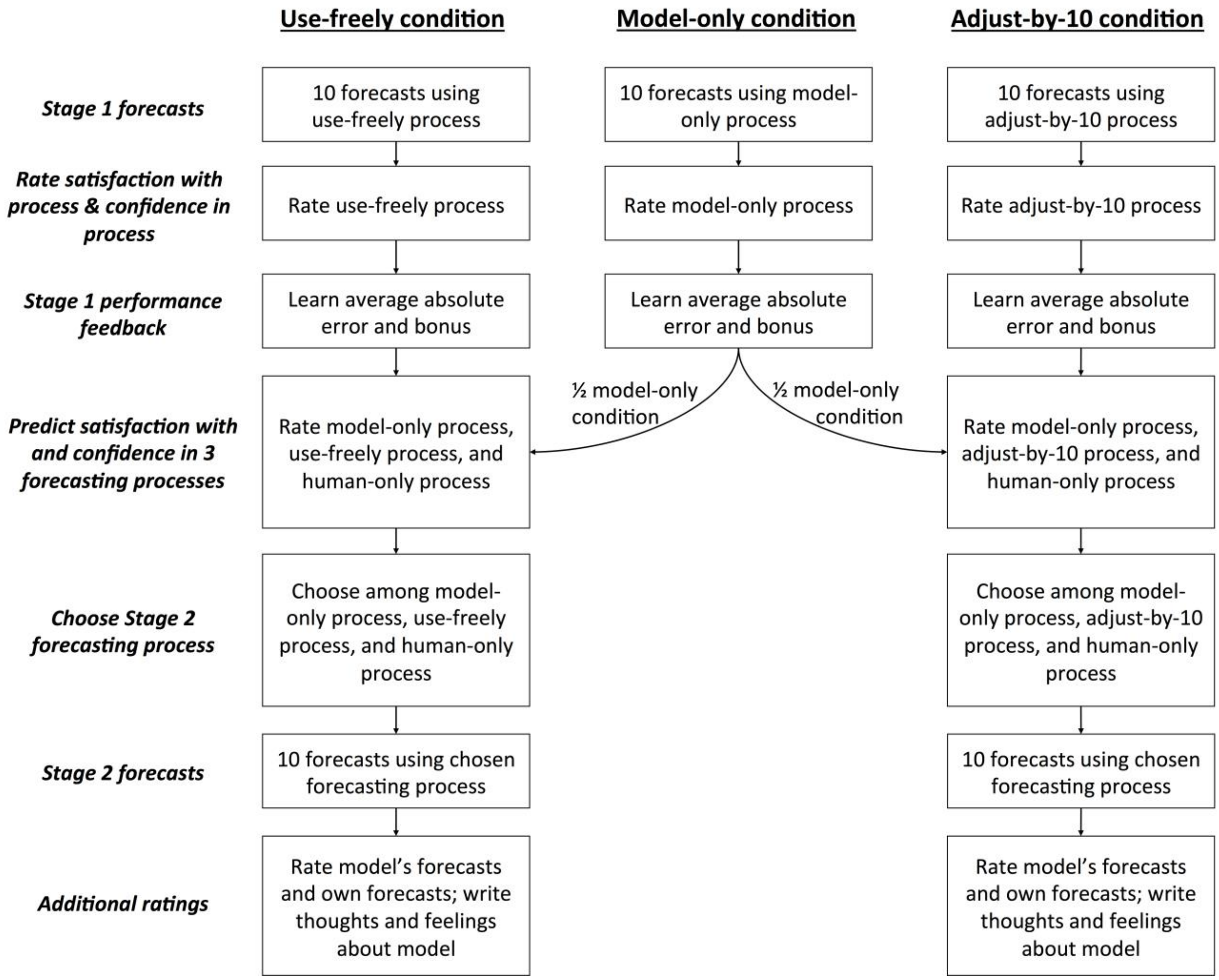
On the following page, participants were presented with three forecasting processes (see Figure 9 for a depiction of which three processes participants in each condition saw) and asked to rate how satisfied they would be with each one (1 = very dissatisfied; 5 = very satisfied), and how much confidence they have that the process would perform well (1 = none; 5 = a lot). This allowed us to see if participants' forecasting experience during the first round of forecasts affected which forecasting processes were attractive to them for the second round. Half of the model-only participants, and all of the use-freely participants, rated the following three processes: (1) using the model for every forecast (model-only), (2) using themselves for every forecast without seeing the model (human-only), and (3) using the model freely (with unlimited adjustment) for every forecast (use-freely). The other half of the model-only participants, and all of the adjust-by-10 participants, rated the following three processes: (1) model-only, (2) human-only, and (3) restrictively adjusting the model's forecasts by up to 10 percentiles (adjust-by-10).

