Selective memory searching does not explain the poor recall of future-oriented feedback

Robert A. Nash[1*], Naomi E. Winstone[2], & Samantha E. A. Gregory[1]

[1] School of Psychology, Aston University

[2] Surrey Institute of Education, University of Surrey

Author note. This research was funded by the Leverhulme Trust (Research Project Grant RPG-2016-189). The authors are grateful to Maryanne Garry for discussions that inspired Experiment 3.

* Corresponding author

Robert A. Nash
School of Psychology
Aston University
Birmingham, B4 7ET
United Kingdom
+44 (0) 121 204 4522
R.Nash1@aston.ac.uk

Abstract

Feedback is invaluable for learning, yet people frequently fail to remember their feedback. Recent studies have demonstrated that people are better at recalling evaluative, past-oriented feedback than directive, future-oriented feedback. This paper tests one possible explanation: namely, that people neglect to search their memory for directive information they have encoded. Participants (total *N* = 759), attempted to recall feedback they had read about their own (Experiment 1) or another person's performance (Experiments 2A-4). We attempted to foster recall of directive feedback via a structured recall task (Experiments 1-2B) or a perspective-taking instruction (Experiment 3). All experiments replicated the preferential recall of evaluative feedback, but our manipulations did not moderate this bias. Experiment 4 replicated the bias using 'non-educational' feedback, and provided tentative indications that it might not translate beyond the feedback domain. The data suggest that selective retrieval processes are not responsible for people's poor recall of directive feedback.

General Audience Summary

Receiving feedback can be an invaluable way of improving our skills and abilities. However, people are not always good at remembering the feedback they receive. In particular, some recent research has provided evidence that people tend to be better at remembering past-oriented feedback comments, which focus on how they have performed, than at remembering future-oriented feedback comments, which focus on how they could improve next time. In this paper we aimed to test one possible explanation of that finding; namely, that when trying to recall feedback, people tend to spontaneously focus on the past and fail to think about the future. In five experiments, participants received feedback either about their own performance (Experiment 1) or someone else's performance (Experiments 2A&B, 3, and 4), and shortly afterwards they took a memory test. In all five experiments we found that people remembered past-oriented comments better than future-oriented comments, replicating the findings of prior research. However, this memory bias remained true even for those participants who were explicitly tasked with remembering past- and future-oriented feedback separately (Experiments 1 and 2A&B), and for those who tried to take the perspective of the person giving the feedback (Experiment 3). In Experiment 4 we found some suggestive—but ultimately inconclusive—evidence that this memory bias might be specific to the way people recall judgmental feedback, rather than other, nonjudgmental kinds of information. Together, these experiments fail to support the idea that people recall future-oriented feedback poorly because they neglect to retrieve it from memory. But these studies provide further corroboration that people's memory for their feedback can depend on subtle differences in wording.

Selective memory searching does not explain the poor recall of future-oriented feedback

We might not always desire it, but people are invariably keen to offer us feedback on the everyday tasks we undertake, from how we perform in the workplace, to the taste of our home cookery. Many experts and coaches who specialize in personal and professional development believe that feedback is especially effective when it is future-oriented: when it emphasizes our next steps for improvement rather than our shortcomings thus far (e.g., Goldsmith, 2010; Hirsch, 2017). Whereas this belief might to some extent be warranted, recent studies demonstrate an important drawback of future-oriented feedback relative to past-oriented feedback: people are less likely to remember future-oriented feedback (Gregory et al., 2020; Nash et al., 2018). Why is this? Here we test the hypothesis that people fail to adequately search for future-oriented feedback during memory retrieval.

Research from cognitive psychology shows that learners often fail to remember the feedback they receive (e.g., Butler & Roediger, 2008). In many circumstances these failures are unlikely to be problematic: if people can repeatedly revisit their feedback, and are motivated to do so, then they might have little need to commit the feedback to memory. But people are not always able or motivated to review their feedback. Indeed, many university students engage minimally with written feedback, affording themselves only a single—often brief—exposure to the advice it contains (Nash & Winstone, 2017; Robinson et al., 2013). In these contexts then, it is clear that memory

processes can be pivotal in determining the effectiveness of feedback. As Irwin et al. (2013, p. 57) put it, "If students cannot remember their feedback when doing future assessments, it would be difficult to use it to improve performance on those assessments."

A wealth of empirical research shows that even minor changes to the wording of feedback comments can have substantive effects on how the recipient interprets and engages with them (Douglas & Skipper, 2012). A recent research programme showed that simple linguistic variations can also influence learners' tendency to remember their feedback. In the first of a series of experiments, Nash et al. (2018) asked participants to write essays on controversial topics, with the understanding that they would receive personalized feedback. Days later, participants indeed received detailed feedback, but unbeknownst to them the comments were generic: all participants saw a variation of the same feedback. Each critical comment was delivered in one of two written styles – either an *evaluative*, past-oriented style (for example "You didn't always demonstrate a sophisticated awareness of the issues you covered"), or a *directive*, future-oriented style ("You should aim to demonstrate a more sophisticated awareness of the issues you cover"). Shortly afterwards, participants completed a surprise memory test in which they attempted to reproduce as much of their feedback as possible from memory. The results revealed what the authors called an *evaluative recall bias* – participants were considerably more likely to retrieve comments that were delivered in the evaluative style, than those delivered in the directive style.

This bias held in all six of Nash et al.'s experiments, and in the three experiments subsequently described by Gregory et al. (2020).

The evaluative recall bias, it seems, is a replicable memory phenomenon. And yet none of the published data offer satisfactory evidence of the cognitive mechanism that underpins this bias. In fact, even though Nash et al. (2018) and Gregory et al. (2020) sought to test numerous theoretical accounts, their data did not support any of these accounts. For instance, Nash et al. observed the evaluative recall bias even when participants read another fictional person's feedback rather than their own, which provides evidence against the causal roles of self-referential encoding (of evaluative information) and motivated avoidance (of directive information; Rogers et al., 1977; Sweeny et al., 2010). Furthermore, the evaluative recall bias emerged even when directive feedback comments were delivered using stern, unfriendly language and the evaluative comments in more supportive, friendly language. This latter finding suggests that the bias is not a consequence of participants preferentially encoding information that seems critical or harsh.

An especially important finding is that evaluative feedback comments do not receive greater visual attention than directive comments, as measured by tracking participants' eye-movements (Gregory et al., 2020). Indeed, in two of their earlier experiments Nash et al. (2018) found that whereas evaluative feedback was better recalled than directive feedback, participants were equally able to *recognize* both feedback types in a two-alternative forced-

choice test. These two sets of findings suggest that both varieties of feedback are processed equivalently during encoding. They therefore lead us to doubt whether memory encoding factors can account for the evaluative recall bias, and lead us to look instead toward retrieval factors.

