Research Article



The Impact of Feedback on Perceptual **Decision-Making and Metacognition: Reduction in Bias but No Change in Sensitivity**



Nadia Haddara and Dobromir Rahnev

School of Psychology, Georgia Institute of Technology

Psychological Science 2022, Vol. 33(2) 259-275 © The Author(s) 2022 Article reuse guidelines: sagepub.com/journals-permissions DOI: 10.1177/09567976211032887 www.psychologicalscience.org/PS



Abstract

It is widely believed that feedback improves behavior, but the mechanisms behind this improvement remain unclear. Different theories postulate that feedback has either a direct effect on performance through automatic reinforcement mechanisms or only an indirect effect mediated by a deliberate change in strategy. To adjudicate between these competing accounts, we performed two large experiments on human adults (total N = 518); approximately half the participants received trial-by-trial feedback on a perceptual task, whereas the other half did not receive any feedback. We found that feedback had no effect on either perceptual or metacognitive sensitivity even after 7 days of training. On the other hand, feedback significantly affected participants' response strategies by reducing response bias and improving confidence calibration. These results suggest that the beneficial effects of feedback stem from allowing people to adjust their strategies for performing the task and not from direct reinforcement mechanisms, at least in the domain of perception.

Keywords

perceptual decision-making, metacognition, criterion, bias, sensitivity, confidence calibration, open data, open materials, preregistered

Received 2/18/20; Revision accepted 6/6/21

The notion that feedback improves performance has been a mainstay of experimental psychology for more than a century (Judd, 1905; Wright, 1906). Indeed, as Pritchard and colleagues (1988) put it, "The positive effect of feedback on performance has become one of the most accepted principles in psychology" (p. 338). However, despite the large amount of research on this topic, the mechanisms through which feedback affects performance are still unclear.

One long-standing view about the mechanism of feedback is that it allows for stimuli and responses to become associated via automatic reinforcement mechanisms. For example, an early version of this theory, dubbed the "law of effect" (Thorndike, 1927), proposed that feedback directly and automatically strengthens internal connections and thus leads to improvements in both animals and humans independent of any cognitive strategies adopted by the organism. Similar theories regarding automatic mechanisms of feedback remain popular to this day (Petrov et al., 2005). On the other hand, a competing view is that feedback is effective only to the extent to which it improves one's strategy for completing the task (Vollmeyer & Rheinberg, 2005). Surprisingly, these two accounts of the mechanisms of feedback have never been distinguished. In fact, adjudicating between these two theories is particularly challenging because both direct changes to observer sensitivity and

Corresponding Authors:

Nadia Haddara, Georgia Institute of Technology, School of Psychology Email: nadiahaddara@gatech.edu

Dobromir Rahnev, Georgia Institute of Technology, School of Psychology Email: rahnev@psych.gatech.edu

indirect strategy adjustment could result in identical outcomes of improved performance.

Fortunately, these two possibilities can be differentiated in the context of simple two-choice perceptual tasks. Behavior in such tasks can be fully captured by two different parameters derived from signal detection theory: perceptual sensitivity (d') and response criterion (c; Green & Swets, 1966). Perceptual sensitivity is determined by sensory processing outside of a participant's conscious control, whereas response criterion is under one's control and can be influenced by a deliberate change in strategy (Macmillan & Creelman, 2005). Therefore, by examining whether feedback affects perceptual sensitivity or response criterion, we can distinguish between a direct effect (based on automatic reinforcement) and an indirect effect (based on strategy change) of feedback on behavior.

However, previous research with perceptual tasks has been surprisingly inconsistent; some studies have reported feedback-related performance enhancements (Ball & Sekuler, 1987; Fahle & Edelman, 1993; Herzog & Fahle, 1997; Seitz et al., 2006), whereas others have shown no difference between feedback and nofeedback groups (Goldhacker et al., 2014; Petrov et al., 2006; Rouault et al., 2019; Shibata et al., 2009). Considering the small sample sizes used in most of these studies (6–36 participants split into two or more groups; except Shiu & Pashler, 1992, N = 59), the inconsistency is not surprising because such designs are likely to lead to both false positives and false negatives while also reflecting the prevailing bias rather than true effects (Ioannidis, 2005).

In addition to simple perceptual tasks, tasks that assess people's metacognition can also be used to investigate the mechanisms of feedback (Metcalf & Shimamura, 1994). Such tasks collect confidence ratings regarding the perceived accuracy of the primary decision and can provide a complementary test of whether feedback has a direct impact on sensitivity or an indirect strategy-mediated influence that can be seen only in one's confidence calibration (Rahnev & Denison, 2018). However, although a relatively sizable literature has at least examined how feedback affects perceptual sensitivity, no studies to date have examined the effects of trial-by-trial feedback on metacognitive sensitivity (although block feedback was examined by Rouault et al., 2019, and Carpenter et al., 2019, and a neurofeedback paradigm was used by Cortese et al., 2016).

Here, we investigated the impact of trial-by-trial feedback on both perceptual and metacognitive judgments in two experiments. Experiment 1 included a single day of training in a sample (N = 443) much larger than those used in previous studies. On the other hand, Experiment 2 (preregistered) employed a much longer training

Statement of Relevance

How can we help people improve their performance? It is often thought that across a variety of domains from education to work settings to sports achievements to various cognitive tasks, performance can be improved by simply providing feedback. However, it remains unclear when and why feedback helps. To address this question, we performed two large experiments in which adult participants made perceptual and metacognitive judgments. We found that trial-by-trial feedback had no effect on perceptual or metacognitive sensitivity but reduced both perceptual and metacognitive bias. These results suggest that, contrary to popular beliefs, feedback may not have direct, automatic effects on performance. Instead, the beneficial effects of feedback could be exclusively driven by the fact that it allows participants to change their strategy for performing the task. If so, feedback can be expected to be effective only if people are willing and able to change the way they perform a specific task.

period (7 days, N = 75). In both experiments, each participant was randomly assigned to a feedback or nofeedback group. We found that feedback had no effect on perceptual or metacognitive sensitivity, but it reduced both perceptual and metacognitive biases, suggesting that it affected performance via strategy adjustments rather than direct reinforcement. These findings reveal the mechanisms of feedback in perceptual decisionmaking and metacognition and have potential implications for the use of feedback in applied settings.

Method

Participants

Participants for both experiments were adults recruited on Amazon's Mechanical Turk and were compensated 7.25 per hour. For Experiment 1, we chose a sample size (N = 443) much larger than those used in previous studies examining the role of feedback in perceptual decision-making to ensure sufficient power for our analyses. For Experiment 2, we preregistered a sample size of 60 after exclusions. The preregistration can be accessed at https://osf.io/efymg/ and was completed after data from Experiment 1 were analyzed. A total of 75 new participants completed all 7 days of Experiment 2 (participants from Experiment 1); 15 of them were excluded on the basis of predetermined criteria. Participants in Experiment 2 received a bonus for completing all 7 days of the experiment; they gained \$0.01 for every three correct trials (average bonus = \$8.70). Recruiting Mechanical Turk research participants is considered a form of convenience sampling. All participants reported normal or corrected-to-normal vision and provided written informed consent. Experimental procedures were approved by the Georgia Institute of Technology Institutional Review Board.