[15] 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." 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." 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."

[16] Unlike Studies 1-3, 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.

Figure 9

Study 4's Procedure



After making these ratings, participants learned that their incentives for a second set of 10 forecasts were the same as the first round, and then chose among the same three forecasting processes that they had rated to use for the next set of forecasts. Thus, half of the participants chose among a model-only, human-only, or use-freely process, whereas the other half chose among a model-only, human-only, or adjust-by-10 process.

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 5-

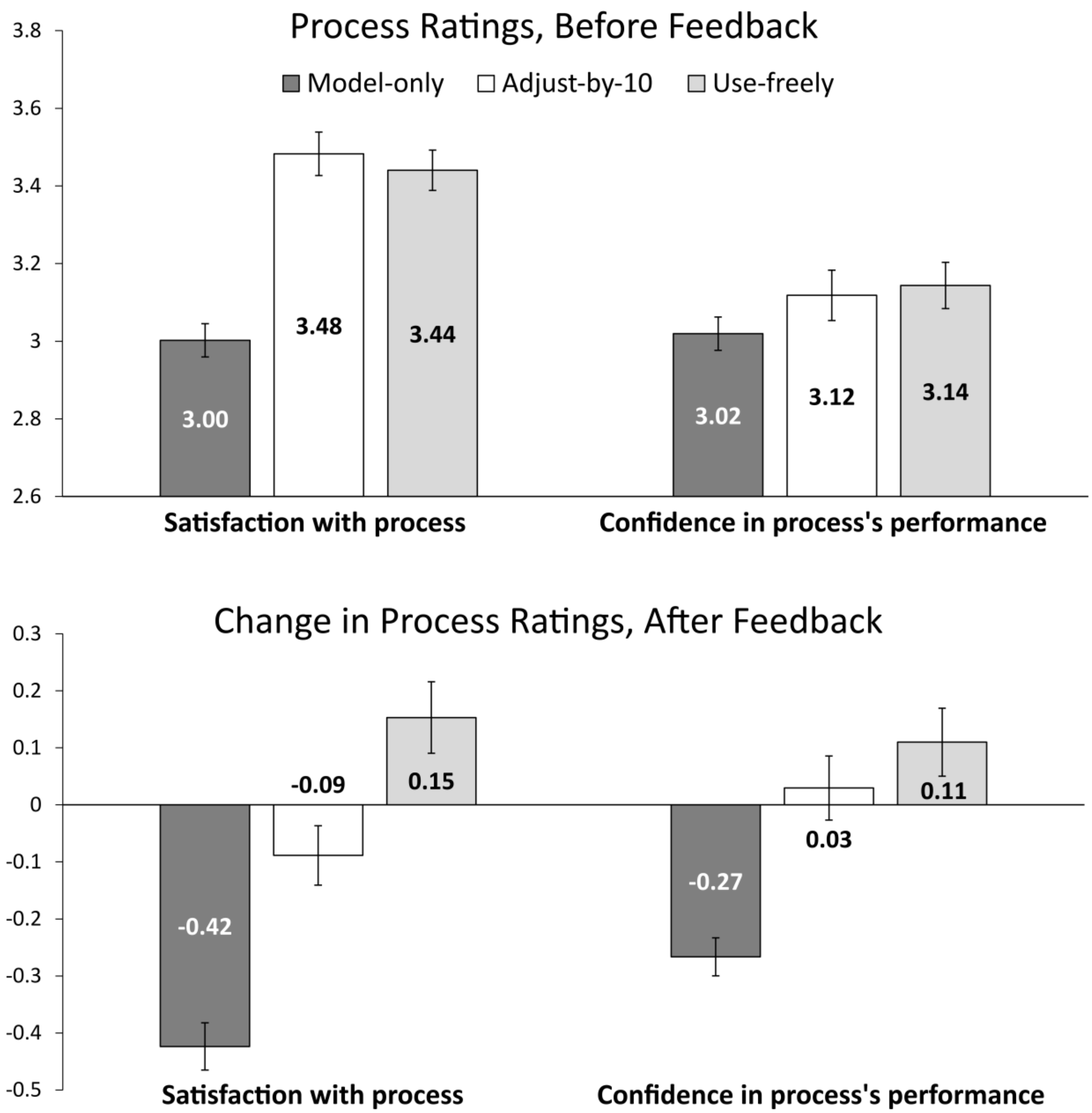
point scales (1 = none; 5 = a lot), and reported their thoughts and feelings about the statistical model. Finally, participants reported their age, gender, and highest level of education.