Selective memory search

A thus far untested account of the evaluative recall bias is that people fail to adequately search their memory for directive information during their retrieval attempts. The cognitive psychology literature is replete with empirical examples of such biased memory search processes. In one classic study, participants read a story about some boys playing in a house, and were asked to mentally take the perspective of either a burglar or a prospective home-buyer (Anderson & Pichert, 1978). After a delay, participants attempted to recall the story. The researchers found that participants tended to recall information that was relevant to their mental perspective (e.g., home-buyers recalled that the basement was damp) more reliably than information relevant to the alternative perspective. After a further delay, participants recalled the story a second time, this time either from the same mental perspective, or from the alternative perspective—burglars became home-buyers, or vice versa. On this second recall attempt, people who changed perspectives now recalled significantly more new information than did people who kept the same perspective. These findings tell us that people's goals and schemas can guide their memory search strategies, and that goal-irrelevant information is not always reliably retrieved despite being successfully encoded.

Anderson and Pichert's (1978) findings have been replicated numerous times (e.g., Kaakinen et al., 2011; Surber, 1983; although see e.g., Ginet et al., 2018 for null findings), and several other studies similarly demonstrate how a switch of task-framing or goals can successfully glean additional information from memory. Gilbert and Fisher (2006), for instance, showed participants a video of a robbery shortly before asking them to freely recall the details. Whereas some participants were given no particular recall strategy to follow, others took the perspective of a police officer, and some took the perspective of the robber. Similar to Anderson and Pichert's (1978) procedure, two days later participants were interviewed again, sometimes with the same recall perspective but sometimes the alternate perspective. The authors found that participants who adopted a different retrieval perspective during their second interview recalled significantly more new information than did those who adopted the same, or no, retrieval perspective. Indeed, reputed memory-elicitation techniques such as the Cognitive Interview have sometimes promoted the value of perspective-change instructions for increasing memory output (Fisher et al., 2011).

Why might these kinds of findings be relevant to people's tendency to recall evaluative feedback better than directive feedback? We reasoned that when recalling feedback, people's retrieval processes might similarly be guided by pre-existing beliefs about the general purpose of feedback. Specifically, even though people typically say they prefer receiving future-oriented feedback (Winstone et al., 2016), we predicted that people might

nevertheless typically understand feedback as a past-oriented process: one wherein a person communicates their judgment and critique of an output or performance. Indeed, one recent study found that university students were far less likely to access the written feedback on their assignments if their grade was available separately to this feedback (Mensink & King, 2020). These kinds of findings indicate that despite people's reported preferences, they in fact tend to see feedback rather more as a past-oriented, judgmental process. Supplementary data from Nash et al. (2018, p. 1874) also support this reasoning: when asked to rate individual evaluative and directive feedback comments, participants tended to perceive the past-oriented, "evaluativeness" of directive comments (i.e., the extent to which they were 'about' the quality of the work) as significantly greater than the future-oriented, "directiveness" of evaluative comments (i.e. the extent to which they were 'about' how to improve next time).

Based on this line of reasoning, we might expect that the evaluative recall bias is a product of selective, schema-driven memory search. Specifically, if people understand feedback as a past-oriented process, then this schema could lead them to selectively search their memories for evaluative comments, and to neglect memories of directive comments.

The present research

If our reasoning above were correct, then we should be able to eliminate the evaluative recall bias by asking people to separately retrieve directive information and evaluative information – that is, by interrupting

their tendency to neglect directive feedback by directly probing this information. In the first three experiments here, we tested this prediction by comparing the recall performance of participants who retrieved feedback in an unstructured free recall test (as per Gregory et al., 2020; Nash et al., 2018), to those who completed a structured recall test that probed separately for memories of past-oriented and future-oriented feedback.

Experiment 1

Method

We pre-registered the procedure and analytic plan for Experiment 1 at https://aspredicted.org/blind.php?x=548an4. The data associated with all the experiments in this paper can be downloaded from https://osf.io/cn2ju/

Participants. All participants were university students who participated in exchange for either course credit or for a cash incentive. Our pre-registered target sample was 159 participants; we slightly over-sampled and recruited 171 participants (145 females, 20 males, and 6 who did not specify their gender; mean age = 20.44, *SD* = 3.62, range 18-52). Any respondents who failed to complete the study in full were automatically removed from the dataset and replaced; likewise, in accordance with our pre-registered protocol, any participant who failed to retrieve at least one correct item of feedback in Session 2 was removed from the dataset and replaced.

Materials. We used the two feedback script versions published in Nash et al. (2018; Script 1A and 1B) as materials in this experiment, and to enhance the generalisability of the materials we also used the two script versions

subsequently published in Gregory et al. (2020; Script 2A and 2B). During the experiment each participant received one of these four standardized feedback script versions (Script 1, version A = 418 words; Script 1, version B = 411 words; Script 2, version A = 415 words; Script 2, version B = 407 words). Each script was divided into three subsections titled "substance", "style", and "format," and each subsection contained critical feedback comments prefixed and suffixed by brief points of praise. The praise in these scripts served solely to make the feedback seem more palatable and realistic to participants, and was not relevant to the study aims.

There were 20 critique comments in each feedback script, which were always presented to participants in the same order. Half of these critique comments were written in an *evaluative style* (for example, "You did not always try to provoke your reader's thinking, and focused instead on arguments that they would expect"); the other half were written in a *directive style* ("You could try to provoke your reader's thinking more, by focusing on arguments that they would find unexpected"), and the style of each comment was counterbalanced across the two versions of Script 1 and Script 2. As the examples here illustrate, we manipulated the style of feedback comments through minimal re-wording, in ways that maintained the comments' general meaning while approximately equating their length and complexity. In each

feedback script we presented critique comments in pairs that alternated between the evaluative and directive style throughout.[1]

Procedure. Participants were recruited for a study on "Personality and persuasive writing", which involved two sessions in the laboratory spaced 2-3 days apart.

Session 1. In the first session, participants completed a persuasive writing task. Each participant chose four essay topics from a list of ten contentious titles that appeared on their computer screen (e.g., "Should students have to pay for their university education?"). Next, a random one of the chosen titles appeared on the screen, and the participant spent 5 mins typing a persuasive argument on that topic. A timer on the screen reminded participants of how much time they had remaining, and the page automatically changed when the time was up. At this point the next essay title appeared, and this process was repeated until the participant had written all four essays. Participants were told that a member of staff would mark their essays, and that they would receive detailed feedback in Session 2. They were also told, falsely, that they would be asked to write more essays in Session 2, and that they would need to apply the feedback to improve their writing.

[1] When Script (1 vs. 2) was included in our main analyses as a between-subject variable, there were no significant two- or three-way interactions involving Script and Feedback Type, in any of Experiments 1, 2A, or 2B. That is to say, none of the effects of feedback type depended statistically upon which script the participants had received. In Experiment 2B there was a significant overall main effect of Script, and in Experiment 2A there was a significant Script x Test Format interaction, but neither of these effects is relevant to our predictions and we do not discuss these further.

Session 2. At the start of Session 2, participants were told that the feedback on their essays had been prepared; however, instead of giving them genuine feedback, they were instead shown at random one of our four generic feedback scripts on their computer screen. Participants were told they could take as long as they wished to read "their" feedback, and that they should click to continue once they had finished.