Procedure

In Experiment 1, each participant was randomly assigned to either a feedback group or no-feedback group and completed two separate perceptual tasks adapted from the study by Rahnev et al. (2015). In Task 1, participants made a perceptual judgment by pressing a key to indicate whether the letter X or O occurred more frequently in a 7×7 grid presented for 500 ms. Participants were given an untimed response period for the perceptual decision. After indicating their response, participants were prompted to rate their confidence in the accuracy of their decision on a 4-point scale (1 = low confidence), $4 = high \ confidence$; untimed response). After the confidence rating, the feedback group received feedback concerning the accuracy of their perceptual judgment ("correct" or "wrong"; 500 ms), whereas the no-feedback group just saw a fixation cross for 500 ms, thus keeping the time between the confidence response and the onset of the next trial the same between the two groups (Fig. 1). Participants completed 11 blocks of 30 trials each (330 total trials) and were given the opportunity to take a break at the end of each block. The first trial of each block began with a longer 2-s fixation to give participants time to focus their attention, and each subsequent trial began with a shorter 500-ms fixation.

Task 2 was designed to be very similar to Task 1; the only difference was the perceptual dimension being discriminated. Unlike Task 1, in which participants discriminated between Xs and Os, Task 2 featured a discrimination between red and blue circles and consisted of five blocks of 30 trials (150 total trials with opportunities for breaks after each block). Critically, in Task 2, no participant received performance feedback. This was done to determine whether any putative feedbackrelated improvements in Task 1 would generalize to a new task.

The dominant stimulus class (X or O in Task 1, red or blue in Task 2) was randomly determined on each trial, and the proportion of the dominant stimulus within the 7 \times 7 grid was fixed at 30/49 for Task 1 and 27/49 for Task 2. The proportions were different for the two

tasks to ensure that the two tasks had similar difficulty levels because our pilot data suggested that the red/ blue discrimination was easier for participants. The average accuracy was 80% for Task 1 and 78% for Task 2.

Experiment 2 employed Task 1 from Experiment 1 (Fig. 1), but participants performed the task over 7 days. Each day of the experiment included 20 blocks of 25 trials each (for a total of 500 trials per day and 3,500 in total). The first six blocks of Day 1 were used to adjust the task difficulty for each participant. We ran a standard two-up, one-down staircase procedure that started with 30 (of 49) characters from the dominant stimulus. The final difficulty for each participant was determined as the average of the number of dominant items at the time of staircase reversals (ignoring the first four reversals; Bang et al., 2018, 2019), rounded to the nearest integer. The resulting difficulty value was then fixed for the remainder of the experiment (14 blocks on Day 1 and all trials in subsequent days). After completing Day 1, participants who passed quality checks (see the Analyses section) were invited to complete the remainder of the experiment. The average accuracy across all days of the experiment was 75%.

The experiments were designed using the *jsPsych* library (Version 5.0.3; de Leeuw, 2015), and grid stimuli were created using in-house JavaScript code. To ensure that the stimulus size was similar for participants who completed the experiment on different screens, we used a procedure previously established in our lab (Bang et al., 2019) in which participants were asked to adjust the stimuli on the screen to match the size of real-life objects (a credit card or a quarter).

Sensitivity and bias measures

We computed the signal detection theory (Green & Swets, 1966; Macmillan & Creelman, 2005) parameters stimulus sensitivity (d') and response criterion (c) to determine participants' performance and degree of response bias on the tasks. These measures were calculated on the basis of the observed hit rate and falsealarm rate as follows:

and

$$c = -\frac{1}{2}(\phi^{-1}(\text{hit rate}) + \phi^{-1}(\text{false-alarm rate})),$$

 $d' = \phi^{-1}$ (hit rate) – ϕ^{-1} (false-alarm rate)

where
$$\phi^{-1}$$
 is the inverse of the cumulative standard normal distribution that transforms hit rate and false-alarm rate into *z* scores. Hit rate and false-alarm rate were

ma n rat defined by treating the letter X in Task 1 and blue circles in Task 2 as the target. Therefore, negative c values



Fig. 1. Experimental tasks. Experiment 1 consisted of two tasks conducted in a single session. For both tasks, participants were required to indicate which of two stimulus classes occurred more frequently in a 7×7 grid (Task 1: dominant shape [O or X], Task 2: dominant color [red or blue]). After providing their response, participants were prompted to give confidence ratings on a scale from 1 to 4. In Task 1, approximately half of the participants received trial-by-trial feedback (feedback group), whereas the rest of the participants received no feedback at all (no-feedback group). In Task 2, neither group received feedback; this second task was used to investigate which effects of feedback would generalize to a new task. Experiment 2 consisted of only Task 1 but was conducted over 7 days.

indicate a bias for the letter X (Task 1) or the color blue (Task 2), whereas positive *c* values indicate a bias for the letter O (Task 1) or the color red (Task 2). The appropriateness of applying signal detection theory to the data here was ascertained by examining individual and group-level receiver operating characteristic (ROC) curves (see Fig. S1 in the Supplemental Material available online).

We further computed metacognitive sensitivity, which is a measure of one's ability to distinguish between one's own correct and incorrect judgments (Fleming & Lau, 2014). Similar to perceptual sensitivity, which reflects the strength of the relationship between one's choices and the stimulus, metacognitive sensitivity reflects the strength of the relationship between one's confidence ratings and the accuracy of the perceptual decisions. Higher metacognitive sensitivity implies that one's confidence ratings are more informative regarding the accuracy of one's judgments. We quantified metacognitive sensitivity using the measure meta-*d*' developed by Maniscalco and Lau (2012), which was derived on the basis of a signal detection model and is expressed

in the same units as the measure d'. Finally, we computed metacognitive bias as one's tendency to have confidence ratings that are too low or too high relative to one's level of perceptual sensitivity (Fleming & Lau, 2014). Metacognitive bias was calculated as the acrossparticipants Pearson's correlation between d' and average confidence for each participant. This measure was used to indicate the degree of confidence bias across participants; higher correlation coefficients are indicative of better confidence calibration.

Analyses

All statistical analyses for both experiments were conducted in MATLAB (The MathWorks, Natick, MA), and figures were generated in MATLAB and the R software environment (Version 4.0.2; R Core Team, 2020).

Experiment 1. Participants who performed at less than 55% correct or greater than 95% correct were excluded from the analyses. These exclusions were made separately for each of the two perceptual tasks; 48 participants were excluded from Task 1, and 47 were excluded from Task 2 (~11% exclusion rate in each task). Additionally, trials with reaction times (RTs) that were either too fast (< 200 ms) or too slow (> 2,000 ms) were excluded from the analyses (12.7% of individual trials were excluded from Task 1 and 12.5% were excluded from Task 2).

To determine the effect of feedback on perceptual sensitivity, we performed independent-samples t tests comparing d' for the feedback and no-feedback groups. Response bias was assessed by comparing the acrossparticipants variability of the criterion c for each of the two groups using a two-tailed F test for equality of variances. (Note that because the equality-of-variances F test is two tailed whereas a standard analysis-ofvariance [ANOVA] F test is one tailed, the p value associated with an F statistic from an equality-of-variances test with F > 1 is 2 times higher than the p value for the same F statistic obtained from an ANOVA.) Because strong bias results in very large positive or negative values of the criterion c, a distribution with low variability in the criterion values indicates the presence of smaller bias in the group, whereas a distribution with high variability in the criterion values indicates the presence of larger bias.