Results

Are people more satisfied with a forecasting process that allows them to modify a model's forecasts than with one that does not? As shown in Figure 10, participants in the adjust-by-10 and use-freely conditions, who were able to modify the model's forecasts, rated their assigned forecasting process more favorably than participants in the model-only condition, who could not modify the model's forecasts. Participants in the model-only condition were significantly less satisfied with their forecasting process than those assigned to the use-freely condition, $t(620) = -6.17$, $p < .001$, and adjust-by-10 condition, $t(614) = -6.59$, $p < .001$. Also, participants in the model-only condition were directionally less confident in the performance of their forecasting process than participants in the use-freely condition, $t(620) = -1.68$, $p = .093$, and adjust-by-10 condition, $t(614) = -1.29$, $p = .196$. Interestingly, participants in the adjust-by-10 condition were about equally satisfied with, $t(410) = 0.56$, $p = .578$, and confident in, $t(410) = -0.29$, $p = .774$, their assigned forecasting process compared to participants in the use-freely condition, even though they had less freedom to adjust the model's forecasts. Consistent with our findings in Study 3, this suggests that people want some freedom to adjust a model's forecasts but are relatively insensitive to how much freedom they have.

Figure 10

Study 4: Participants who could modify the model's forecasts were more satisfied with their forecasting process (top panel) and reacted less harshly after learning that the process had erred (bottom panel).



Note: Errors bars indicate ±1 standard error.

Are people more forgiving of forecasting errors when they were able to modify the model's forecasts than when they were not? To answer this question, we computed the change between participants' satisfaction/confidence with their Stage 1 process before vs. after receiving performance feedback. Positive values indicate that people's satisfaction/confidence increased after learning how well they performed, whereas negative values indicate that people's satisfaction/confidence decreased after learning how well they performed. As shown in Figure 10, analyses of these measures revealed that participants in the adjust-by-10 and use-freely conditions, who were able to modify the model's forecasts, were less sensitive to performance feedback than participants in the model-only condition, who could not modify the model's forecasts.[17] This arose even though those in the model-only condition performed directionally better (see Figure 14). Strikingly, though past research shows that people judge algorithms much less favorably after seeing them perform (Dietvorst et al., 2015), in this study we find that only the model-only condition became significantly less satisfied and less confident with the process after learning how well they performed. Evidently, giving people some control over the model's forecasts not only increases their satisfaction with the forecasting process; it renders that satisfaction more impervious to performance feedback.[18] (See Study S3 in the supplement for a replication of these results).

