


Utility of Feedback Has a Greater Impact on Learning than Ease of Decoding

Chelsea Dainton¹ , Naomi Winstone², Peter Klaver^{1,3}, and Bertram Opitz^{1,4}

ABSTRACT— While feedback is a key facilitator of learning, researchers have yet to determine the ideal feedback process for optimal performance in learners. The current study investigates the combined effects of ease of decoding, and utility of feedback during learning. Accuracy and rate of learning were recorded alongside changes to the feedback related negativity (FRN), an event-related potential (ERP) elicited by feedback stimuli. This study investigates the FRN within the context of future-focused directive feedback (DF), in addition to past-focused evaluative feedback (EF) typically seen in the neuroscience literature. Results indicate a main effect of utility together with an interaction with ease of decoding on the accuracy data, but only the main effect of utility on learning rate. DF produced an FRN, like EF, which was then larger during high-utility feedback, specifically following negative EF or when hard-to-decode. Implications and future research directions are discussed.

Feedback is a crucial facilitator of learning which is considered one of the principal means for an individual to effect learning gains (Hattie & Timperley, 2007). By closing the gap between current and desired performance, feedback serves to reduce the uncertainty a learner may experience regarding their proficiency (Shute, 2008). The effectivity of feedback depends upon students' ability to understand (decode) the message conveyed and apply (utilize) it in future work (Jonsson & Panadero, 2018). Studies have found that feedback which is easy-to-decode, or provides highly

applicable information, improves performance (Butler, Godbole, & Marsh, 2013; Evans, 2013), facilitates engagement (Winstone, Nash, Parker, & Rowntree, 2016; Winstone, Nash, Rowntree, & Parker, 2017) and is favored by students (Winstone, Nash, Rowntree, & Menezes, 2016).

In contrast, feedback messages that are difficult to understand can lead to frustration and are perceived to be less effective in facilitating learning (e.g. Jonsson, 2013). This frustration can induce a state of learned helplessness (Teodorescu & Erev, 2014), whereby students become disenfranchised with their feedback due to its perceived difficulty. As well as adding to cognitive load, feedback that is hard-to-decode makes a limited contribution to reducing learners uncertainty (Shute, 2008).

In terms of feedback utility, cognitive models also inform our understanding of the likely effectiveness of different types of feedback. Kulhavy and Stock (1989) distinguished between the functions of verification and elaboration in feedback. Verification feedback informs the learner whether their response was incorrect or correct, without further explanation. In contrast, elaborated feedback provides guidance, for example indicating why a particular response was correct, or providing recommendations to improve performance. In a comprehensive review of formative feedback literature, Shute (2008) concludes that verification feedback in isolation has limited utility for learning, whereas elaborative feedback supports learners to self-correct and advance their understanding. Different types of elaborative feedback exist on a spectrum from lower to higher utility, depending on the degree of explanation and direction it provides.

Extending this cognitive perspective, there is robust literature on the outcome measures of feedback, particularly in neural responses to feedback processing. The feedback related negativity (FRN) signal; an event-related potential (ERP) elicited 200–400 ms after receiving feedback (Nieuwenhuis, Holroyd, Mol, & Coles, 2004) shows sensitivity to multiple aspects of feedback processing, including feedback valence (Holroyd & Coles, 2002;

¹School of Psychology, University of Surrey

²Department of Higher Education, University of Surrey

³Applied University of Special Needs Education

⁴Centre for Digital Learning and Teaching, Martin Luther University Halle-Wittenberg

Address correspondence to Chelsea Dainton, School of Psychology, University of Surrey, Surrey, GU2 7XH, UK; e-mail: c.m.dainton@surrey.ac.uk

Nieuwenhuis et al., 2004), magnitude (Holroyd, Larsen, & Cohen, 2004; San Martín, Manes, Hurtado, Isla, & Ibañez, 2010), probability (Holroyd & Coles, 2002; Nieuwenhuis et al., 2004), and informative value (Du, Cao, He, & Li, 2018).