To determine the effect of feedback on metacognitive sensitivity, we compared the feedback and nofeedback groups using independent-samples t tests to assess differences in meta-d' (i.e., metacognitive sensitivity). On the other hand, to compare the size of metacognitive bias between the two groups, we computed the d'-confidence correlation coefficient for each group, and the two values were compared using a standard z test for comparing two independent r values as follows:

$$z = \frac{z_1 - z_2}{\sqrt{\frac{1}{N_1 - 3} + \frac{1}{N_2 - 3}}},$$

where $z_i = \frac{1}{2} \log \left(\frac{1 + r_i}{1 - r_i} \right)$ is the *z*-transformed correlation

coefficient (*r*) and N_i is the sample size for the corresponding group (i = 1, 2).

We further investigated whether there were any feedback-related changes in decision and confidence RTs by conducting independent-samples t tests. In the analysis of confidence RTs, we additionally excluded trials with confidence RTs outside the range of 200 ms to 2,000 ms. In a separate analysis, we checked whether the difference in decision and confidence RTs for the feedback and no-feedback groups changed from Task 1 to Task 2; for that analysis, we included participants with greater than 55% and less than 95% accuracy for both tasks and performed an independent-samples t test.

One limitation of our main signal detection theory analyses is that they do not take RT into account. Therefore, it is possible that the feedback manipulation affected perceptual sensitivity but also resulted in emphasizing speed over accuracy (via a speed/accuracy trade-off), thus resulting in no d' difference between the two feedback groups. To check for such a possibility, we fitted the drift-diffusion model (Ratcliff, 1978) to the response and RT data and computed the parameters drift rate (v), boundary separation (a), and nondecision time (Ter) separately for the feedback and no-feedback groups. Note that the drift rate (v) reflects the perceptual sensitivity, the boundary (a) reflects the speed/accuracy trade-off, and the nondecision time (Ter) reflects the sensory and motor delays involved in the decision process. Therefore, comparing these three parameters between the two feedback groups allowed us to determine whether perceptual sensitivity was truly matched even when controlling for speed/accuracy trade-off. We used a simple version of the drift-diffusion model without variability parameters (Wagenmakers et al., 2007) in which the starting point of the accumulation (z) is kept constant at a/2. This version has been previously found to better recover parameter changes in the drift-diffusion model (van Ravenzwaaij et al., 2017) and provides good fits to the data here (see Fig. S2 in the Supplemental Material).

We also computed *d'* and meta-*d'* in a hierarchical estimation framework that is more robust to small sample sizes and edge effects. We employed the *HMeta-d* toolbox (Fleming, 2017), which uses Markov chain

Monte Carlo sampling to estimate posterior distribution over model parameters. To assess the existence of group differences in d' and meta-d', we calculated the 95% highest-density intervals (HDIs) of the posteriors.

Finally, in addition to using frequentist statistics, we performed Bayesian analyses using JASP (Version 0.9.2; JASP Team, 2020) and MATLAB. When comparing groups, we report either the Bayes factor (BF) that denotes support for the null over the alternative hypothesis (BF₀₁) or the BF that denotes support for the alternative over the null hypothesis (BF₁₀); higher values indicate stronger evidence in both cases (Masson, 2011). For ANOVAs, we report BF_{inclusion}, which indicates whether the observed data are more probable under models that include a particular factor. For all Bayesian analyses, the default priors in JASP were used.

Experiment 2. All analyses for Experiment 2 were conducted in accordance with our preregistration and largely followed the analysis steps from Experiment 1. We first excluded participants and trials on the basis of our preregistration criteria. We removed the data from the first six blocks of Day 1 for all participants because these blocks contained the trials from our staircase procedure. After completion of Day 1, data quality for each participant was assessed using the remaining data from that day, and we excluded participants from participating in Days 2 to 7 if they had (a) accuracy lower than 60% correct, (b) accuracy higher than 85% correct, or (c) more than 10% of decision RTs that were less than 200 ms or more than 2,000 ms. A total of 75 participants successfully completed all 7 days of the experiment, and 15 participants were excluded from analyses on the basis of the predetermined criteria, which included (a) performing at less than 55% correct across all 7 days or (b) having more than 15% of decision RTs less than 200 ms or more than 2,000 ms. Lastly, as in Experiment 1, individual trials with decision RTs considered too short (< 200 ms) or too long (> 2,000 ms) were excluded from the analyses (4.9% of individual trials).

To assess whether the feedback group had increased perceptual and metacognitive sensitivity compared with the no-feedback group over the 7 days of the experiment, we conducted independent-samples t tests on the average d' and meta-d' across all seven sessions. Additionally, we performed a linear regression on task accuracy (percentage correct) as a function of block over the time course of the experiment for each participant, followed by an independent-samples t test comparing the slope of the regression between the feedback and no-feedback groups. We also conducted mixed ANOVAs with a between-subjects factor of group (feedback and no feedback) and a within-subjects factor of session (1-7) on d' and meta-d'. Similar analyses on decision and confidence RTs are reported in Supplementary

Results and in Fig. S3 (both in the Supplemental Material). Lastly, Bayesian analyses were implemented to determine the strength of the evidence in support of the null hypothesis for all relevant nonsignificant effects.

Data and code

All data and codes for the analyses are freely available online at https://osf.io/94r87/. In addition, the complete data sets for both experiments have been uploaded to the Confidence Database (Rahnev et al., 2020).

Results

Experiment 1

To uncover whether feedback acts via automatic reinforcement mechanisms or deliberate changes in strategy, we examined the effect of trial-by-trial feedback on participants' sensitivity and bias in both perceptual and metacognitive judgments. Participants completed a perceptual task that required them to indicate whether more Xs or Os were presented in a 7×7 grid (Fig. 1). Each participant was randomly assigned to a feedback or no-feedback group; the former group received trialby-trial feedback, and the latter did not receive any feedback. To investigate how the effects of feedback generalize to a novel task, we had all participants complete a second perceptual task where no feedback was given, in which they judged whether more red or blue circles appeared in a 7×7 grid (Fig. 1). We first examined the effects of feedback in Task 1 and later explored which of these effects generalized to Task 2.

Trial-by-trial feedback has no effect on perceptual or metacognitive sensitivity. We first sought to determine whether trial-by-trial feedback can affect behavior via automatic reinforcement mechanisms, which would be manifested in feedback increasing perceptual sensitivity. To do so, we compared d' values for the feedback and no-feedback groups. We found that perceptual sensitivity computed over all trials was virtually identical across the two groups (feedback: d' = 1.79, no feedback: d' = 1.80, t(392) = -0.16, p = .87, Cohen's d = -0.02 (Fig. 2a). We further investigated whether this null effect was due to a lack of power or whether our data can provide positive evidence for a lack of difference between the two groups. We therefore conducted a Bayesian independentsamples t test, which showed that the data strongly supported the null hypothesis of no difference between the two groups (BF₀₁ = 8.9). We also replicated the lack of d'difference in the two groups using a hierarchical estimation



Fig. 2. Effects of feedback on perceptual and metacognitive sensitivity in Task 1, Experiment 1. Density distributions and box plots for each of the two feedback groups are shown separately for (a) perceptual sensitivity (d') and (b) metacognitive sensitivity (meta-d'). Box plots show the median (vertical line) and the interquartile (25%–75%) range, and the whiskers indicate the 2% to 98% range. Dots indicate individual participant data. Raincloud plots were adapted from the study by Allen et al. (2019).

framework (feedback: *d*′ = 1.78, no feedback: *d*′ = 1.72; 95% HDI = [-0.33, 1.03]).