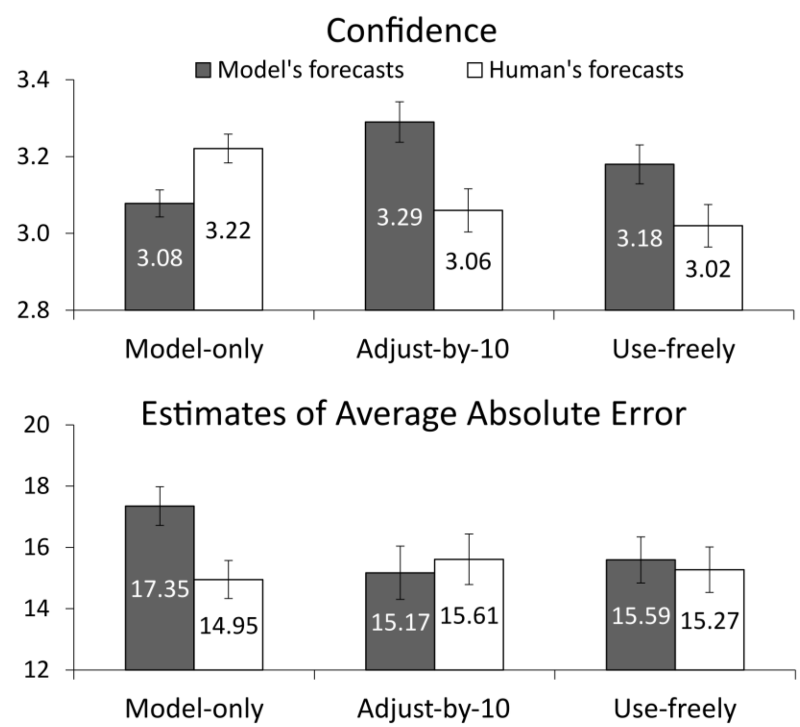
Does giving people the opportunity to modify a model's forecasts make them think the model is better? Participants' confidence ratings and performance estimates suggest that allowing people to modify an algorithm's forecasts may improve their perceptions of the algorithm relative to themselves. Participants who could not modify the model's forecasts during the first set of forecasts had less confidence in the model's forecasts than their own, $t(411) = -3.31$, $p = .001$, and thought that their average absolute error was lower than the model's, $t(412) = 5.54$, $p < .001$ (see Figure 11). In contrast, participants who could adjust the model's forecasts by up to 10 percentiles had more confidence in the model's forecasts than their own, $t(200) = 4.51$, $p < .001$, and thought that their average absolute error was similar to the model's, $t(200) = -0.67$, $p = .506$. Similarly, participants who could modify the model's

[17] Participants in the adjust-by-10 condition lost significantly less satisfaction with, t(614) = -4.82, p < .001, and confidence in, t(614) = -4.79, p < .001, their assigned forecasting process compared to participants in the model-only condition. Participants in the use-freely condition lost significantly less satisfaction with, t(620) = -7.86, p < .001, and confidence in, t(620) = -5.96, p < .001, their assigned forecasting process compared to participants in the model-only condition.
[18] All participants saw their forecasting process err. The best performing participant had an average absolute error of 9.2.

forecasts by an unlimited amount had more confidence in the model's forecasts than their own, $t(203) = 2.77$, $p = .006$, and thought that their average absolute error was similar to the model's, $t(203) = 0.51$, $p = .612$. 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, perhaps because their increased satisfaction with the process bleeds into their feelings about the model or because their forecasting experience gives them greater appreciation for the difficulty of the task.

Figure 11

Study 4: 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.