However, the FRN has only been recorded using feedback regarding past performance, termed evaluative feedback (EF) in educational sciences. EF is similar in function to verification feedback. In contrast, directive feedback (DF), a form of elaborative feedback, focuses on future performance, which can vary in ease of decoding and utility, but typically provides greater utility than EF (Carless, 2006; Hattie & Timperley, 2007). The FRN is typically regarded as measuring the amount, or complexity of, feedback information (Du et al., 2018). Therefore, how easy feedback is to decode, and how easy it is to utilize, are expected to influence this signal, with both hard-to-decode and high-utility feedback engaging more cognitive processes than their counterparts.

The present study aims to test whether utility and ease of decoding of DF affected learning and neural mechanisms of feedback processing. We hypothesized that:

- 1 Easy-to-decode feedback better facilitates learning than hard-to-decode feedback;
- 2 High-utility feedback better facilitates learning than low-utility feedback;
- 3 DF will elicit an FRN signal like EF;
- 4 The peak-to-peak FRN will be larger for high-utility, and hard-to-decode feedback, due to greater cognitive demands.
- 5 Valence of EF will influence the FRN produced by subsequent DF due to differential levels of informative value.

METHODS

Participants

Data were collected from 27 volunteers recruited through social media snowballing and through a participant recruitment system (SONA). One participant's data were excluded because of chance performance in the learning tasks. The final sample consisted of 26 students (20 female, aged 18–26, $M = 20.65$ years), who reported normal or corrected-to-normal vision, no contraindications of EEG, and no color-blindness. Participants gave written informed consent before the experiment. Undergraduate psychology students ($N = 21$) received a course incentive in exchange for participation, while other participants received no compensation.

Materials

The task was an adaptation of a category learning task (Knowlton, Squire, & Gluck, 1994). Participants aimed to ascertain which flavor ice cream various potato head figures

(henceforth called “the character”) preferred. This setup allowed for the variation of three different features (hat, eyes or shoes), with three variations each (e.g., black, white or blue shoes) that were systematically varied to create the 27 character images used in this experiment. The different variations of a single feature dictated which flavor ice cream (vanilla, chocolate, or strawberry) the character would prefer. For example, when the shoe color indicated preferred flavor, each color shoe directly mapped onto one ice cream flavor. Participants were asked to firstly deduce which feature, and then which variation of said feature, predicted the preferred ice cream flavor based on the EF and DF they received.

EF consisted of a stylized frowning red face for incorrect responses, a green stylized smiling face for correct responses, and the words “No response” if the participant did not respond within the time given.

DF consisted of a cartoon hat, eyes and shoes on a white background (see Figure 1). Colored bars were used to construct the various forms of DF. High-utility screens contained a green bar behind the feature relevant for flavor preference, while low-utility screens contained a red bar behind an irrelevant feature. The easy-to-decode screens included just a singular green or red colored bar, while the hard-to-decode screens included two other colored bars to increase perceptual load (Lavie, Hirst, De Fockert, & Viding, 2004). Therefore, the easy-to-decode with high-utility (EH) condition, contained a single green bar, whereas the easy-to-decode with low-utility (EL) condition contained a single red bar. The hard-to-decode with high-utility (HH) condition contained purple and pink bars alongside the green bar, while the hard-to-decode with low-utility (HL) condition displayed a red bar among a yellow and blue bar. The noninformative bar colors were chosen to match the informative bar in terms of brightness and intensity, while being distinct from red or green.

Procedure

Trials began with a crosshair for 500 ms followed by the character stimulus for the following 2,500 ms. Participants were asked to respond during this 2,500 ms, after which they were presented with the EF screen for 500 ms, and finally the DF screen for 2,000 ms. Instructions required participants to respond with “1,” “2” or “3” on the number pad of keyboard to designate the preferred ice cream flavor. The correct response key was counterbalanced across participants. On-screen instructions informed participants that they would receive two feedback screens; one informing them whether their choice was correct (EF), and one that would help them uncover the relevant feature for that block (DF). Instructions noted that green feedback bars indicated a relevant feature, while red bars highlighted an irrelevant feature, but there could be other, noninformative, colors present