Despite the absence of overall differences in perceptual sensitivity between the feedback and no-feedback groups, it could be argued that trial-by-trial feedback may still have had a beneficial effect on perceptual accuracy. Specifically, it is possible that the effects of trial-by-trial feedback take time to manifest, and therefore, analyzing all 330 task trials together may mask the beneficial effects of feedback on perceptual sensitivity that may emerge only later in the task. To investigate this possibility, we considered the data only from the second half of the task and again found no difference between the two groups (feedback: d' = 1.88, no feedback: *d*′ = 1.84), *t*(388) = 0.59, *p* = .56, Cohen's *d* = 0.06, $BF_{01} = 7.6$. Similar results were further obtained if d' for the last n blocks of 30 trials were analyzed for nbetween 1 and 10 (all ps > .72; all $BF_{01}s > 8.4$). We also considered the rate of learning by performing a linear regression using trial number to predict task accuracy (percentage correct) and found that the rate of learning correlated with both perceptual and metacognitive sensitivity (see Fig. S4 in the Supplemental Material). Critically, there was significant learning, as indicated by a positive regression slope for both the feedback group $(\text{slope} = 1.2 \times 10^{-3}), t(199) = 6.65, p = 2.8 \times 10^{-10}, \text{Cohen's}$ d = 0.47, and no-feedback group (slope = 1.1×10^{-3}), $t(193) = 4.49, p = 1.2 \times 10^{-5}$, Cohen's d = 0.33; there was no difference between the two groups, t(392) = 1.45, p = .15, Cohen's d = 0.15, BF₀₁ = 3.3 (Fig. 3).

Finally, it is possible that a gain in sensitivity for the feedback group was obscured by faster responding driven by an altered speed/accuracy trade-off. To check for this possibility, we fitted the drift-diffusion model to the data and found no difference between the feedback and no-feedback groups in drift rate v (feedback: v = .110, no feedback: v = .106), t(392) = 0.89, p = .37, Cohen's d = 0.09, BF₀₁ = 6.1. Overall, our results show that trial-by-trial feedback had virtually no effect on perceptual sensitivity regardless of how sensitivity was assessed, thus suggesting that feedback does not affect choice behavior via direct reinforcement mechanisms.

However, even if feedback does not affect the perceptual decisions via automatic reinforcement mechanisms, it is possible that it affects higher level metacognitive judgments via such mechanisms. Such an effect would be manifested in trial-by-trial feedback improving participants' metacognitive sensitivity. To investigate this possibility, we computed the measure of metacognitive sensitivity meta-d' (Maniscalco & Lau, 2012) and compared it between the feedback and nofeedback groups. We again found no effect of feedback (feedback: meta-d' = 1.23, no feedback: meta-d' = 1.33), t(392) = -1.29, p = .20, Cohen's d = -0.13, BF₀₁ = 4.0 (Fig. 2b). Similar effects were again obtained for the second half of the task (feedback: meta-d' = 1.21, no feedback: meta-d' = 1.30, t(388) = -1.12, p = .26, Cohen's d = -0.11, BF₀₁ = 4.9, as well as when using a hierarchical estimation framework to estimate metacognitive efficiency (M-ratio = meta-d'/d'; feedback:



Fig. 3. Effects of feedback on learning rate in Task 1, Experiment 1. The mean proportion of correct responses in each feedback group is shown as a function of trial number. Shaded areas indicate standard errors of the mean. The plotted lines are smoothed with eight-trial moving average windows for display purposes; all statistics were computed using unsmoothed data.

M-ratio = .73, no feedback: M-ratio = .75; 95% HDI = [-0.09, 0.06]). Taken together, these findings demonstrate that trial-by-trial feedback did not affect either perceptual or metacognitive sensitivity in our task and therefore did not act via direct reinforcement mechanisms at the level of either the perceptual or metacognitive judgments.

Trial-by-trial feedback reduces bias in perceptual and metacognitive judgments. The results so far demonstrate that feedback does not act via automatic reinforcement mechanisms but do not elucidate whether feedback affects participants' strategies. To examine how feedback affects participants' strategic behavior, we analyzed how the presence of trial-by-trial feedback influenced bias in perceptual and metacognitive judgments.

We quantified the bias in the perceptual task using the signal detection theory measure *c*, which indicates the location of the response criterion. Large negative values of *c* indicate a strong bias toward the letter X, large positive values indicate a strong bias toward the letter O, and values close to zero indicate a lack of bias. Both the feedback and no-feedback groups had a mean value of *c* close to zero and were not different from each other (feedback: *c* = -.03, no feedback: *c* = -.02), *t*(392) = -0.50, *p* = .62, Cohen's *d* = -0.05, BF₀₁ = 8.0, indicating that neither group as a whole had a bias toward a particular stimulus category (neither the Os nor the Xs were preferred in the whole sample of participants).

Critically, we tested for a reduction in bias due to feedback, which should manifest as criterion values in



Fig. 4. Effects of feedback on perceptual bias in Task 1, Experiment 1. Density distributions and box plots for each of the two feedback groups are shown for response criterion *c*. Box plots show the median (vertical line) and the interquartile (25%–75%) range, and the whiskers indicate the 2% to 98% range. Dots indicate individual participant data.

the feedback group becoming less extreme and therefore having smaller variance. This is exactly what we found: The feedback group (SD = .21) had smaller variability of criterion scores than the no-feedback group (SD = .25), and the difference was statistically significant, F(199, 193) = 0.74, p = .035, F test of equality of variances (Fig. 4). An alternative way to test for a decrease in bias is to examine whether the absolute value of the criterion c decreased with feedback. However, the absolute values are not normally distributed (given that they are bound by zero), which necessitates the use of the less powerful Wilcoxon signed-rank test. Consequently, we found that although the absolute value of the bias was smaller for the feedback group (0.165) than for the no-feedback group (0.194), the difference was only marginally significant (z = -1.85, p = .065). Control analyses in which bias was analyzed via the variance of the distribution of criterion values in mini blocks of 10 trials suggested that this reduction in bias emerged very quickly and could already be detected by the end of the first block of 30 trials (see Fig. S5 in the Supplemental Material). These results are consistent with the notion that trial-by-trial feedback allowed participants to adjust their response strategy, thus reducing bias.

Given that trial-by-trial feedback induced a strategic change in bias in the perceptual judgment, we investigated whether a similar strategic change would also occur for the metacognitive judgment. Conceptually, metacognitive bias is the propensity of the overall



Fig. 5. Effects of feedback on metacognitive bias in Task 1, Experiment 1. The scatterplots show the correlation between d' and average confidence for the (a) no-feedback group and (b) feedback group. Each point represents a single participant; the dark gray line indicates the line of best fit. The bar graph (c) shows the correlation coefficients (r values) in each of the two feedback groups, together with the p value from a statistical test that compares their magnitude. Error bars depict standard errors of the mean.

confidence ratings of a given participant to deviate from the overall sensitivity of that same participant (Fleming & Lau, 2014). For example, participants with a high metacognitive bias may have either low perceptual sensitivity but high average confidence or high perceptual sensitivity but low average confidence. Conversely, participants with low metacognitive bias would have both high perceptual sensitivity and high average confidence or low perceptual sensitivity and low average confidence. Note that because confidence was collected on a scale ranging from 1 to 4, in which the different options on the confidence scale were not directly associated with different accuracy levels, it is impossible to quantify an individual participant's metacognitive bias in isolation. However, it is possible to determine whether the feedback and no-feedback groups differ in how strongly average perceptual sensitivity and average confidence are related across participants; a weaker relationship would indicate the presence of larger metacognitive biases (note that this measure captures relative bias across participants but not other possible biases such as overconfidence or underconfidence within each group).