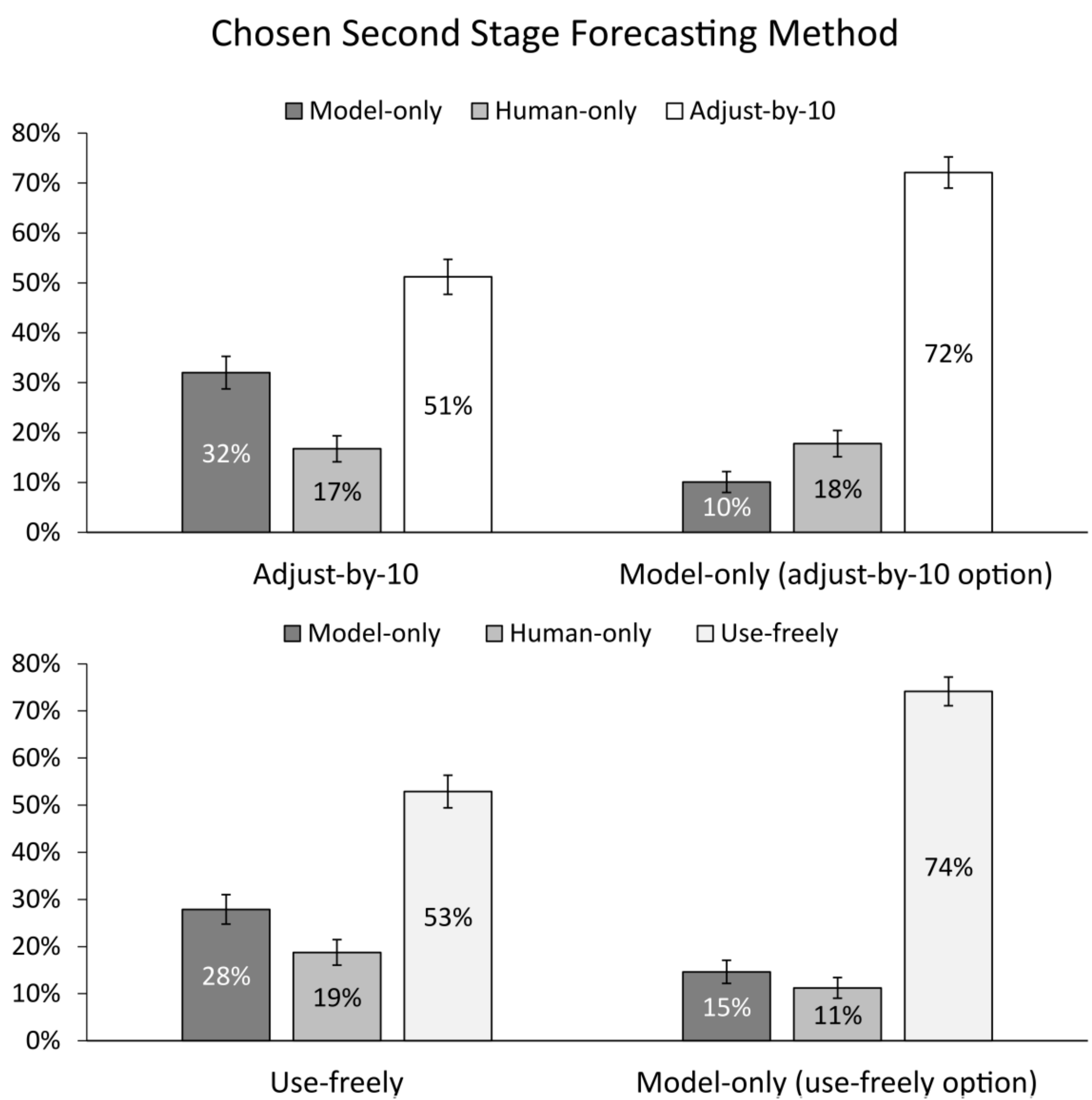


Note. Error bars indicate ±1 standard error.

Are people more likely to choose to use a model when they were previously able to modify its forecasts than when they were not? Participants in all conditions were most likely to choose the forecasting process that allowed them to modify the model's forecast (see Figure 12). However, 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 participants who could not modify the model’s forecasts in Stage 1 (12%), $\chi^2(1, N = 823) = 38.45$, $p < .001$. Thus, not only were those who were able to modify the model’s forecasts in Stage 1 more confident in the model’s ability, they were also more likely to *completely* rely on the model to make subsequent forecasts.

Figure 12

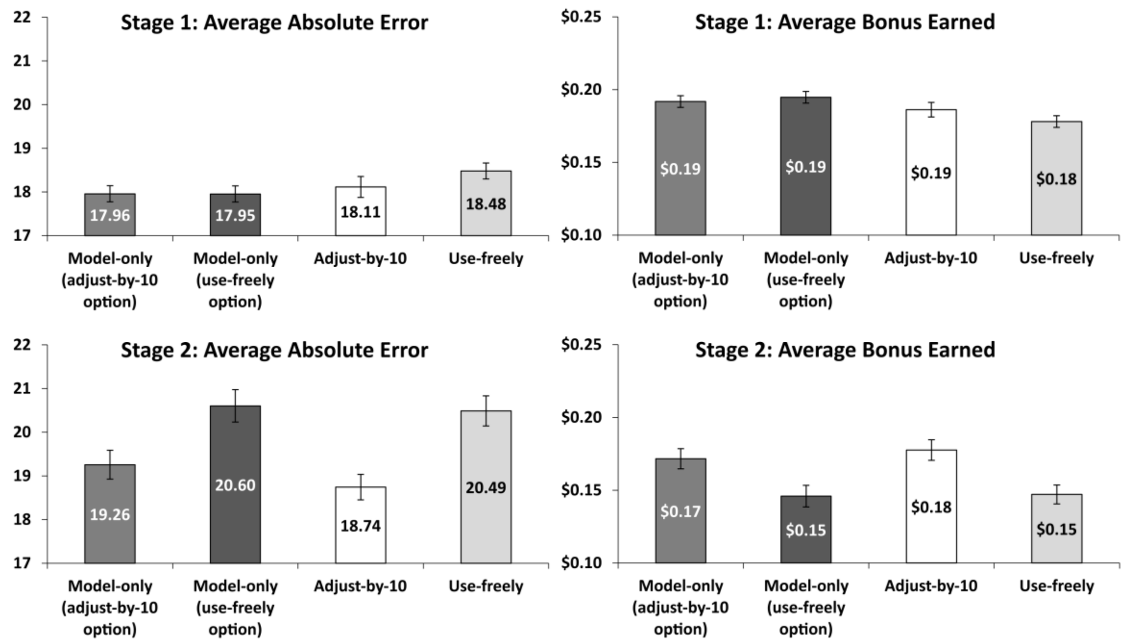
Study 4: Participants in each condition chose to modify the model’s forecasts most often, instead of only using their own forecasts or only using the model’s forecasts.