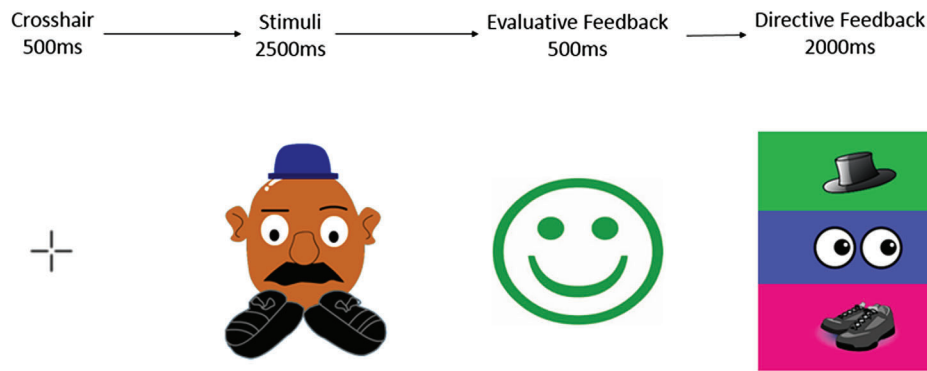


Fig. 1. A sample trial where the hat is the relevant feature, displaying positive evaluative feedback and the hard to decode with high utility directive feedback screen.

alongside these cues. After finishing these instructions participants were shown a diagram of a trial (Figure 1) including EH DF to familiarize them with the stimuli.

The experiment was presented using E-Prime 2.0 software (Psychology Software Tools, Pittsburgh, PA) and consisted of three runs of four blocks, one per DF conditions (EH, EL, HH, HL), with a self-directed break after each block. Each block contained 27 trials, each presenting a different character. Each flavor was correct in nine trials per block, resulting in a 33% probability of guessing. The first run was presented in a fixed order of supposedly increasing difficulty: EH, EL, HH, HL, but then proceeded in a pseudorandom manner, with each condition presented once before the first condition was repeated. The experimental task took an average of 30 min to complete and the whole study, including EEG set up and debriefing, was not longer than 2 hr. Participant accuracy was recorded for each trial to comprise the behavioral data in the analysis. Learning curves were created after the study using a rolling three-point average. These values were then fitted by an exponential rise to maximum model to estimate rate of learning. See Supporting Information for model formulae.

EEG Recording and Preprocessing

EEG data were recorded from 30 Ag/AgCl electrodes mounted on an elastic cap (EasyCap) and positioned according to the extended international 10–20 system, with online reference at FCz and ground at FPz. Data were amplified from DC to 70 Hz at a sampling rate of 500 Hz using a BrainAmp MR (Brain Products GmbH, Germany). To control for eye movements, the EOG was recorded from the suborbital ridge of the left eye. Impedance of all electrodes was kept below 10 k Ω . Data were processed offline using Brain Vision Analyser. After, data were referenced to linked mastoids so that FCz could be analyzed, a band pass filter of 0.5 Hz to 30 Hz was applied. Artifacts resulting from eye blinks were removed using an automatic ocular correction

independent components analysis (Jung et al., 2000) and remaining artifacts were automatically excluded from further analyses using a 50 μ V/ms gradient and a 100 μ V/200 ms maximal amplitude change criterion. Data were segmented from 100 before to 700 ms after feedback (EF and DF) onset. All segments were baseline corrected from –100 to 0 ms before averaging. FRN was quantified for each participant as peak-to-peak amplitude by calculating the difference between the most negative amplitude from 230 to 330 ms post feedback presentation and the amplitude of the preceding positive peak in a time window from 170 to 230 ms (Ferdinand, 2019; Holroyd et al., 2004; Nieuwenhuis et al., 2004). Each dataset was manually checked for accuracy of peak detection before data were extracted for all feedback conditions separately. For all conditions and participants the peaks were accurately detected in the respective time windows. After artifact extraction an average of 37.80 (SD 3.10) trials remained in each condition, an appropriate number for FRN analysis (Marco-Pallares, Cucurell, Münte, Strien, & Rodriguez-Fornells, 2011).