We therefore investigated the presence of metacognitive bias in the feedback and no-feedback groups by correlating the perceptual sensitivity (d') and the average confidence across participants for each group. We found no significant correlation between d' and average confidence in the no-feedback group (r = .006, p = .94; Fig. 5a) but a significant positive correlation in the



Fig. 6. Effects of feedback on both decision and confidence reaction times (RTs) in Task 1, Experiment 1. Density distributions and box plots for each of the two feedback groups are shown separately for (a) decision RTs and (b) confidence RTs. Box plots show the median (vertical line) and the interquartile (25%–75%) range, and the whiskers indicate the 2% to 98% range. Dots indicate individual participant data.

feedback group (r = .27, p = .0001; Fig. 5b). Critically, the difference between the r values of the two feedback groups was statistically significant (p = .008, z test for comparing r values; Fig. 5c). These results are again consistent with the notion that participants in the feedback group were able to strategically reduce their metacognitive biases, which led to a stronger across-participants association between sensitivity and confidence in that group.

Trial-by-trial feedback decreases RTs. Beyond examining the effect of trial-by-trial feedback on sensitivity and bias in perceptual and metacognitive judgments, we also explored the effect of feedback on RTs. We first verified that decision RTs were faster for trials with correct responses than trials with errors, t(393) = 19.6, $p = 3.5 \times 10^{-60}$, Cohen's d = 0.98, and for high compared with low confidence ratings, t(393) = -21.8, $p = 1.5 \times 10^{-69}$, Cohen's d =-1.097, thus confirming that we could replicate known effects from the literature. Critically, we found that participants in the feedback group exhibited faster decision RTs (mean RT = 607 ms) than participants in the no-feedback group (mean RT = 699 ms), which is a highly significant difference, t(392) = -5.57, $p = 4.7 \times 10^{-8}$, Cohen's d =-0.56, BF₁₀ = 2.2×10^5 (Fig. 6a). Additional analyses with the drift-diffusion model (which found no difference in drift rate between the groups; see above) suggested that this effect was driven primarily by a feedback-induced decrease in nondecision time Ter (feedback: Ter = .23, no feedback: Ter = .30), t(392) = -5.05, $p = 6.7 \times 10^{-7}$, Cohen's d = -0.51, BF₁₀ = 1.8×10^4 , although there was also a small effect on boundary separation *a* (feedback: *a* = .133, no feedback: *a* = .137), *t*(392) = -2.81, *p* = .005, Cohen's *d* = -0.28, BF₁₀ = 4.9. Moreover, we found a similar effect for the RTs for the confidence response; specifically, the feedback group had lower confidence RTs (mean confidence RT = 443 ms) than the no-feedback group (mean confidence RT = 501 ms), *t*(392) = -4.54, *p* = 7.4 × 10⁻⁶, Cohen's *d* = -0.46, BF₁₀ = 1.9×10^3 (Fig. 6b), demonstrating that the feedback-induced response speedup was present for both the perceptual and confidence judgments.

Generalization of the effect of feedback to a second task. Beyond establishing the effects of trial-by-trial feedback on different variables of interest, we wanted to explore whether any of these effects would generalize to a new task. For this reason, we included a second task that was designed to be very similar to the first one while tapping into a different perceptual dimension. We therefore chose a task that requires participants to discriminate color rather than shape. Specifically, participants indicated whether more circles were red or blue (other aspects of the task were identical to the first; Fig. 1). None of the participants received feedback in Task 2. All analyses tested whether the effects of feedback in Task 1 extended to Task 2 in the absence of task-specific feedback.

Given that trial-by-trial feedback had no effect on perceptual or metacognitive sensitivity in Task 1, the

feedback and no-feedback groups from Task 1 also predictably did not differ significantly in Task 2 in either perceptual sensitivity (feedback: d' = 1.74, no feedback: d' = 1.67, t(393) = 1.31, p = .19, Cohen's d = 0.13, $BF_{01} = 4.0$, or metacognitive sensitivity (feedback: metad' = 1.15, no feedback: meta-d' = 1.20), t(393) = -0.57, p = .57, Cohen's d = -0.06, BF₀₁ = 7.7 (Figs. 7a and 7b). Unlike in Task 1, we did not observe significant learning over the course of the 150 trials in Task 2, potentially because of the smaller number of trials in that task. Interestingly, the accuracy in Task 2 was slightly higher in the initial trials for the feedback group, but this effect was not statistically significant (see Fig. S6 in the Supplemental Material). Thus, trial-by-trial feedback in Task 1 had no effect on perceptual or metacognitive sensitivity in either Task 1 or Task 2.

More importantly, the presence of trial-by-trial feedback in Task 1 also had no effect on bias in the perceptual and metacognitive judgments in Task 2. Specifically, we found no difference in the variability of response criterion c (feedback: SD = .37, no feedback: SD = .36), F(200, 193) = 1.03, p = .81, F test of equality of variances (Fig. 7c), or in the absolute value of the criterion c (z = .53, p = .6) between the two feedback groups in Task 2. Similarly, the d'-confidence correlation that was modulated by feedback in Task 1 did not differ between the two groups in Task 2 (feedback: r = .002, p = .98; no feedback: r = -.08, p = .30; difference between the two r values: p = .45; Fig. 7d). Therefore, trial-by-trial feedback appears to allow participants to reduce their bias in the specific task in which such feedback is provided but does not allow for generalization even to a very similar task as long as the new task employs a different perceptual dimension.

Despite the lack of generalization of the effects of trial-by-trial feedback on bias, we found such generalization for RT. Indeed, participants who received feedback in Task 1 were significantly faster on Task 2 even though participants in Task 2 did not receive feedback (feedback: RT = 610 ms, no feedback: RT = 657 ms), t(393) = -2.69, p = .007, Cohen's d = -0.27, BF₁₀ = 3.6 (Fig. 7e). However, it is worth noting that the decision-RT difference between the two groups was 92 ms in Task 1 and decreased by almost half to 47 ms in Task 2, and the difference in the RT effect in the two tasks was significant, t(376) = -4.93, $p = 1.2 \times 10^{-6}$, Cohen's d = -0.51. The effects of feedback on decision RTs in Task 2 also extended to confidence RTs (feedback: RT = 392 ms, no feedback: RT = 434 ms, t(393) = -3.49, p =.0005, Cohen's d = -0.35, BF₁₀ = 37.2 (Fig. 7f). Thus, receiving trial-by-trial feedback allows participants to respond faster not only on the task in which feedback is given but also on a new task in which the perceptual dimension changes.

Experiment 2

Experiment 1 demonstrated that trial-by-trial feedback provided over the course of 330 trials did not improve either perceptual or metacognitive sensitivity. Nevertheless, it is possible that the effects of feedback take longer to manifest and possibly require exposure over multiple days. Therefore, in Experiment 2, we examined the effects of feedback over the course of 7 days of training with 500 trials per day. As noted in our preregistration, because Experiment 2 included only 60 participants (after exclusions), it did not have sufficient power to examine bias effects, and our goal was solely to test whether feedback affected perceptual and metacognitive sensitivity. Nevertheless, for completeness, we examine criterion variability and RT in Supplementary Results and also explore whether initial performance during the thresholding was related to the rate of improvement (see Fig. S7 in the Supplemental Material).