Note: Errors bars indicate ±1 standard error.

Figure 13

Study 4: 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.

Forecasting Performance. As shown in Figure 13, the more participants were required to use the model in the first forecasting stage, the better they performed. Those in the use-freely condition had significantly larger average absolute errors than participants in the model-only condition, $t(622) = -2.32$, $p = .021$, and directionally larger average absolute errors than participants in the adjust-by-10 condition, $t(410) = -1.22$, $p = .224$. As a result, participants in the use-freely condition earned significantly smaller bonuses than participants in the model-only condition, t(622) = 3.03, p = .003, and directionally smaller bonuses than participants in the adjust-by-10 condition, t(410) = 1.27, p = .207.

Much more interesting is how participants fared in Stage 2, when they could choose to either completely use the model, completely use their own forecasts, or to adjust the model's forecasts. 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), $t(510) = -6.28$, $p < .001$, lower average absolute errors than participants who chose to use their own forecasts (M = 24.50), $t(382) = 14.01$, $p < .001$, and similar average absolute errors to participants who chose to use the model's forecasts (M = 18.20), $t(424) = 0.97$, $p = .331$. Participants who used the adjust-by-10 process outperformed those who used the model freely or not at all specifically because they provided forecasts that were closer to the model's (M = 5.04) compared to participants who used the human-only (M = 14.72), $t(382) = 26.58$, $p < .001$, and use-freely (M = 7.67), $t(510) = -8.10$, $p < .001$, processes. As a result of these performance differences, participants who had the option to use the adjust-by-10 process in Stage 2 (i.e. the adjust-by-10 condition, ½ 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 ½ of the model-only condition).[19]

Discussion. Taken together, these results highlight the substantial and surprising benefits of having people modify an algorithm's forecasts. It increases their satisfaction with the process, their tolerance of errors, their confidence in the model, and their use of the model on subsequent forecasts. Moreover, restricting people's adjustments to the model, rather than allowing them to use it freely, prevents forecasts that deviate greatly from the model and thus prevents large errors.

General Discussion

Our studies show that providing people with the option to modify an algorithm's forecasts significantly increases their likelihood of using it. Moreover, people are insensitive to the amount by which they can modify the algorithm's forecasts when making this decision. Participants in our studies chose to use the algorithm substantially more often when they could modify its forecasts, and this propensity to use the algorithm persisted even when their ability to modify the algorithm was severely

[19] Participants in the adjust-by-10 condition had lower average absolute errors than, and earned more money than, participants in the use-freely condition had, $t(403) = -3.86$, $p < .001$, and earned, $t(403) = 3.59$, $p < .001$, or participants in the model-only condition who had the use-freely option had, $t(404) = 3.92$, $p < .001$, and earned, $t(404) = -3.68$, $p < .001$. Participants in the model-only condition who had the adjust-by-10 option had lower average absolute errors than, and earned more money than, participants in the use-freely condition had, $t(410) = -2.86$, $p = .01$, and earned, $t(410) = 2.99$, $p = .003$, or participants in the model-only condition who had the use-freely option had, $t(411) = -2.71$, $p = .007$, and earned, $t(411) = 3.10$, $p = .002$.

restricted. Further, allowing people to adjust an algorithm's forecasts has additional important benefits. Participants who were able to modify an algorithm's forecasts reported higher satisfaction with their forecasting process and higher tolerance for forecasting errors. Participants who adjusted the algorithm's forecasts also had relatively more confidence in the algorithm's forecasts than their own forecasts, while the opposite was true for participants who couldn't modify the algorithm's forecasts. Finally, we found that restricting the amount by which people can modify an algorithm's forecasts leads to better performance as compared to allowing people to freely modify an algorithm's forecasts.