RESULTS

Behavioral Results

Overall accuracy and learning rates are depicted in Table 1. Accuracy was significantly above chance level for each condition (EH $t(26) = 26.60, p < .001$; EL $t(26) = 10.09, p < .001$; HH $t(26) = 16.34, p < .001$; HL $t(26) = 11.89, p < .001$), indicating that participants were able to deduce the feature required to select the correct response across each condition. A 2×2 repeated-measures analysis of variance (ANOVA) was used to investigate the effects of decoding and utility on accuracy. Utility yielded a significant main effect ($F(1, 25) = 35.00, p < .001, \eta_p^2 = .59$), but not decoding ($F(1, 25) = 2.06, p = .16, \eta_p^2 = .08$), with participants performing more accurately on high-utility blocks. The decoding by utility interaction was significant ($F(1, 25) = 22.52, p < .001$,

Table 1

Descriptive Statistics of the Overall Accuracy Data, Accuracy for Final Eight Blocks, and Learning Rate (Gradient of Learning Curve) for Each Directive Feedback Condition

	<i>Easy decode (high utility)</i>	<i>Easy decode (low utility)</i>	<i>Hard decode (high utility)</i>	<i>Hard decode (low utility)</i>
Mean overall accuracy (%)	81.77	63.85	78.08	73.41
Standard deviation	0.10	0.16	0.14	0.18
Mean final eight blocks accuracy (%)	85.96	70.37	83.25	75.98
Standard deviation	0.08	0.18	0.14	0.19
Mean learning rate	.36	.18	.35	.23
Standard deviation	0.21	0.12	0.14	0.13

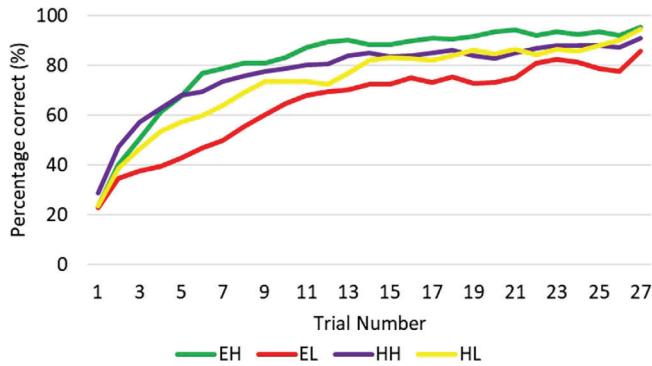


Fig. 2. The average percentage of correct responses on each trial separated by directive feedback condition.

$\eta_p^2 = .47$). Post hoc *t*-tests revealed a significant difference between the EL and EH conditions ($p < .001$, $d = 1.33$), but not between the HL and HH conditions ($p = .11$, $d = 0.29$).

Next, learning curves were analyzed to assess within-block learning (see Table 1 and Figure 2). A 2×2 repeated measures ANOVA revealed a significant effect of utility on learning rate ($F(1, 25) = 26.98$, $p < .001$, $\eta_p^2 = .52$), with high-utility feedback eliciting steeper learning curves. Neither an effect of decoding ($F(1, 25) = .24$, $p = .63$, $\eta_p^2 = .01$), nor an interaction between the two factors ($F(1, 25) = .56$, $p = .46$, $\eta_p^2 = .02$) were observed.

Finally, the first run was presented in a fixed order, while the remaining eight blocks were presented in a pseudo-randomized order. To accommodate for these effects the accuracy data were re-analyzed after excluding the first run. Utility ($F(1, 25) = 31.59$, $p < .001$, $\eta_p^2 = .56$) and the interaction between utility and decoding ($F(1, 25) = 8.09$, $p = .009$, $\eta_p^2 = .24$) remained significant, while decoding remained nonsignificant ($F(1, 25) = 0.57$, $p = .46$, $\eta_p^2 = 0.02$). Hence, the effects of utility and the interaction of both factors on accuracy were independent of order effects as the direction of the effect persisted even when the first four blocks were removed.