Next, participants completed a 3-min filler task involving simple logic puzzles, before being told there would be an unexpected memory test. At this point they were randomly assigned to either the *free recall* condition, or to one of our two *structured recall* conditions. Participants in the free recall condition were asked to type as much of their feedback they could remember. They were unable to finish this task until they had spent at least 6 min on their recall attempt. Participants in the structured recall conditions, in contrast, were asked two separate questions in the memory test. Specifically, on one page of the test they were asked to recall as much as possible about how they had performed in the writing task (i.e., evaluative feedback). On the other page of the test they were asked to recall as much as possible about how they could improve in the second writing task (i.e., directive feedback). For the purpose of counterbalancing, structured recall participants viewed these two pages in a random order, and we analysed these two groups' data separately (per our pre-registered plan) to check for potentially informative order effects. That is to say, *evaluative-first* participants were asked about evaluative feedback and then about directive feedback, whereas the reverse was true for

directive-first participants. Participants in both structured recall groups were required to spend at least 3 min on each of their two recall attempts.

After completing the memory test, participants were asked to rate the fairness (1 = very unfair; 5 = very fair) and helpfulness (1 = very unhelpful; 5 = very helpful) of their feedback. They were also asked to guess what percentage grade they would have received for their essays, and what grade they might achieve next time having received the feedback. They were then given an opportunity to provide open comments about the feedback, and were finally debriefed. Note that whereas four participants indicated some degree of suspicion here over whether the feedback was personalized to themselves, the vast majority of comments suggested that participants accepted it as individual feedback (e.g. "I believe the feedback was fair, and I would likely have given myself similar feedback").

Memory coding. One researcher coded all of the memory responses, blind to experimental condition. This researcher knew whether each participant had received Script 1 or Script 2, but did not know which version (A or B) of the script they had seen, and therefore did not know in which style each individual critique comment had been presented.

To code the data, the researcher judged which of the 20 critique comments from the feedback each participant had recalled. If participants recalled any of the praise comments, these were ignored. After the researcher had coded every response, we revealed which feedback script version each participant had seen, which permitted us to establish (1) the number of

evaluative feedback comments that had been recalled, and (2) the number of directive comments recalled. To assess the reliability of the data coding, a second researcher also coded 20% of these data in the same manner as the first researcher. The inter-rater agreement was good (total evaluative comments recalled, r = .89; total directive comments recalled, r = .88). Therefore, all analyses are based on the first researcher's coding.

Deviations from pre-registered protocols. In the present experiments, as well as assessing which feedback comments participants recalled, our original intention was to also assess the style (evaluative or directive) in which they reproduced each comment. Doing this would have allowed us to test—in line with our pre-registered protocols—whether the between-group manipulations in each experiment reduced a second kind of memory bias seen in prior research: the tendency to reproduce feedback comments in an overly-evaluative style (Gregory et al., 2020; Nash et al., 2018).

However, a statistical artifact in Experiments 1-3 prevented us from examining this secondary bias as planned. In Nash et al.'s and Gregory et al.'s work, participants sometimes recalled particular feedback comments twice: once in an evaluative style and once in a directive style. Because this happened rarely in their work (up to ~4% of the comments recalled), those authors took the approach of 'double-counting' each such comment: that is, these comments were coded as being recalled once in the "correct style" and once in the "incorrect style". In Experiments 1-3 of the present work though, this "double-counting" happened much more frequently, and was far more

common in certain experimental conditions than in others (see Table S1 in the supplemental materials for details). Taking the 'double-counting' approach would therefore have introduced a significant confound into the data, artificially boosting the recall rates in certain conditions over others.

In addition to this statistical problem, in Experiment 4 we found a different problem. There, we found that some parts of participants' free recall could not be coded as either evaluative or as directive, and that this coding problem happened exclusively in one experimental condition (the 'description' condition; M = 0.21 comments or 3.0% of the comments recalled). In short, looking at retrieval style caused a statistically significant confound in this experiment, too.

To resolve these statistical issues, we decided to deviate from our pre-registered analysis plans for all of the present experiments. Specifically, we chose to prioritize answering our primary research question by assessing only which feedback comments were recalled, regardless of whether these were recalled in the correct style, the incorrect style, or indeed in both styles. Therefore, whereas we had planned to include retrieval style as an independent variable in our statistical analyses, this variable was omitted from all analyses throughout this paper. For more information, Tables S2-S6 of the supplemental materials provide more detail about the style in which participants retrieved the feedback within each of the experiments reported in this paper. In addition, the variables required for conducting our original, pre-registered analysis approach (e.g. number of evaluative comments

recalled in a directive style, including double-counting) can be found in our open datasets.

Results and Discussion

Main analysis. Our main question was to what extent the evaluative recall bias would be statistically smaller among participants who completed a structured recall test, than among those who took a free recall test. To answer this question we examined how many of the 20 individual feedback comments participants recalled in total. We analyzed these data using a 3 (test format: free recall vs. evaluative-first vs. directive-first) x 2 (feedback type: evaluative vs. directive) mixed-factor ANOVA, with repeated measures on the latter variable.

This analysis first revealed a statistically significant main effect of feedback type: that is, an evaluative recall bias whereby evaluative comments were better recalled than directive comments, $F(1, 168) = 43.93$, $p < .001$, $\eta^2_p = .21$, $d = 0.51$, $M_{diff} = 1.18$ comments, 95% CI [0.83, 1.52]. As Figure 1 shows though, there was no significant main effect of test format, $F(2, 168) = 1.38$, $p = .26$, $\eta^2_p = .02$, and most importantly, the predicted two-way interaction between feedback type and test format was not significant, $F(2, 168) = 0.30$, $p = .74$, $\eta^2_p = .00$. In other words, participants showed a similar evaluative recall bias regardless of the format of the memory test.

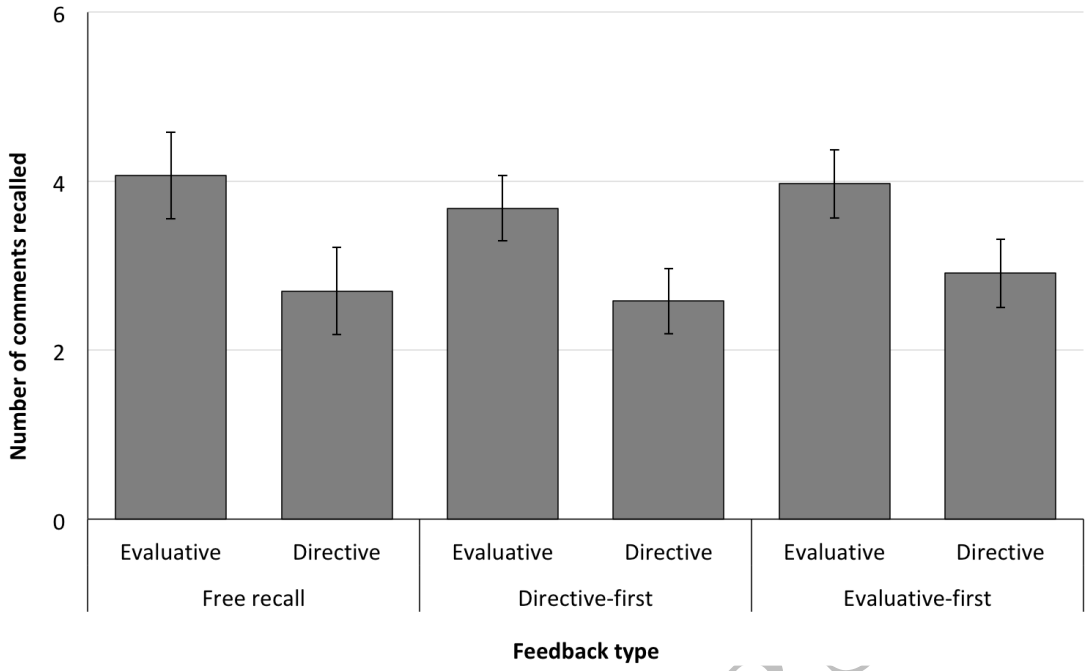

Figure 1. Number of evaluative and directive feedback comments recalled by participants in Experiment 1, according to test format (maximum possible = 10 for each feedback type). Error bars are 95% within-participant confidence intervals, calculated separately for each between-participant condition (Loftus & Masson, 1994).