Multiple forecasting processes that allowed participants to adjust algorithms' forecasts decreased algorithm aversion and increased performance; however, allowing participants to modify all of an algorithm's forecasts by a restricted amount was the most promising solution. Restricting the amount that participants could adjust an algorithm's forecasts kept participants' forecasts closer to the model's, and increased their performance as a result. Additionally, restricting the amount that participants could modify the algorithm's forecasts did not decrease their willingness to use the algorithm. In fact, participants in Study 4 chose to use the model's forecasts freely and adjust the model's forecasts by up to 10 percentiles at a very similar rate, even though the latter gave them less freedom to use their own judgment.

In our studies, allowing participants to adjust some of the algorithm's forecasts by an unlimited amount (i.e. the change-x process) did not consistently improve their performance. However, when people have important information that an algorithm does not have, allowing them to make large adjustments to the algorithm's forecasts may increase accuracy (see Lawrence, Goodwin, O'Connor, & Önkal, 2006). Specifically, Fildes, Goodwin, Lawrence, and Nikolopoulos, (2009) found that experts who adjusted algorithms' forecasts in supply chain companies made large and beneficial adjustments that reduced forecasting error; however, these experts also made small and unnecessary adjustments that were not beneficial. Allowing forecasters to adjust some of an algorithm's forecasts by an unlimited amount may be a useful tool in similar domains because it allows forecasters to make important adjustments to the algorithm's forecasts, but keeps them from making unnecessary changes.

These findings have many important implications. Practitioners should use algorithms to complete many types of forecasting tasks, such as forecasting demand, scheduling inventory, diagnosing patients, hiring employees, and admitting students. However, they often fail to do so. Large surveys of professional forecasters have shown that they frequently fail to use an algorithm or over-adjust an algorithm's forecasts (Fildes & Goodwin, 2007; Sanders & Manrodt, 2003). The studies in this paper suggest that having these forecasters adjust an algorithm by a restricted amount may increase their use of the algorithm and improve their forecasting performance without making them dissatisfied with their forecasting process. Thus, allowing forecasters to restrictively adjust algorithms could be a long term fix for algorithm aversion. Also, laypeople should use recommendations from algorithms to improve their decision making in many important domains, such as investment and healthcare decisions. Our studies suggest that presenting an algorithm's recommendation in a format that allows people to adjust it may increase the weight that people give to that recommendation.

Limitations and future directions.

The studies in this paper leave some questions unanswered. First, there could be conditions under which the effects we found would be diminished or eliminated. For example, participants were willing to use two imperfect algorithms in our experiments; however, these algorithms only erred by 17.5 percentiles and 4.3 ranks on average. We can't be sure that people would be willing to use an algorithm that performs significantly worse than these. 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 to 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. 1 percentile). Also, while we did use two different forecasting tasks and populations in our studies, it is possible that the effects we found are dependent on some characteristics of those tasks or populations.

Future work could investigate these limitations and expand upon the studies in this paper. First, future work could investigate the effects shown in this paper with different populations of participants, different algorithms that are more or less accurate than the two used in our studies, and different forecasting domains. Research with a population of professional forecasters would be especially informative. Second, future research could investigate how describing algorithms differently or using different types of algorithms affects people's use of those algorithms. Finally, future research could investigate the optimal way to have people adjust an algorithm's forecasts. We only tested three processes for allowing participants to adjust an algorithm's forecasts – it is inevitable that other effective methods exist.

In conclusion, we found that letting people adjust an algorithm's forecasts increases their likelihood of using the algorithm, improves their forecasting performance, heightens their tolerance of errors, and increases their confidence in the algorithm. 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 frequently worsen the algorithm's forecasts when given the ability to adjust them; however, we may have to accept this small increase in error in order to have people make less error overall.

References

Adams, I. D., Chan, M., Clifford, P. C., Cooke, W. M., Dallos, V., De Dombal, F. T., ... & McIntyre, N. (1986). Computer aided diagnosis of acute abdominal pain: a multicentre study. *British Medical Journal*, 293(6550), 800-804.