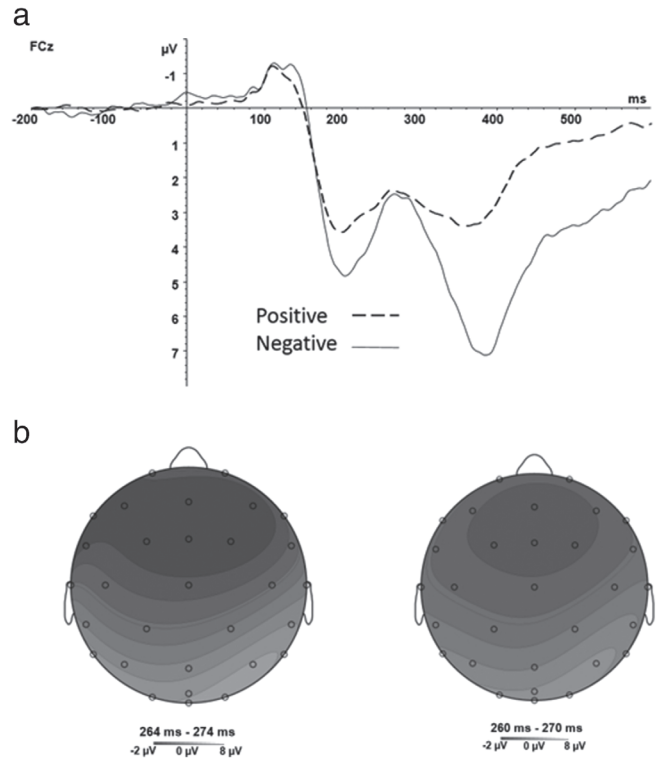


Fig. 3. (a) Feedback-related negativity signal (260-280 ms) to positive (dashed) and negative (solid) evaluative feedback recorded at the FCz electrode. (b) Topography for the Feedback-related negativity signal baseline corrected to the peak of the P2, for positive (left) and negative (right) evaluative feedback.

EEG Results

EF was analyzed to replicate standard findings within the literature and ensure we had a valid comparison for the FRN elicited by DF. The peak-to-peak FRN amplitude at FCz for negative EF ($M = 3.50$, $SD = 2.07$; see Figure 3) was larger than the FRN to positive EF ($M = 2.15$, $SD = 1.59$), as confirmed by a two-way repeated measures ANOVA ($F(1, 24) = 27.94$, $p < .001$, $\eta_p^2 = .40$).

Within the DF condition an FRN signal was found at FCz between 200 and 400 ms after stimuli onset with a

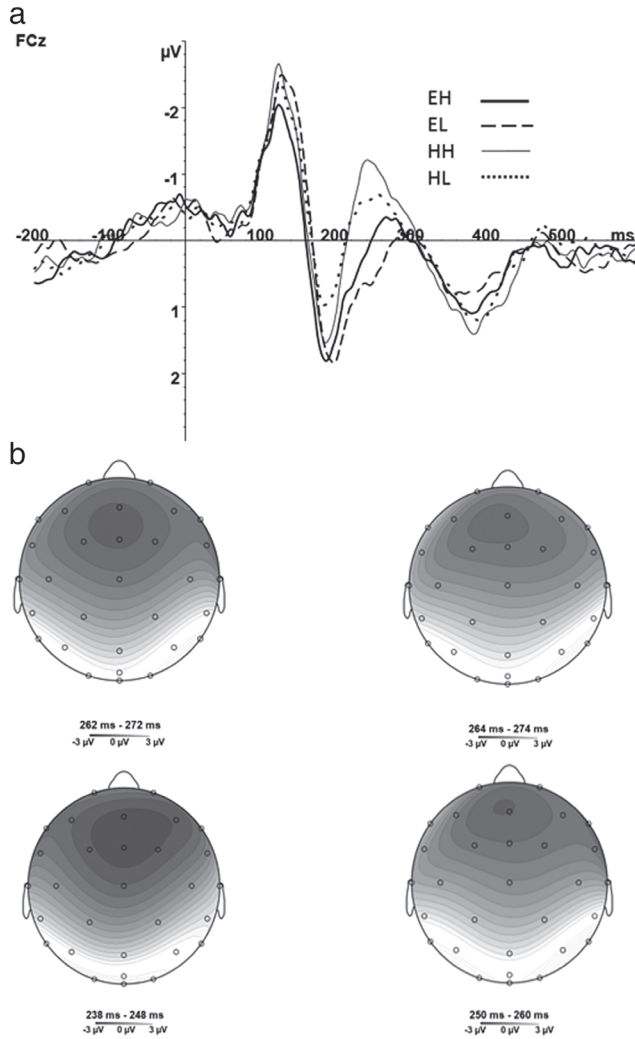


Fig. 4. (a) Feedback-related negativity signal (240–290 ms) to four directive feedback conditions at the FCz electrode: easy to decode with high utility (EH), easy to decode with low utility (EL), hard to decode with high utility (HH), and hard to decode with low utility (HL). (b) Topography of Feedback-related negativity signal, baseline corrected to the peak of the P2, for the EH (top left), EL (top right), HH (bottom left), and HL conditions (bottom right).