This set of results, in short, therefore provides little support for the argument that people might selectively neglect to search their memories for directive feedback. In particular, when participants were explicitly prompted to recall the comments about how they could improve, separately from the comments about how they had performed, they nonetheless still exhibited an evaluative recall bias of similar magnitude to that seen in the free recall condition.

Additional analysis. There were no main effects of test format on participants' ratings of the feedback's fairness (M = 3.97 out of 5 across conditions, SD = 0.76), helpfulness (M = 4.11, SD = 0.66), estimates of the

writer's grade (M = 49.5%, SD = 13.0%), or estimates of the future grade the writer could achieve after taking the feedback on board (M = 63.4%, SD = 11.7%), all $p > .19$.

Experiments 2A and 2B

In Experiments 2A and 2B we conducted replications using the same general design as Experiment 1, but this time our participants did not complete an initial persuasive writing task. Instead, participants were merely given a piece of feedback and told it belonged to another student who had completed a writing task. Nash et al. (2018) found that the evaluative recall bias can be obtained using this more straightforward study procedure; we therefore set out to examine the extent to which the recall test format would affect people's recollection of another fictional person's feedback.

Method

We pre-registered the protocol and analytic plan at https://aspredicted.org/blind.php?x=ia4rt4. Deviations from this protocol are explained in Experiment 1.

Participants. Participants were recruited via an online panel provider. Our pre-registered target sample for each replication was 159 participants. In Experiment 2A we slightly over-sampled and recruited 169 participants. Due to a technical error, participants' age and sex data were not collected in that experiment. In Experiment 2B our final sample comprised 159 participants (110 females, 39 males, and 10 who did not specify their gender; mean age = 42.7, SD = 16.3, range = 18-78). Any respondents who failed to complete the

studies in full were automatically removed from the dataset and replaced; likewise, in accordance with our pre-registered protocol, any participant who failed to retrieve at least one correct item of feedback was removed from the dataset and replaced.

Procedure and materials. The procedure and design of Experiments 2A and 2B mirrored Session 2 of Experiment 1. That is to say, our participants did not complete any kind of writing task (as they had in Session 1 of Experiment 1); instead they were simply given—at random—one of the four feedback scripts used in Experiment 1, and told that it was the feedback received by a student who had completed a set of short essays. Participants were asked to carefully read this student's feedback at their own pace, and to click to continue once they were ready. Next, participants solved logic puzzles for 3 min, before being randomly assigned to one of the same three test format conditions used in Experiment 1 (i.e., free recall, evaluative-first, and directive-first). All participants were given the surprise memory test as in Experiment 1; free recall participants were required to spend at least 6 min on this test, whereas those in the two structured-recall conditions were required to spend at least 3 min on each of the two pages. Finally, participants completed the same ratings as were used in Experiment 1, except that we removed the 'fairness' item.

Memory coding. One researcher coded all of the memory responses for each experiment, blind to experimental condition and script version, and a second researcher also coded 20% of the data for each experiment in the same

way. The agreement was good (Experiment 2A: total evaluative comments recalled, r = .77; total directive comments recalled, r = .83; Experiment 2B: total evaluative comments recalled, r = .85; total directive comments recalled, r = .88), therefore all analyses are based on the first researcher's coding.

Results and Discussion

Main analysis. For Experiment 2A we analyzed the recall data using a 3 (test format: free recall vs. evaluative-first vs. directive-first) x 2 (feedback type: evaluative vs. directive) mixed-factor ANOVA, with repeated measures on the latter variable. This analysis, illustrated in Figure 2, revealed a statistically significant main effect of feedback type, with evaluative feedback recalled better than directive feedback, $F(1, 166) = 6.11$, $p = .01$, $\eta^2_p = .04$, $d = 0.19$, $M_{diff} = 0.33$ comments, 95% CI [0.07, 0.60]. However, just as in Experiment 1, there was no overall main effect of test format, $F(2, 166) = 1.54$, $p = .22$, $\eta^2_p = .02$, and no significant interaction of feedback type and test format, $F(2, 166) = 1.37$, $p = .26$, $\eta^2_p = .02$.

When we conducted the same analyses for the Experiment 2B data, the results were identical. Specifically, there was a significant overall advantage in the recall of evaluative feedback as compared to directive feedback, $F(1, 156) = 10.55$, $p < .01$, $\eta^2_p = .06$, $d = 0.26$, $M_{diff} = 0.48$ comments, 95% CI [0.19, 0.77] (see Figure 2), but there was no overall main effect of test format, $F(2,$

156) = 0.70, p = .50, η^2_p < .01, and no significant interaction, $F(2, 156) = 0.64$, p = .53, η^2_p < .01.