To determine whether feedback enhanced perceptual sensitivity over the course of training, we conducted a mixed ANOVA with a between-subjects factor of group (feedback and no feedback) and a withinsubjects factor of day (1-7). We found no significant difference in d' between the feedback and no-feedback groups, F(1, 58) = 0.43, p = .51, $\eta_p^2 = .007$, BF₀₁ = 1.9 (Fig. 8a). Critically, the feedback group exhibited slightly lower d' values, and a directed, one-sided Bayesian independent-samples t test that specifically tested whether trial-by-trial feedback increased d' found substantial evidence for the null hypothesis ($BF_{01} = 5.8$). Further, perceptual sensitivity d' increased over the course of the 7 days of training-main effect of day: $F(6, 348) = 8.31, p < .0001, \eta_p^2 = .13, BF_{10} = 4.3 \times 10^5$ -but this increase was similar for both groups-interaction between group and day: F(6, 348) = 1.50, p =.18, $\eta_p^2 = .03$, BF_{inclusion} = .21. Examining the rate of increase over the course of training again showed that there was no difference between the two groups (feedback: slope = .026, no feedback: slope = .044), t(58) =-1.14, p = .26, Cohen's d = 0.30, BF₀₁ = 2.2, and the slope for the no-feedback group was numerically higher. Conducting a directed, one-sided test confirmed that there is substantial evidence against the hypothesis that feedback increased the rate of learning ($BF_{01} = 8.5$). Together, these findings indicate that trial-by-trial feedback, even when provided over the course of multiple days and a total of 3,500 trials, does not improve sensitivity on a perceptual task. It should be noted that the



Fig. 7. (continued on next page)

Fig. 7. Effects of not providing feedback in Task 2, Experiment 1. Density distributions and box plots for each of the two feedback groups are shown separately for (a) perceptual sensitivity (d'), (b) metacognitive sensitivity (meta-d'), and (c) response bias (variability of criterion *c*). The scatterplots (d) show the correlation between d' and average confidence, separately for the no-feedback and feedback groups. Density distributions and box plots for each of the two feedback groups are shown separately for (e) decision RTs and (f) confidence RTs. Box plots show the median (vertical line) and the interquartile (25%–75%) range, and the whiskers indicate the 2% to 98% range. Dots indicate individual participant data. Each point in (d) represents a single participant; the dark gray line indicates the line of best fit.

 BF_{01} scores were lower than 3 in both two-sided tests, suggesting that although we have substantial evidence that feedback did not improve perceptual sensitivity, there is only anecdotal evidence that it did not change perceptual sensitivity because we cannot convincingly reject the (arguably nonsensical) possibility that feedback decreased sensitivity.

In addition to perceptual sensitivity (*d'*), we examined whether feedback had an effect on metacognitive sensitivity (meta-*d'*). We conducted an ANOVA equivalent to the one for the *d'* analyses above and found no evidence that feedback affected meta-*d'*—main effect of group: F(1, 58) = 2.29, p = .14, $\eta_p^2 = .04$, BF₀₁ = 1.2; interaction between group and day: F(6, 348) = 0.43, p = .86, $\eta_p^2 = .007$, BF_{inclusion} = .02 (Fig. 8b). As with *d'*, the feedback group showed slightly lower meta-*d'* scores, and a directed, one-sided Bayesian independent-samples *t* test confirmed that there is substantial evidence against the hypothesis that feedback increased meta-*d'* (BF₀₁ = 8.6). The rate of increase in meta-*d'* was also similar between the two groups (feedback: slope = .015, no feedback: slope = .023), *t*(58) = -0.58,

p = .56, Cohen's d = -0.18, BF₀₁ = 3.3, and again, the rate of increase was slightly higher for the no-feedback group. Taken together, these results show that even a much longer period of trial-by-trial feedback does not affect either perceptual or metacognitive sensitivity, thus providing additional evidence against the notion that feedback acts via direct reinforcement mechanisms.

Discussion

Feedback is one of the most universal and powerful ways in which behavior can be improved across domains as diverse as educational, work, and sports settings, as well as a myriad of cognitive tasks (Hattie & Timperley, 2007). However, the mechanisms through which feedback affects behavior remain unclear. Specifically, two very different accounts coexist in the literature: Feedback could have a direct effect on performance through automatic reinforcement mechanisms (Petrov et al., 2005; Thorndike, 1927), or alternatively, feedback may have only an indirect effect mediated by



Fig. 8. Effects of feedback on perceptual and metacognitive sensitivity in Experiment 2, in which feedback was provided over 7 days. Mean perceptual sensitivity (d'; a) and (b) metacognitive sensitivity (meta-d; b) are shown for each day across the feedback and no-feedback groups. Error bars depict standard errors of the mean.

a deliberate change in strategy (Vollmeyer & Rheinberg, 2005). Perceptual decision-making and metacognition provide ideal test beds for adjudicating between these two possibilities. In both cases, a direct effect via reinforcement mechanisms would predict an increase in perceptual and metacognitive sensitivity, whereas an indirect effect mediated by altered strategies would predict an exclusive effect on perceptual and metacognitive bias. Our results demonstrate that feedback does not affect perceptual or metacognitive sensitivity but reduces both perceptual and metacognitive bias. These findings strongly suggest that, in the context of simple perceptual tasks, explicit feedback affects behavior not via direct reinforcement mechanisms but by allowing participants to deliberately adjust their response strategy.

It is hardly surprising that trial-by-trial feedback allows people to alter their strategies. In fact, a large literature has already shown that different types of feedback can readily debias confidence ratings in diverse tasks (Benson & Önkal, 1992; Lichtenstein & Fischhoff, 1980; Stone & Opel, 2000). In the context of our experiment, participants biased toward choosing X can easily notice that they are wrong more frequently when picking the X response and consequently decide to select O more often, thus reducing their original bias. Similarly, when overconfident participants receive feedback about being wrong, this feedback allows them to lower their confidence ratings, thus improving their confidence calibration.

A more controversial question is whether feedback also has a direct effect on behavior via automatic reinforcement mechanisms. For example, much of the behaviorist tradition was able to achieve substantial success in training animals under the assumption that feedback exclusively acts to strengthen desirable stimulus–response associations outside of any need to posit cognitive constructs such as strategy (Skinner, 1938; Thorndike, 1927). Nevertheless, it is important to note that whether or not the experimenter poses the existence of cognitive strategies does not preclude their existence, for both humans and animals. In fact, our results strongly question the theory that feedback has a direct, automatic influence on behavior outside of the strategies adopted by the participant.

The current findings help resolve a longstanding controversy in the field of perceptual decision-making about the effects of feedback on sensitivity. A close examination of this literature reveals several issues that preclude drawing strong conclusions from previous studies. First and foremost, with total sample sizes ranging from six to 36 participants further split into groups (except Shiu & Pashler, 1992), no study was sufficiently powered. Second, researchers claiming to find a positive effect of feedback on sensitivity sometimes did not conduct a direct statistical comparison between the feedback and no-feedback groups (Seitz et al., 2006). Finally, in the rare cases in which a statistically significant difference was observed, the difference occurred for some but not other conditions and only for certain time points during training (Ball & Sekuler, 1987). These factors suggest that previous claims that feedback improves sensitivity may have been based more on the seeming plausibility of this claim rather than the strength of the evidence. The converging evidence from our two well-powered experiments (consisting of more than 200,000 trials each) implies that previous positive results could have been statistical aberrations caused by a combination of small sample sizes and preexisting assumptions that feedback must improve perceptual sensitivity.