mid-frontal distribution in all four DF conditions (See Figure 4). All conditions were compared to investigate the effect of utility and ease of decoding on the FRN. Furthermore, the valence of preceding EF was included to inspect whether this also influences the FRN elicited by DF (see Table 2 for descriptives). A $2 \times 2 \times 2$ repeated-measures ANOVA revealed a significant main effect of valence ($F(1, 24) = 22.35, p < .001, \eta_p^2 = .48$), and utility ($F(1, 24) = 11.52, p = .002, \eta_p^2 = .32$), with higher utility and negative valence eliciting larger amplitudes, while the effect of decoding was not significant ($F(1, 24) = .94, p = .34, \eta_p^2 = .04$). Significant interactions were found between valence and utility ($F(1, 24) = 12.28, p = .002, \eta_p^2 = .34$), and decoding and utility

($F(1, 24) = 5.45, p = .028, \eta_p^2 = .19$), but not between valence and decoding ($F(1, 24) = 0.08, p = .78, \eta_p^2 = .003$), nor for a three way interaction ($F(1, 24) = 0.07, p = .79, \eta_p^2 = .003$).

Follow-up Wilcoxon tests revealed that there was a significant effect of utility on hard-to-decode trials ($z = 2.46, p = .014, r = .35$), and those following negative feedback ($z = 3.00, p = .003, r = .42$), but not those that were easy-to-decode ($z = 0.23, p = .82, r = .03$) or following positive feedback ($z = 1.39, p = .17, r = .20$). Peak-to-peak difference scores were higher for high-utility feedback in both significant comparisons.

DISCUSSION

The current study aimed to investigate how the utility and ease of decoding of DF affected learning and neural mechanisms of feedback processing. In agreement with previous findings (Butler et al., 2013), high-utility feedback consistently enhanced the overall accuracy and the speed of learning. However, contrary to our predictions, ease of decoding did not. Nonetheless, both factors did interact within the accuracy data, but not in the learning rate data. This suggests that there is a large utility effect for easy-to-decode information, but when information is hard-to-decode, utility is less pertinent to learning. This interaction indicates that decoding could act as a “gatekeeper” for the effects of utility, rather than exert an influence on learning on its own grounds. These findings could be interpreted in line with the attentional load theory (Lavie et al., 2004). According to this theory high perceptual load in the hard-to-decode conditions would limit the attentional processing of irrelevant information, consequently reducing fixation on low-utility information.

Therefore, our findings support calls for feedback to be both easy-to-understand, and highly applicable (Shute, 2008; Winstone, Nash, Parker, & Rowntree, 2016; Winstone, Nash, Rowntree, & Menezes, 2016), as this seems to be the optimum combination for learning speed and outcome. Hard-to-decode feedback reduced differentiation between high and low utility feedback on learning outcome. Such feedback could be associated with an increase in learned helplessness (Teodorescu & Erev, 2014), and a barrier to engagement (Jonsson, 2013; Winstone et al., 2017). By contrast, easy-to-decode with low-utility feedback yielded low accuracy and learning rate, suggesting that with clear but noninformative meaning, students may disengage as they are unsure of how to improve in future.