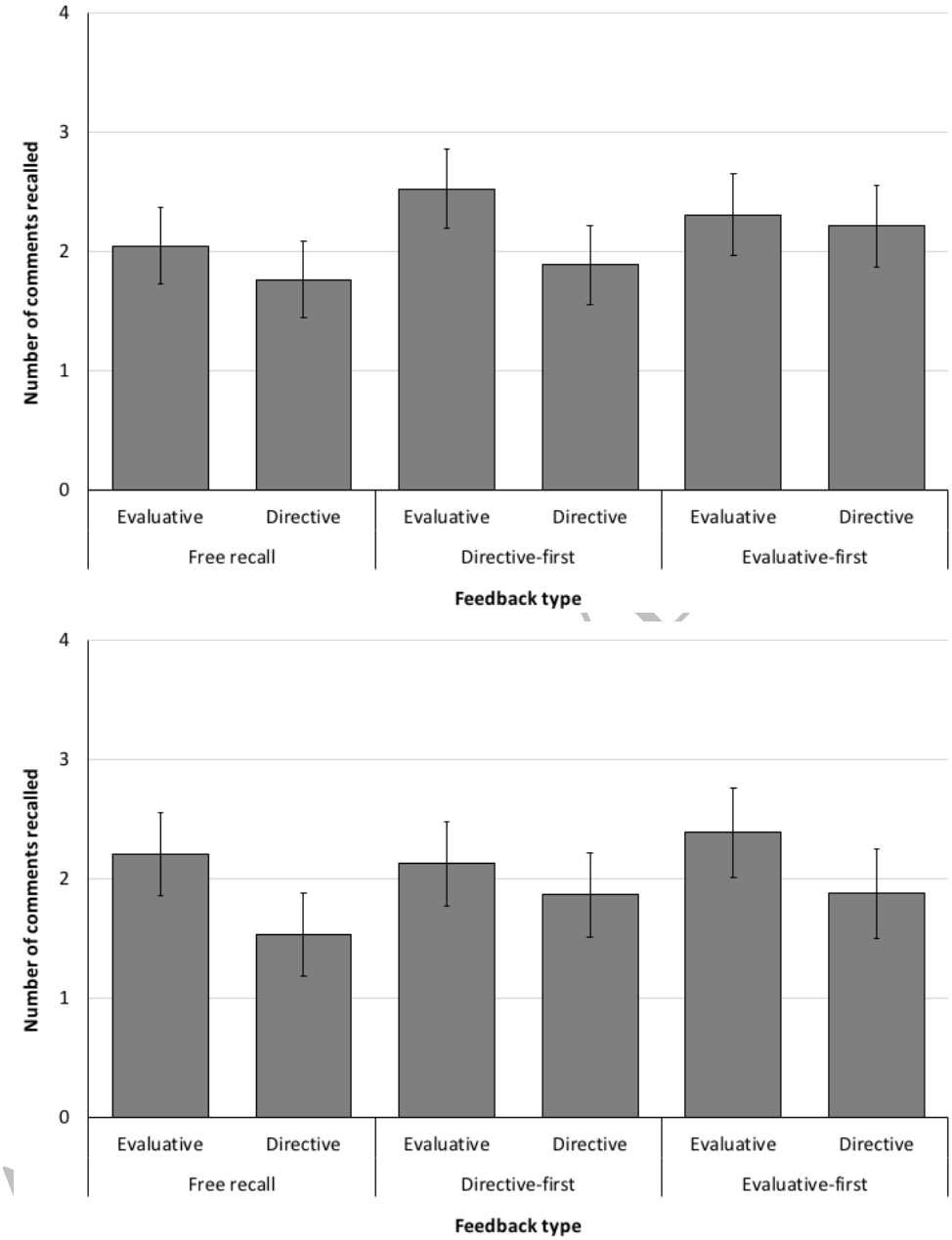

Figure 2. Number of evaluative and directive feedback comments recalled by participants in Experiments 2A (top) and 2B (bottom), according to test format (maximum possible = 10 for each feedback type). Error bars are 95% within-participant confidence intervals, calculated separately for each between-participant condition (Loftus & Masson, 1994).

Overall, the results of Experiments 2A and 2B mirror very closely those of Experiment 1. In short, even when participants were explicitly asked to recall details of how they could improve in future (separately from details of how they had performed), they still exhibited an evaluative recall bias. These three experiments therefore provide minimal evidence to support the hypothesis that people fail to spontaneously search their memories for future-oriented information.

Additional analysis. There were no main effects of test format on participants' ratings of feedback helpfulness (Experiment 2A: M = 4.02 out of 5 across conditions, SD = 1.00; Experiment 2B: M = 4.07, SD = 0.80), estimates of the writer's grade (Experiment 2A: M = 70.3%, SD = 14.1%; Experiment 2B: M = 71.0%, SD = 12.2%), or estimates of the future grade the writer could achieve after taking the feedback on board (Experiment 2A: M = 81.0%, SD = 14.5%; Experiment 2B: M = 82.9%, SD = 12.4%), all $p > .20$.

Experiment 3

A theoretical premise that underpinned our initial hypothesis was that people might generally understand feedback as a past-oriented process. That is to say, the information that people selectively retrieve from memory may be shaped by their schematic representations of what feedback actually is, or what it is for. Before concluding that the selective retrieval hypothesis is unsupported, we wanted to test some alternative approaches to manipulating how people might apply schemas when reading feedback comments, and to test the extent to which these manipulations would influence people's recall.

In Experiment 3 we first used a perspective-taking manipulation, asking participants to take the perspective of either the student receiving the feedback, or the teacher giving the feedback. Based in part on Anderson and Pichert (1978), we speculated that people might adopt different schemas of feedback when taking each of these different perspectives. Specifically, we reasoned that someone approaching feedback from the perspective of a student may adopt a schema that emphasises how it feels to be critiqued, and may consequently be more attuned to performance-related (evaluative) information. In contrast, we reasoned that someone taking a teacher's perspective may adopt a schema that emphasises feedback's developmental intent, and would therefore be more attuned to development-related (directive) information. Research shows that people tend to encode and recall information better when it is relevant to their goals (e.g. Eitam et al., 2009; Montagrin et al., 2013), and we therefore predicted that this perspective-taking manipulation could influence participants' active feedback schemas, and thus moderate the evaluative recall bias.

Method

We pre-registered the protocol and analytic plan for Experiment 3 at https://aspredicted.org/blind.php?x=b9ix97. Deviations from this protocol are explained in Experiment 1.

Participants. Participants were recruited via an online panel provider, and were included in the final sample or replaced using the same criteria as Experiments 2A/B, as well as excluding any participant who failed the

attention check described below. Our final sample comprised 128 participants (87 females, 39 males, 1 other, and 1 who did not disclose their gender; mean age = 53.4, *SD* = 15.2, range = 20-83).

Procedure and materials. The procedure generally mirrored that of Experiments 2A/B. To begin, participants were told they would see some feedback that was received by a student who had completed some writing. Each participant was randomly assigned to either the *Student* condition, or the *Teacher* condition. Those in the Student group were instructed "When you see the feedback, take a moment to imagine you are the student who is receiving it. Try to put yourself in the student's shoes, and reflect on what your teacher wants to help you to understand." Participants in the Teacher condition were instead told: "When you see the feedback, take a moment to imagine you are the teacher who is giving it. Try to put yourself in the teacher's shoes, and reflect on what you want to help your student to understand."

Participants were then shown at random one of the two versions of feedback script 1 (i.e., the scripts originally reported by Nash et al., 2018), and were asked to carefully read the feedback at their own pace. Next, participants solved logic puzzles for 3 min, before being given a surprise free recall test. Participants were required to spend at least 5 min on this test.

Next, participants completed an attention check question in which they were asked to select, from a choice of five options, the perspective-taking instruction they received before reading the feedback. Participants who failed this attention check were excluded and replaced in accordance with our pre-

registration. Finally participants rated the helpfulness and fairness of the feedback as in Experiment 1, and they rated how much they thought the feedback had focused on (a) how the student had performed, and (b) how the student could improve next time.

Memory coding. One researcher coded participants' responses blind to condition and script version, and a second researcher coded 20% of responses. The agreement was good (total evaluative comments recalled, r = .84; total directive comments recalled, r = .76), therefore all analyses are based on the first researcher's coding.

Results and Discussion

A 2 (perspective: student vs. teacher) x 2 (feedback type: evaluative vs. directive) mixed-factor ANOVA revealed a main effect of feedback type, with evaluative feedback recalled better than directive feedback, $F(1, 126) = 7.68$, $p < .01$, $\eta^2_p = .06$, $d = 0.25$, $M_{diff} = 0.40$ comments, 95% CI [0.11, 0.69] (see Figure 3). However, there was no overall main effect of perspective, $F(1, 126) = 0.05$, $p = .83$, $\eta^2_p < .001$, and no significant interaction – the evaluative recall bias was of an equivalent magnitude regardless of which perspective participants took, $F(1, 126) = 0.03$, $p = .85$, $\eta^2_p < .001$. Put differently, Experiment 3 found that people's preferential recall of evaluative feedback held even among participants who were asked to put themselves in the feedback-giver's shoes. Insofar that this perspective manipulation emulated aspects of Anderson and Pichert's (1978) design, our data provide no evidence for a role of schematic retrieval processes in the evaluative recall bias.