The present work is also the first to demonstrate that feedback reduces response bias in perceptual decisionmaking. Note that our design allowed us to examine a true, nondirectional reduction in response bias by investigating whether the variability of the response criterion was smaller in the feedback group (for a similar analysis, see Hu & Rahnev, 2019). Instead, previous studies examined only directional changes in bias by exploring whether the response criterion had an overall increase or decrease (Aberg & Herzog, 2012; Goldhacker et al., 2014; Jones et al., 2015; Petrov et al., 2006; Wenger et al., 2008). However, a change in bias in a specific direction signifies only that participants have started to choose one response option over the other; such an effect cannot in and of itself be interpreted as either a reduction or an increase in bias.

The current study also sheds light on the question of what components of feedback-induced learning generalize to a new task. We found that the effects of feedback on perceptual and metacognitive bias were task specific and did not generalize to a new task. This lack of generalization may be driven by the fact that different kinds of processes lead to the response bias observed in different tasks. For example, the bias in the X/O task is likely independent from the bias in the red/ blue task, and thus reducing the former has no effect on the latter. Nevertheless, in our Experiment 1, feedback helped participants reduce their RTs, and this effect generalized to a new task. It is likely that this generalization was at least in part driven by the fact that the second task had the exact same timing structure as the first one. Future studies should examine whether this feedback-related RT advantage generalizes to tasks that differ in other aspects such as the timing and structure of the task.

An important limitation of our work is that it used only simple perceptual tasks. It is therefore an open question of whether the same effects would be obtained in the context of more complex tasks, and especially outside of the domain of perception. Nevertheless, our results fit well with the findings in fields outside of perceptual decision-making and metacognition. For example, Kantner and Lindsay (2010) showed that trial-by-trial feedback does not affect memory sensitivity across different experimental conditions chosen to create a feedback advantage. Even more tellingly, Kluger and DeNisi (1996) conducted an extensive review and showed that feedback in education can have either a positive or a negative effect on subsequent performance. They argued that the widespread assumption in education research that feedback always leads to improvement is false and suggested that the effect of feedback is determined by a number of mediating factors, which has been subsequently confirmed by other studies, too (Hattie & Timperley, 2007; Vollmeyer & Rheinberg, 2005). Another limitation of our study is that even though the decreases in perceptual and metacognitive bias are consistent with an effect of feedback on participants' conscious strategies, it remains possible that other types of automatic effects contributed to them. Nevertheless, such a possibility is made unlikely by the finding that the perceptualbias effects emerged very quickly within the first 30 trials and remained largely stable afterward (see Fig. S5), which is more consistent with an early strategy adjustment than a slow automatic process. A final limitation is that our participants were recruited via Amazon Mechanical Turk and were based only in the United States; therefore, future studies should test whether our findings generalize to other populations.

In conclusion, we found that trial-by-trial feedback reduces response bias and improves confidence calibration but does not affect perceptual or metacognitive sensitivity. These results strongly suggest that, at least in the context of perception, feedback acts via modifying participants' strategies for completing the task but has no direct, mediator-free effect on learning.

Transparency

Action Editor: Marc J. Buehner

Editor: Patricia J. Bauer

Author Contributions

D. Rahnev and N. Haddara designed the studies. N. Haddara collected and analyzed the data. Both authors interpreted the results. N. Haddara drafted the manuscript, and D. Rahnev provided critical revisions. Both authors approved the final manuscript for submission.

Declaration of Conflicting Interests

The author(s) declared that there were no conflicts of interest with respect to the authorship or the publication of this article.

Funding

This work was supported by the National Institutes of Health (Grant R01MH119189) and the Office of Naval Research (Grant N00014-20-1-2622).

Open Practices

Deidentified data for both experiments along with the code for data analysis and figure scripts have been made publicly available via OSF and can be accessed at https://osf .io/94r87/. The design and analysis plans for Experiment 2 were preregistered at https://osf.io/efymg/. Experiment 1 was not preregistered. This article has received the badges for Open Data, Open Materials, and Preregistration. More information about the Open Practices badges can be found at http://www.psychologicalscience.org/publica tions/badges.

Acknowledgments

We thank Samuel Weiss-Cowie for help with data collection and Sriraj Aiyer, Matthew Davidson, Matthew Jaquiery, and Nicholas Yeung for helpful feedback on our manuscript.

Supplemental Material

Additional supporting information can be found at http://journals.sagepub.com/doi/suppl/10.1177/09567976211032887

References

- Aberg, K. C., & Herzog, M. H. (2012). Different types of feedback change decision criterion and sensitivity differently in perceptual learning. *Journal of Vision*, *12*(3), Article 3. https://doi.org/10.1167/12.3.3
- Allen, M., Poggiali, D., Whitaker, K., Marshall, T. R., & Kievit, R. A. (2019). Raincloud plots: A multi-platform tool for robust data visualization. *Wellcome Open Research*, 4, Article 63. https://doi.org/10.12688/wellcomeopenres .15191.1
- Ball, K., & Sekuler, R. (1987). Direction-specific improvement in motion discrimination. *Vision Research*, 27(6), 953–965. https://doi.org/10.22226/2410-3535-2017-3-209-217
- Bang, J. W., Milton, D., Sasaki, Y., Watanabe, T., & Rahnev, D. (2019). Post-training TMS abolishes performance improvement and releases future learning from interference. *Communications Biology*, 2(1), Article 320. https://doi .org/10.1038/s42003-019-0566-4
- Bang, J. W., Sasaki, Y., Watanabe, T., & Rahnev, D. (2018). Feature-specific awake reactivation in human V1 after visual training. *The Journal of Neuroscience*, 38(45), 9648– 9657. https://doi.org/10.1523/JNEUROSCI.0884-18.2018
- Bang, J. W., Shekhar, M., & Rahnev, D. (2019). Sensory noise increases metacognitive efficiency. *Journal of Experimental Psychology: General*, 148(3), 437–452. https://doi.org/ 10.1037/xge0000511
- Benson, P. G., & Önkal, D. (1992). The effects of feedback and training on the performance of probability forecasters. *International Journal of Forecasting*, 8(4), 559–573. https://doi.org/10.1016/0169-2070(92)90066-I