The present study also examined the electrophysiological correlates of processing DF contingent on EF. Replicating previous findings (Holroyd & Coles, 2002; Nieuwenhuis et al., 2004), negative EF elicited a larger amplitude FRN than positive EF. Furthermore, larger FRN to DF and especially to

Table 2

Mean and Standard Deviation of Peak-to-Peak Difference of the Feedback Related Negativity Signal at FCz Across all Directive Feedback Conditions Following Either Positive or Negative Evaluative Feedback

	<i>Easy decode (high utility)</i>	<i>Easy decode (low utility)</i>	<i>Hard decode (high utility)</i>	<i>Hard decode (low utility)</i>
Positive feedback	3.31 (2.74)	3.41 (2.32)	3.96 (2.82)	3.26 (2.40)
Negative feedback	4.83 (2.40)	3.85 (2.56)	5.66 (3.35)	3.70 (2.37)

high utility feedback after negative compared to positive EF suggests that DF is more informative after an unexpected and unsatisfying result. It also suggests that participants approach the current task in a two staged process in a hierarchical way. Participants first determined the relevant feature using the DF, before matching the variations of that feature to the ice cream preference. In case of negative EF participants seem to engage more with DF to reassure they picked the correct feature for their decision. Thus, high utility DF elicited a FRN signal, which indicates that reinforcement of the relevant feature is a mandatory first step to solve the task. Thus, the information view of the current results is in line with a recent hierarchical reinforcement learning theory (Holroyd & Yeung, 2012) stating that the FRN reflects the reinforcement of higher-level behavioral plans and the sequences of specific behaviors directed toward these plans.

One potential limitation of the methodology is whether the perceptual load (Lavie et al., 2004) used in the current study to simulate the level of decoding difficulty is comparable to the difficulty of decoding complex academic terms in educational settings. If the current operationalization was not an accurate proxy for decoding complex feedback, it could explain why previous reviews indicate that this factor affects learning (Jonsson & Panadero, 2018). Alternatively, the effects of decoding on learning may have been confounded in previous studies by the utility present in the feedback provided. Further studies, comparing different types of feedback, such as common feedback terms or phrases, would address this question. Despite a significant interaction of decoding and utility, the effect size of the main effect of decoding is rather small. Therefore, a replication is required to examine whether decoding has no individual impact upon learning measures.

Finally, future studies should test whether cognitive factors known to influence the FRN elicited by EF, such as perceived fairness (Riepl, Mussel, Osinsky, & Hewig, 2016), have similar effects upon the FRN elicited by DF in this study. Similarly, more in-depth analysis of the neural signature could reveal further similarities or differences between processes underlying EF and DF. We also did not analyze any ERP's beyond the FRN. Future research should consider the P300, a signal also sensitive to various aspects of feedback (San Martín, 2012), such as the expectancy of

feedback (Ferdinand, Becker, Kray, & Gehring, 2016). The current study was unable to investigate this aspect as all DF conditions were identical in terms of probability.

In conclusion, the present study demonstrated that while utility produced a significant main effect on both learning and FRN amplitude, decoding acted as a “gatekeeper” for the effect of utility for accuracy, while both valence and decoding of feedback acted as “gatekeeper” for the effect of utility on the FRN amplitude. To our knowledge, this has been the first experimental study to disentangle the simultaneous effects of ease of decoding, and utility of feedback on learning, alongside uncovering the neural signals associated with DF, and providing a means for future studies to consider the nuances of DF.

Acknowledgments—We would like to acknowledge our gratitude to both the University of Surrey for funding the PhD of the first author, and Dr. Philip Dean for his technical assistance with data collection. BO was partly supported by a Grant of the German Federal Ministry of Research and Education, No. 01PL17065, Quality Pact for Teaching.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article. Data S1. Supporting Information.

CONFLICT OF INTEREST STATEMENT

No conflicts of interested noted.

REFERENCES

- Butler, A. C., Godbole, N., & Marsh, E. J. (2013). Explanation feedback is better than correct answer feedback for promoting transfer of learning. *Journal of Educational Psychology, 105*(2), 290–298.
- Carless, D. (2006). Differing perceptions in the feedback process. *Studies in Higher Education, 31*(2), 219–233.
- Du, B., Cao, B., He, W., & Li, F. (2018). Different electrophysiological responses to informative value of feedback between children and adults. *Frontiers in Psychology, 9*, 346.