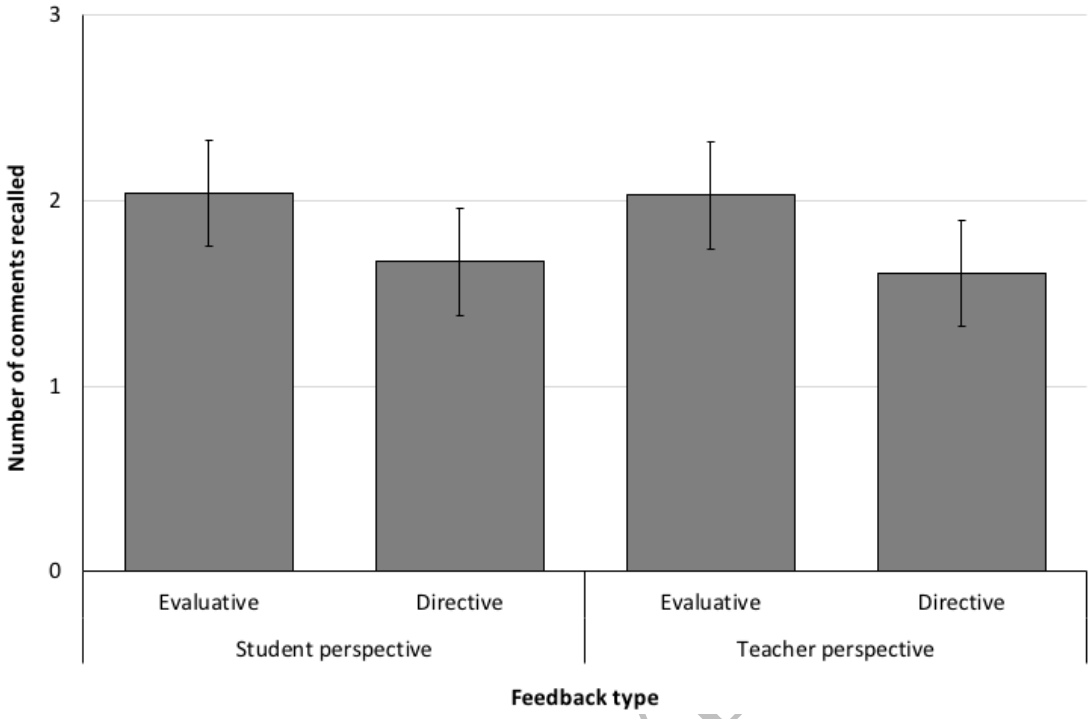

Figure 3. Number of evaluative and directive feedback comments recalled by participants in Experiment 3, according to perspective (maximum possible = 10 for each feedback type). Error bars are 95% within-participant confidence intervals, calculated separately for each between-participant condition (Loftus & Masson, 1994).

Additional analysis. There were no significant main effects of perspective on participants' ratings of how much the feedback focused on past performance (M = 5.72 out of 7 across conditions, SD = 1.06), or future improvement (M = 6.01, SD = 1.19). Participants who took the student's perspective believed the feedback was more fair (M = 5.79 vs. 5.15, $t(126)$ = 2.94, $p < .01$, d = 0.52), and somewhat more helpful (M = 5.88 vs. 5.39, $t(126)$ = 1.95, p = .054, d = 0.35), than did participants who took the teacher's perspective.

Experiment 4

Our perspective-taking manipulation in Experiment 3 was ineffective in influencing the evaluative recall bias (and indeed, perhaps even ineffective in influencing perspective-taking). In Experiment 4 we tried a different approach, by attempting to manipulate whether or not the information participants read would even be considered 'feedback' at all. We reasoned that if a memory bias is driven by schematic retrieval processes, then theoretically it should only occur in circumstances where the to-be-recalled information is schema-relevant. Therefore, if the evaluative recall bias is a consequence of how people understand the concept of feedback (i.e., that feedback serves primarily to communicate an evaluation or judgment), then the bias should be absent when people receive evaluative and directive information that is clearly not feedback.

Method

We pre-registered the protocol and analytic plan for Experiment 4 at https://aspredicted.org/blind.php?x=ex3e4p. Deviations from this protocol are outlined in Experiment 1.

Participants. Participants were recruited via an online panel provider. Our pre-registered target sample was 128 participants; we slightly over-sampled and recruited 132 participants (97 females and 35 males; mean age = 53.35, *SD* = 14.50, range 24-81).

Materials. For the purposes of this experiment we created new scripts. Our aim was to cover subject matter in these scripts that could be communicated either in a judgmental, "feedback" style, or in a

nonjudgmental, purely descriptive style. We therefore produced new scripts that gave a commentary on a fictional person's performance in a cookery task. For the "feedback" scripts, we prepared a set of critical comments, each formulated in both an evaluative and a directive style. Specifically, the evaluative comments described things that could have gone better in the cookery task (e.g., "When preparing the sauce, you could have been more precise by grating the exact amount of cheese required"), and the directive comments described things that could be improved in a future cookery task (e.g., "When preparing the sauce, you need to be more precise and grate the exact amount of cheese required"). For the "description" scripts, we prepared an equivalent number of comments that merely described what the fictional person had already done in the cooking task (evaluative; e.g. "To prepare the sauce, you have used the cheese grater to grate the whole block of cheese") and what they would need to do afterwards (directive; e.g., "To prepare the sauce, you will use the cheese grater to grate the whole block of cheese"). As the examples here illustrate, like in the previous experiments we manipulated the style of the comments in ways that maintained their general meaning while approximately equating their length and complexity.

Using these comments we prepared four scripts, each of which contained 18 comments, split between an evaluative style and a directive style and with the comments presented in pairs that alternated between the evaluative and directive style throughout (Feedback Script A = 291 words; Feedback Script B = 287 words; Description Script A = 293 words; Description

Script B = 294 words; the scripts are available in the online supplemental materials). Due to an oversight, the number of evaluative and directive comments in each script was not exactly equal – the Version A's contained 10 evaluative and 8 directive comments, whereas the Version B's contained 8 evaluative and 10 directive comments. As a further deviation from our pre-registered plan then, we therefore calculated each dependent variable as a proportion of the total number of comments received, rather than as a raw frequency.

Procedure. As in Experiments 2A and 2B, participants completed the study in a single online session where they were told they would read some advice that had been written for another person. Participants were randomly assigned to one of the two narrative format conditions—either the Feedback condition or the Description condition—and were randomly assigned to see one of the two counterbalanced script variants for their assigned condition. They were asked to read this script carefully and to click to confirm once they were ready to proceed.

Next participants moved directly to the unexpected recall test – they did not complete a filler task. All participants were asked to recall as much of the text about cookery they had read a moment before. They were required to spend at least 5 min on this recall task, but otherwise had no time limit.