Arkes, H. R., Dawes, R. M., & Christensen, C. (1986). Factors influencing the use of a decision rule in a probabilistic task. *Organizational Behavior and Human Decision Processes*, 37, 93–110.

Beck, A. H., Sangoi, A. R., Leung, S., Marinelli, R. J., Nielsen, T. O., van de Vijver, M. J., ... & Koller, D. (2011). Systematic analysis of breast cancer morphology uncovers stromal features associated with survival. *Science translational medicine*, 3(108), 108ra113-108ra113.

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

Dawes, R. M. (1971). A case study of graduate admissions: Application of three principles of human decision making. *American psychologist*, 26(2), 180.

Dawes, R. M. (1979). The robust beauty of improper linear models in decision making. *American Psychologist*, 34(7), 571–582.

Dawes, R. M., Faust, D., & Meehl, P. E. (1989). Clinical versus actuarial judgment. *Science*, 243(4899), 1668-1674.

Diab, D. L., Pui, S. Y., Yankelevich, M., & Highhouse, S. (2011). Lay Perceptions of Selection Decision Aids in US and Non-US Samples. *International Journal of Selection and Assessment*, 19(2), 209-216.

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General*, 144(1), 114-126.

Eastwood, J., Snook, B., & Luther, K. (2012). What People Want From Their Professionals: Attitudes Toward Decision-making Strategies. *Journal of Behavioral Decision Making*, 25(5), 458-468.

Einhorn, H. J. (1986). Accepting error to make less error. *Journal of 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. *International Journal of Forecasting*, 25(1), 3-23.

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

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.

Highhouse, S. (2008). Stubborn reliance on intuition and subjectivity in employee selection. *Industrial and Organizational Psychology*, 1(3), 333-342.

Hogarth, R. M., & Makridakis, S. (1981). Forecasting and planning: An evaluation. *Management Science*, 27(2), 115-138.

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

Lim, J. S., & O'Connor, M. (1995). Judgemental adjustment of initial forecasts: its effectiveness and biases. *Journal of Behavioral Decision Making*, 8(3), 149-168.

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

Önkal, D., Goodwin, P., Thomson, M., Gönül, S., & Pollock, A. (2009). The relative influence of advice from human experts and statistical methods on forecast adjustments. *Journal of Behavioral Decision Making*, 22(4), 390-409.

Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior research methods*, 40(3), 879-891.

Promberger, M., & Baron, J. (2006). Do patients trust computers?. *Journal of Behavioral Decision Making*, 19(5), 455-468.

Sanders, N. R., & Manrodt, K. B. (2003). The efficacy of using judgmental versus quantitative forecasting methods in practice. *Omega*, 31(6), 511-522.

Schweitzer, M. E., & Cachon, G. P. (2000). Decision bias in the newsvendor problem with a known demand distribution: Experimental evidence. *Management Science*, 46(3), 404-420.

Thompson, R. E. (1952). A validation of the Glueck Social Prediction Scale for proneness to delinquency. *Journal of Criminal Law: Criminology and Police Science*, 43, 451-470.

Willemain, T. R. (1991). The effect of graphical adjustment on forecast accuracy. *International Journal of Forecasting*, 7(2), 151-154.

Wormith, J. S., & Goldstone, C. S. (1984). The clinical and statistical prediction of recidivism. *Criminal Justice and Behavior*, 11, 3-34.

Appendix A: Payment Rule for Study 1

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

Appendix B: Payment Rule for Study 2

Participants in Study 2 were paid as follows:
$0.60 - within 1 rank of state's actual rank on average
$0.50 - within 2 ranks of state's actual rank on average
$0.40 - within 3 ranks of state's actual rank on average
$0.30 - within 4 ranks of state's actual rank on average
$0.20 - within 5 ranks of state's actual rank on average
$0.10 - within 6 ranks of state's actual rank on average

Appendix C: Payment Rule for Studies 3 and 4

Participants in Studies 3 and 4 were paid as follows:
$0.50 - within 5 percentiles of student's actual percentiles on average
$0.40 - within 10 percentiles of student's actual percentiles on average
$0.30 - within 15 percentiles of student's actual percentiles on average
$0.20 - within 20 percentiles of student's actual percentiles on average
$0.10 - within 25 percentiles of student's actual percentiles on average
*Participants in study 4 were paid separately for each 2 rounds of 10 forecasts