- Evans, C. (2013). Making sense of assessment feedback in higher education. *Review of Educational Research*, 83(1), 70–120.
- Ferdinand, N. K. (2019). The influence of task complexity and information value on feedback processing in younger and older adults: No evidence for a positivity bias during feedback-induced learning in older adults. *Brain Research*, 1717, 74–85.
- Ferdinand, N. K., Becker, A. M. W., Kray, J., & Gehring, W. J. (2016). Feedback processing in children and adolescents: Is there a sensitivity for processing rewarding feedback? *Neuropsychologia*, 82, 31–38.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81–112.
- Holroyd, C. B., & Coles, M. G. H. (2002). The neural basis of human error processing: Reinforcement learning, dopamine, and the error-related negativity. *Psychological Review*, 109(4), 679–709.
- Holroyd, C. B., Larsen, J. T., & Cohen, J. D. (2004). Context dependence of the event-related brain potential associated with reward and punishment. *Psychophysiology*, 41(2), 245–253.
- Holroyd, C. B., & Yeung, N. (2012). Motivation of extended behaviors by anterior cingulate cortex. *Trends in Cognitive Sciences*, 16(2), 122–128.
- Jonsson, A. (2013). Facilitating productive use of feedback in higher education. *Active Learning in Higher Education*, 14(1), 63–76.
- Jonsson, A., & Panadero, E. (2018). Facilitating students' active engagement with feedback. In A. A. Lipnevich, & J. K. Smith (Eds.), *The Cambridge handbook of instructional feedback*. Cambridge, England: Cambridge University Press.
- Jung, T. P., Makeig, S., Humphries, C., Lee, T. W., Mckeown, M. J., Iragui, V., & Sejnowski, T. J. (2000). Removing electroencephalographic artifacts by blind source separation. *Psychophysiology*, 37(2), 163–178.
- Knowlton, B. J., Squire, L. R., & Gluck, M. A. (1994). Probabilistic classification learning in amnesia. *Learning & Memory*, 1(2), 106–120.
- Kulhavy, R. W., & Stock, W. A. (1989). Feedback in written instruction: The place of response certitude. *Educational Psychology Review*, 1(4), 279–308.
- Lavie, N., Hirst, A., De Fockert, J. W., & Viding, E. (2004). Load theory of selective attention and cognitive control. *Journal of Experimental Psychology: General*, 133(3), 339–354.
- Marco-Pallares, J., Cucurell, D., Münte, T. F., Strien, N., & Rodriguez-Fornells, A. (2011). On the number of trials needed for a stable feedback-related negativity. *Psychophysiology*, 48(6), 852–860.
- Nieuwenhuis, S., Holroyd, C. B., Mol, N., & Coles, M. G. (2004). Reinforcement-related brain potentials from medial frontal cortex: Origins and functional significance. *Neuroscience & Biobehavioral Reviews*, 28(4), 441–448.
- Riepl, K., Mussel, P., Osinsky, R., & Hewig, J. (2016). Influences of state and trait affect on behavior, feedback-related negativity, and P3b in the ultimatum game. *PLoS One*, 11(1), e0146358.
- San Martín, R. (2012). Event-related potential studies of outcome processing and feedback-guided learning. *Frontiers in Human Neuroscience*, 6(304), 1–17.
- San Martín, R., Manes, F., Hurtado, E., Isla, P., & Ibañez, A. (2010). Size and probability of rewards modulate the feedback error-related negativity associated with wins but not losses in a monetarily rewarded gambling task. *NeuroImage*, 51(3), 1194–1204.
- Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, 78(1), 153–189.
- Teodorescu, K., & Erev, I. (2014). Learned helplessness and learned prevalence: Exploring the causal relations among perceived controllability, reward prevalence, and exploration. *Psychological Science*, 25(10), 1861–1869.
- Winstone, N. E., Nash, R. A., Parker, M., & Rowntree, J. (2016). Supporting learners' agentic engagement with feedback: A systematic review and a taxonomy of recipience processes. *Educational Psychologist*, 52(1), 17–37.
- 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 & Evaluation in Higher Education*, 41(8), 1237–1253.
- Winstone, N. E., Nash, R. A., Rowntree, J., & Parker, M. (2017). 'It'd be useful, but I wouldn't use it': Barriers to university students' feedback seeking and recipience. *Studies in Higher Education*, 42(11), 2026–2041.