Memory coding. One researcher coded all responses blind to which script version each participant had seen, although they were aware of which narrative format condition each participant was assigned. A second

researcher coded 20% of responses in the same way. The agreement was good for both measurements (total evaluative comments recalled, r = .74; total directive comments recalled, r = .92). Therefore, all analyses are based on the first researcher's coding.

Results and Discussion

We conducted our pre-registered analysis: a 2 (narrative format: feedback vs. description) x 2 (comment type: evaluative vs. directive) x 2 (retrieval style accuracy: correct vs. incorrect) mixed-factor ANOVA with repeated measures on the latter two factors. Combined across narrative format conditions, the overall main effect of comment type was significant, $F(1, 130) = 7.03$, $p < .01$, $\eta^2_p = .05$, $d = 0.23$, $M_{diff} = 0.05$, 95% CI [0.01, 0.08] with a greater proportion of the evaluative comments recalled than of the directive comments. There was also a main effect of narrative format, with participants in the Description condition recalling more than those in the Feedback condition, $F(1, 130) = 6.52$, $p = .01$, $\eta^2_p = .05$, $d = 0.22$, $M_{diff} = 0.08$, 95% CI [0.02, 0.15]. Figure 4 shows that the pattern of data mirrored our predictions, with a notably smaller evaluative recall bias in the Description condition than in the Feedback condition. However, this predicted two-way interaction was not statistically significant, $F(1, 130) = 3.11$, $p = .08$, $\eta^2_p = .02$. In sum then, the Experiment 4 data provide additional evidence in support of an evaluative recall bias, even with these very different feedback materials. However, whereas we had predicted that this bias would only occur in the Feedback

condition, the suggestive evidence in support of this specific prediction was not conclusive.

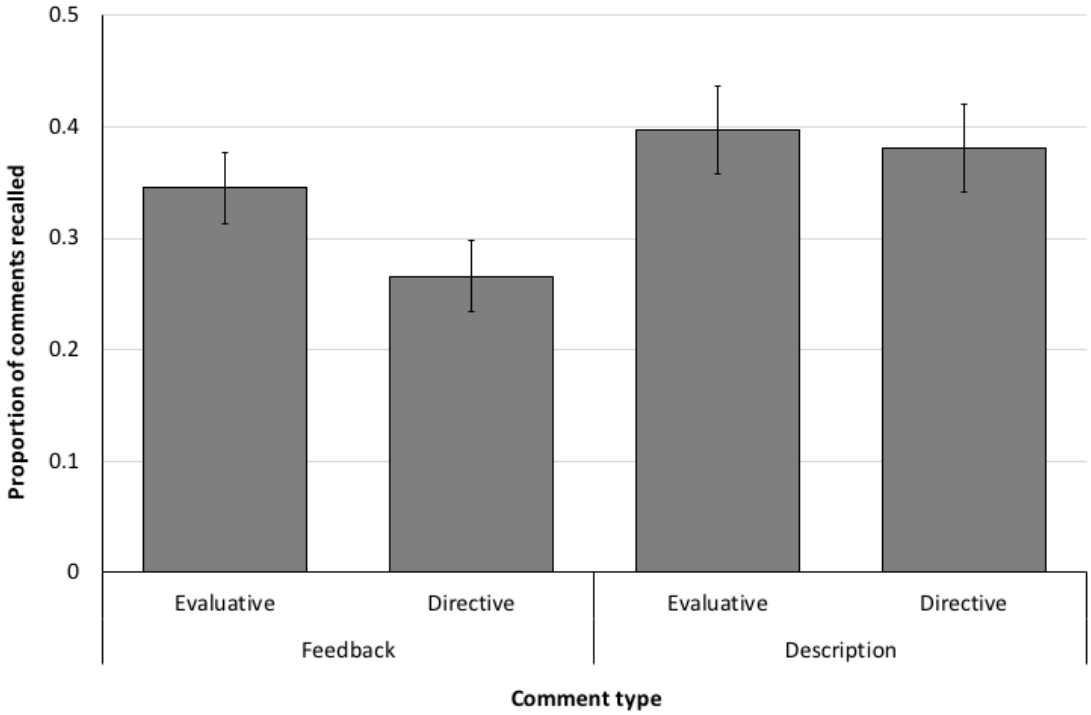

Figure 4. Proportion of evaluative and directive comments recalled by participants in Experiment 4, according to narrative format. Data are collapsed across retrieval style accuracy. Error bars are 95% within-participant confidence intervals calculated separately for each narrative format condition (Loftus & Masson, 1994).

General Discussion

Prior research demonstrates that when people receive feedback from others, they tend to preferentially remember past-oriented (evaluative) comments over future-oriented (directive) comments. All five experiments in this paper obtained further evidence of this evaluative recall bias, but one of the paper's key goals was to continue efforts to identify the cognitive mechanism that underpins it.

In our first three experiments we tested the hypothesis that although people successfully encode directive feedback in memory equivalently to evaluative feedback, they neglect to access directive feedback during retrieval. To this end, we asked whether people would be better able to retrieve directive feedback if explicitly prompted to do so via a structured recall task. However, consistently across Experiments 1, 2A, and 2B the evaluative recall bias persisted even in a structured recall test. These experiments therefore provide evidence against the selective neglect hypothesis.

In Experiment 3 we tested the hypothesis that participants who took the perspective of a teacher would recall the directive aspects of the feedback better than would those taking the perspective of a student. However, we found that they were just as likely to exhibit an evaluative recall bias when taking the perspective of the teacher, as when taking the perspective of the student. The findings of this experiment are limited by the absence of a satisfactory manipulation check that validated participants' engagement with the perspective-taking. Nonetheless the data fit with our broader conclusion from Experiments 1-2, insofar that efforts to influence how people accessed their memories were unsuccessful.

Finally in Experiment 4 we tested whether the findings were specific to the domain of feedback, by manipulating whether the text was written either in a way that could be interpreted as feedback or written in a descriptive manner, yet still containing the fundamental directive vs. evaluative

elements. We found suggestive but nonsignificant evidence that the evaluative recall bias may be specific to the context of feedback, and was less evident when participants read and recalled a non-judgmental, descriptive account of a person's actions. We should be cautious in drawing strong conclusions from these data, but this avenue of inquiry warrants further replication and investigation since it leaves open the possibility that schematic representations might play a role in this memory bias. In conceptual replications it would be important to ensure that readers judge the 'feedback' and 'description' comments to convey similar degrees of evaluative information. Doing so would help to rule out the possibility that the apparently smaller effect in the Description condition was due merely to a weaker manipulation of written style.

The findings of Experiment 4 extend those of Experiments 1-3, and indeed those of Nash et al. (2018) and Gregory et al. (2020) in at least two other important ways. First, due to the specific manipulation in Experiment 4, we avoided using the word "feedback" in our task instructions, opting instead to describe both the feedback scripts and the description scripts as "advice". These choices of linguistic framing can have sizeable effects on the kinds of information that assessors provide to learners (Yoon et al., 2019), and one might predict too that the word "feedback"—by virtue of being past-oriented—could similarly account for the evaluative recall bias in memory. The fact that we were able to replicate this memory bias without priming

participants with the word "feedback", though, provides some evidence against this explanation.