- Carpenter, J., Sherman, M. T., Kievit, R. A., Seth, A. K., Lau, H., & Fleming, S. M. (2019). Domain-general enhancements of metacognitive ability through adaptive training. *Journal* of *Experimental Psychology: General*, 148(1), 51–64. https://doi.org/10.1037/xge0000505
- Cortese, A., Amano, K., Koizumi, A., Kawato, M., & Lau, H. (2016). Multivoxel neurofeedback selectively modulates confidence without changing perceptual performance. *Nature Communications*, 7, Article 13669. https://doi .org/10.1038/ncomms13669
- de Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a Web browser. *Behavior Research Methods*, 47(1), 1–12. https://doi.org/10.3758/ s13428-014-0458-y
- Fahle, M., & Edelman, S. (1993). Long-term learning in vernier acuity: Effects of stimulus orientation, range and of feedback. *Vision Research*, 33(3), 397–412. https://doi .org/10.1016/0042-6989(93)90094-D
- Fleming, S. M. (2017). HMeta-d: Hierarchical Bayesian estimation of metacognitive efficiency from confidence ratings. *Neuroscience of Consciousness*, 2017(1), Article nix007. https://doi.org/10.1093/nc/nix007
- Fleming, S. M., & Lau, H. C. (2014). How to measure metacognition. *Frontiers in Human Neuroscience*, 8, Article 443. https://doi.org/10.3389/fnhum.2014.00443
- Goldhacker, M., Rosengarth, K., Plank, T., & Greenlee, M. W. (2014). The effect of feedback on performance and brain activation during perceptual learning. *Vision Research*, 99, 99–110. https://doi.org/10.1016/j.visres.2013.11.010
- Green, D. M., & Swets, J. A. (1966). Signal detection theory and psychophysics. John Wiley.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81–112. https:// doi.org/10.3102/003465430298487
- Herzog, M. H., & Fahle, M. (1997). The role of feedback in learning a vernier discrimination task. *Vision Research*, 37(15), 2133–2141. https://doi.org/10.1016/S0042-6989 (97)00043-6
- Hu, M., & Rahnev, D. (2019). Predictive cues reduce but do not eliminate intrinsic response bias. *Cognition*, 192, Article 104004. https://doi.org/10.1016/j.cognition.2019 .06.016
- Ioannidis, J. P. A. (2005). Why most published research findings are false. *PLOS Medicine*, 2(8), Article e124. https:// doi.org/10.1371/journal.pmed.0020124
- JASP Team. (2020). *JASP* (Version 0.9.2) [Computer software]. https://jasp-stats.org/
- Jones, P. R., Moore, D. R., Shub, D. E., & Amitay, S. (2015). The role of response bias in perceptual learning. *Journal of Experimental Psychology: Learning Memory and Cognition*, 41(5), 1456–1470. https://doi.org/10.1037/xlm0000111
- Judd, C. H. (1905). Practice without knowledge of results. *Psychological Review Monographs*, 2(Suppl. 7), 185–198.
- Kantner, J., & Lindsay, D. S. (2010). Can corrective feedback improve recognition memory? *Memory and Cognition*, 38(4), 389–406. https://doi.org/10.3758/MC.38.4.389
- Kluger, A. N., & DeNisi, A. (1996). Effects of feedback intervention on performance: A historical review, a metaanalysis, and a preliminary feedback intervention theory.

Psychological Bulletin, *119*(2), 254–284. https://doi .org/10.1037//0033-2909.119.2.254

- Lichtenstein, S., & Fischhoff, B. (1980). Training for calibration. Organizational Behavior and Human Performance, 26, 149–171.
- Macmillan, N., & Creelman, C. (2005). *Detection theory*. LEA Publishing.
- Maniscalco, B., & Lau, H. (2012). A signal detection theoretic approach for estimating metacognitive sensitivity from confidence ratings. *Consciousness and Cognition*, 21(1), 422–430. https://doi.org/10.1016/j.concog.2011.09.021
- Masson, M. E. J. (2011). A tutorial on a practical Bayesian alternative to null-hypothesis significance testing. *Behavior Research Methods*, 43(3), 679–690. https://doi.org/10 .3758/s13428-010-0049-5
- Metcalf, J. E., & Shimamura, A. P. (1994). *Metacognition: Knowing about knowing*. MIT Press.
- Petrov, A. A., Dosher, B. A., & Lu, Z. L. (2005). The dynamics of perceptual learning: An incremental reweighting model. *Psychological Review*, 112(4), 715–743. https:// doi.org/10.1037/0033-295X.112.4.715
- Petrov, A. A., Dosher, B. A., & Lu, Z. L. (2006). Perceptual learning without feedback in non-stationary contexts: Data and model. *Vision Research*, 46(19), 3177–3197. https://doi.org/10.1016/j.visres.2006.03.022
- Pritchard, R. D., Jones, S. D., Roth, P. L., Stuebing, K. K., & Ekeberg, S. E. (1988). Effects of group feedback, goal setting, and incentives on organizational productivity. *Journal of Applied Psychology*, 73(2), 337–358.
- Rahnev, D., & Denison, R. N. (2018). Suboptimality in perceptual decision making. *Behavioral and Brain Sciences*, 41, Article e223. https://doi.org/10.1017/S0140525X18000936
- Rahnev, D., Desender, K., Lee, A. L. F., Adler, W. T., Aguilar-Lleyda, D., Akdoğan, B., Arbuzova, P., Atlas, L. Y., Balcı, F., Bang, J. W., Bègue, I., Birney, D. P., Brady, T. F., Calder-Travis, J., Chetverikov, A., Clark, T. K., Davranche, K., Denison, R. N., Dildine, T. C., . . . Zylberberg, A. (2020). The Confidence Database. *Nature Human Behaviour, 4*, 317–325. https://doi.org/10.1038/s41562-019-0813-1
- Rahnev, D., Koizumi, A., McCurdy, L. Y., D'Esposito, M., & Lau, H. (2015). Confidence leak in perceptual decision making. *Psychological Science*, 26(11), 1664–1680. https://doi.org/10.1177/0956797615595037
- Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, *85*(2), 59–108. https://doi.org/10.1037/0033-295X.85.2.59
- R Core Team. (2020). *R: A language and environment for statistical computing* (Version 4.0.2) [Computer software]. http://www.R-project.org
- Rouault, M., Dayan, P., & Fleming, S. M. (2019). Forming global estimates of self-performance from local confidence. *Nature Communications*, 10(1), Article 1141. https://doi.org/10.1038/s41467-019-09075-3
- Seitz, A. R., Nanez, J. E., Holloway, S., Tsushima, Y., & Watanabe, T. (2006). Two cases requiring external reinforcement in perceptual learning. *Journal of Vision*, 6(9), 966–973. https://doi.org/10.1167/6.9.9
- Shibata, K., Yamagishi, N., Ishii, S., & Kawato, M. (2009). Boosting perceptual learning by fake feedback. *Vision*

Research, 49(21), 2574–2585. https://doi.org/10.1016/j .visres.2009.06.009

- Shiu, L. P., & Pashler, H. (1992). Improvement in line orientation discrimination is retinally local but dependent on cognitive set. *Perception & Psychophysics*, 52(5), 582–588. https://doi.org/10.3758/BF03206720
- Skinner, B. F. (1938). *The behavior of organisms: An experimental analysis*. Appleton-Century.
- Stone, E. R., & Opel, R. B. (2000). Training to improve calibration and discrimination: The effects of performance and environmental feedback. Organizational Behavior and Human Decision Processes, 83(2), 282–309. https:// doi.org/10.1006/obhd.2000.2910
- Thorndike, E. L. (1927). The law of effect. *The American Journal of Psychology*, *39*(1/4), 212–222.
- van Ravenzwaaij, D., Donkin, C., & Vandekerckhove, J. (2017). The EZ diffusion model provides a powerful test of simple

empirical effects. *Psychonomic Bulletin and Review*, 24(2), 547–556. https://doi.org/10.3758/s13423-016-1081-y

- Vollmeyer, R., & Rheinberg, F. (2005). A surprising effect of feedback on learning. *Learning and Instruction*, 15(6), 589–602. https://doi.org/10.1016/j.learninstruc.2005 .08.001
- Wagenmakers, E.-J., Van der Maas, H. L. J., & Grasman, R. P. P. P. (2007). An EZ-diffusion model for response time and accuracy. *Psychonomic Bulletin & Review*, 14(1), 3–22.
- Wenger, M. J., Copeland, A. M., Bittner, J. L., & Thomas, R. D. (2008). Evidence for criterion shifts in visual perceptual learning: Data and implications. *Perception and Psychophysics*, 70(7), 1248–1273. https://doi.org/10.3758/ PP.70.7.1248
- Wright, W. R. (1906). Some effects of incentive on work and fatigue. *Psychological Review*, 13, 23–34.