Second, by using cookery scripts, Experiment 4 is the first to demonstrate this evaluative recall bias in the recall of kinds of "feedback" that are unrelated to educational performance. This is an important finding because it indicates that the effects of these biases—if they affect behavior outside of laboratory contexts—could be relevant not just to the classroom but also to other training contexts and professions. Researchers and practitioners from other contexts involving feedback, such as business or sports coaching, might therefore explore the robustness and behavioral consequences of this memory bias in their own specialist contexts. Indeed, one of the key gaps in our understanding of this effect remains the extent to which it influences people's subsequent performance, and which advice they actually act upon.

A clear limitation of the present work is its artificiality: our participants were unlikely to be engaging with the feedback in a way that mirrors realistic learning and development contexts. Future work that tackles this limitation, and thereby enhances the applied generalizability of these findings, would be invaluable for this reason. For example, one question of relevance to applied contexts is how people's memory for evaluative and directive feedback might interact with other contextual performance information, such as high vs. low grades. However, despite the artificiality of our methods, it is interesting to note that the size of the evaluative recall bias was substantially larger in Experiment 1—where participants were told the feedback was personalized—

than in Experiments 2-4 where participants were told the feedback related to another student. This finding raises the prospect that the effect could remain, and potentially be bigger in more personally relevant contexts. Nash et al. (2018) previously dismissed several self-oriented accounts of this memory bias, on the basis that the bias survives even in non-self-relevant paradigms. Nevertheless we should consider the possibility that self-motives could contribute to the evaluative recall bias even if they are not the only mechanism. Future studies might, for example, randomly assign participants to be told they are receiving feedback on their own work vs. another student's work. The memory effects of this manipulation would be particularly interesting in light of recent research that found people were less likely to learn from "failure feedback" than from equally informative "success feedback", but only when the feedback was self-relevant (Eskreis-Winkler & Fishbach, 2019). This different bias—which the authors found to also influence participants' memory reports—seems intuitively difficult to reconcile with the evaluative recall bias. Research paradigms that allowed both biases to be examined together, perhaps through comparing people's memory for evaluative vs. directive praise, as well as critique, might shed further light on key mechanisms.

In sum, these experiments provide further evidence for the evaluative recall bias. The robustness of this counterintuitive memory bias might give pause for thought to feedback-givers who have been convinced of the superiority of future-oriented, directive feedback. Nevertheless, the most

effective response to this bias in practice is unlikely to be to simply avoid giving directive feedback; identifying the underpinning mechanism(s) will be key to knowing how to ensure the memorability and impact of future-oriented comments. The present experiments offer little evidence that selective retrieval mechanisms underpin this bias; however, they do add to the list of putative mechanisms that have been disputed by empirical data.

References

Anderson, R. C., & Pichert, J. W. (1978). Recall of previously unrecallable information following a shift in perspective. *Journal of Verbal Learning and Verbal Behavior*, *17*, 1-12.

Butler, A. C., & Roediger, H. L. (2008). Feedback enhances the positive effects and reduces the negative effects of multiple-choice testing. *Memory & Cognition, 36*, 604–616.

Douglas, K. M., & Skipper, Y. (2012). Subtle linguistic variation in feedback. In R. M. Sutton & K. Douglas (Eds.), *Feedback: The communication of praise, criticism and advice* (pp. 73-88). Peter Lang.

Eitam, B., Schul, Y., & Hassin, R. R. (2009) Goal relevance and artificial grammar learning, *Quarterly Journal of Experimental Psychology*, *62*, 228–238.

Eskreis-Winkler, L., & Fishbach, A. (2019). Not learning from failure-the greatest failure of all. *Psychological Science, 30*, 1733-1744.

Fisher, R. P., Milne, R., & Bull, R. (2011). Interviewing cooperative witnesses. *Current Directions in Psychological Science*, *20*, 16-19.

Gilbert, J. A., & Fisher, R. P. (2006). The effects of varied retrieval cues on reminiscence in eyewitness memory. *Applied Cognitive Psychology, 20*, 723-739.

Ginet, M., Dodier, O., Bardin, B., Désert, M., Greffeuille, C., & Verkampt, F. (2018). Perspective effects on recall in a testimony paradigm. *Journal of General Psychology*, *145*, 313-341.

Goldsmith, M. (2010). Leadership coaching with feedforward. In J. Passmore (Ed.), *Leadership coaching* (pp.177-187). Kogan Page.

Gregory, S. E. A., Winstone, N. E., Ridout, N., & Nash, R. A. (2020). Weak memory for future-oriented feedback: investigating the roles of attention and improvement focus. *Memory*, *28*, 216-236.

Hirsch, J. (2017). *The feedback fix: Dump the past, embrace the future, and lead the way to change*. Lanham, MD :Rowman & Littlefield.

Irwin, B., Hepplestone, S., Holden, G., Parkin, H. J., & Thorpe, L. (2013). Engaging students with feedback through adaptive release. *Innovations in Education and Teaching International*, *50*, 51-61.

Kaakinen, J. K., Hyönä, J., & Viljanen, M. (2011). Influence of a psychological perspective on scene viewing and memory for scenes. *Quarterly Journal of Experimental Psychology*, *64*, 1372-1387.

Loftus, G. R., & Masson, M. E. (1994). Using confidence intervals in within-subject designs. *Psychonomic Bulletin & Review*, *1*, 476-490.

Mensink, P. J., & King, K. (2020). Student access of online feedback is modified by the availability of assessment marks, gender and academic performance. *British Journal of Educational Technology, 51,* 10–22.

Montagrin, A., Brosch, T., & Sander, D. (2013) Goal conduciveness as a key determinant of memory facilitation. *Emotion*, *13*, 622–628.

Nash, R. A., & Winstone, N. E. (2017). Responsibility-sharing in the giving and receiving of assessment feedback. *Frontiers in Psychology*, *8*, 1519.

Nash, R. A., Winstone, N. E., Gregory, S. E. A., & Papps, E. (2018). A memory advantage for past-oriented over future-oriented performance feedback. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*, 1864-1879.

Robinson, S., Pope, D., & Holyoak, L. (2013). Can we meet their expectations? Experiences and perceptions of feedback in first year undergraduate students. *Assessment and Evaluation in Higher Education*, *38*, 260-272.

Rogers, T. B., Kuiper, N. A., & Kirker, W. S. (1977). Self-reference and the encoding of personal information. *Journal of Personality and Social Psychology*, *35*, 677-688.

Sweeny, K., Melnyk, D., Miller, W., & Shepperd, J. A. (2010). Information avoidance: Who, what, when, and why. *Review of General Psychology*, *14*, 340-353.

Surber, J. R. (1983). The influence of decision factors on what is reported in free recall of a brief narrative. *Contemporary Educational Psychology*, *8*, 119-126.

Winstone, N. E., Nash, R. A., Rowntree, J., & Menezes, R. (2016). What do students want most from written feedback information? Distinguishing necessities from luxuries using a budgeting methodology. *Assessment and Evaluation in Higher Education, 41*, 1237–1253.

Yoon, J., Blunden, H., Kristal, A., & Whillans, A. (2019). *Framing feedback giving as advice giving yields more critical and actionable input*. Unpublished manuscript retrieved from

https://www.hbs.edu/faculty/Publication%20Files/20-021_b907e614-e44a-4f21-bae8-e4a722babb25.pdf