HOGREFE PUBLISHING - AUTHOR QUERY FORM

Journal Code: ZEA	Article DOI: 10.1027/1618-3169/a000242
Article Number: a000242	First Author: Alva Appelgren
Article Title: Impact of Feedback on Three Phases of Performance Monitoring	Corresponding Author: Sara Bengtsson

AUTHOR QUERIES – TO BE ANSWERED BY THE CORRESPONDING AUTHOR

Dear Author,

During the preparation of your manuscript for typesetting, the queries listed below have arisen. Please answer these queries by marking the required corrections at the appropriate point in the text.

Electronic file usage

Sometimes we are unable to process the electronic file of your article and/or artwork. If this is the case, we have proceeded by:

☐ Scanning (parts of) your article ☐ Rekeying (parts of) your article

Precopyedited Manuscript

☐ While processing the precopyedited manuscript some oversights were noted. These have been listed below.

Queries and/or remarks

Location in article	Query/remark	Response
Q1	Please check and approve the insertion of expansion for ACC.	
Q2	The reference "Hattie & Timperley, 2007" is cited in the text but not listed in the References. Please provide full reference details.	
Q3	As per style we are supposed to retain the name of the first six authors and the last author, so please provide the list of the remaining authors for the reference "Bengtsson et al. (2009)".	
Q4	Please provide publisher location for reference "Friston et al. (2007)"	

CE: Deepa
Thanks for your assistance.

Impact of Feedback on Three Phases of Performance Monitoring

Alva Appelgren,[1] William Penny,[2] and Sara L Bengtsson[1]

[1]Department of Clinical Neuroscience, Karolinska Institutet, Stockholm, Sweden, [2]Wellcome Trust Center for Neuroimaging, University College London, UK

Abstract. We investigated if certain phases of performance monitoring show differential sensitivity to external feedback and thus rely on distinct mechanisms. The phases of interest were: the error phase (FE), the phase of the correct response after errors (FEC), and the phase of correct responses following corrects (FCC). We tested accuracy and reaction time (RT) on 12 conditions of a continuous-choice-response task; the 2-back task. External feedback was either presented or not in FE and FEC, and delivered on 0%, 20%, or 100% of FCC trials. The FCC_{20} was matched to FE and FEC in the number of sounds received so that we could investigate when external feedback was most valuable to the participants. We found that external feedback led to a reduction in accuracy when presented on all the correct responses. Moreover, RT was significantly reduced for FCC_{100}, which in turn correlated with the accuracy reduction. Interestingly, the correct response after an error was particularly sensitive to external feedback since accuracy was reduced when external feedback was presented during this phase but not for FCC_{20}. Notably, error-monitoring was not influenced by feedback-type. The results are in line with models suggesting that the internal error-monitoring system is sufficient in cognitively demanding tasks where performance is $\sim$ 80%, as well as theories stipulating that external feedback directs attention away from the task. Our data highlight the first correct response after an error as particularly sensitive to external feedback, suggesting that important consolidation of response strategy takes place here.

Keywords: error-monitoring, external feedback, internal feedback, working memory, information theory

Error-monitoring is thought to be of particular importance for successful performance, since error signals directly call for adjustment of actions (Botvinick, Braver, Barch, Carter, & Cohen, 2001; Holroyd, Yeung, Coles, & Cohen, 2005; Ridderinkhof, Ullsperger, Crone, & Nieuwenhuis, 2004). One observation that has been made in support of this claim is that whereas RTs on most correct responses in a learned continuous choice task are fast, a characteristic of error-monitoring is a post-error slowing in RTs (Danielmeier, Eichele, Forstmann, Tittgemeyer, & Ullsperger, 2011; King, Korb, Von Cramon, & Ullsperger, 2010; Rabbitt, 1969). Rabbitt (1969) suggested that the slowing of responses immediately after errors is due to the validation of an error, and thus transient changes in response strategy to minimize the possibility of further errors. This proposal is supported by empirical findings that post-error slowing lowers the probability of committing a subsequent error in the post-error trial (Danielmeier et al., 2011; Rabbitt, 1969; Rabbitt & Rodgers, 1977). The conflict monitoring model by Botvinick et al. (2001) specifies that the Anterior cingulate conflict (ACC) plays a central role in error detection, serving as a learning signal that increases the threshold for executing the subsequent response. ACC has been found to register errors both when they are detected by the individual and when external error-feedback is provided and is thus regarded as a general error-monitoring module (Holroyd et al., 2004; Ullsperger, Nittono, & von Cramon, 2007).

Post-error slowing has not always direct bearing on the subsequent performance leading to improved performance (Hajcak, McDonald, & Simons, 2003). Notebaert et al. (2009) have proposed an alternative account for the post-error slowing where the slowing is caused by the error being a rare outcome and therefore grasping attention. Thus, it may take attentional resources from the task, which may result in reduced performance (Huettel & McCarthy, 2004). They found that when correct responses outnumbered error responses, post-error slowing was observed, whereas when the majority of the trials were incorrect post-correct slowing was observed (Notebaert et al., 2009). Regardless of whether external error-feedback was present or not, they found the same pattern of prolonged post-error RT when errors were rare outcomes and the absence of post-error slowing when error frequency reached 50%, which made them argue that the internal error-monitoring system is more important than the external. The accuracy levels were fixed and therefore the impact of feedback on accuracy was not investigated (Houtman, Castellar, Notebaert, & Nu, 2012).

External feedback on trial outcomes informs us on task success. We use this feedback to confirm, restructure, or tune information so that behavior meets the task goals (Hattie & Timperley, 2007). Feedback signals are designed to minimize the risk that a participant would miss the outcome and as such the feedback may grasp attention. It is unclear whether this is beneficial for performance or if it directs attention away from the task. A meta-analysis on feedback interventions showed that one third of the studies reported reduced performance upon external feedback (Kluger & DeNisi, 1996). No consistent conclusion could be drawn as to whether feedback played a different role dependent on the type of task, for example, vigilance tasks or problem-solving tasks. The main factors contributing to the impact of explicit feedback on performance were if outcome was measured on a trial-to-trial basis or after a time of consolidation (Goodman, 1998; Schmidt, Young, Swinnen, & Shapiro, 1989), if outcome was measured in terms of the intention of the participants to invest effort (motivation) (Van-Dijk & Kluger, 2004), or if feedback was given on errors or corrects (Wade, 1974). Goodman (1998) showed that detailed task-feedback when solving a puzzle helped the participants to perform better, but the absence of explicit feedback had beneficial learning effects in the long run, that is, to solve a later puzzle. A similar pattern of results was observed in a study by Schmidtet al. (1989) where the frequency of feedback was manipulated and they observed that error rate increased when feedback was delivered after every trial, compared to when feedback was delivered after every 15th trial. They concluded that feedback after every trial may eliminate the participant's internal evaluation process. Van-Dijk and Kluger (2004) demonstrated that the participants' intention to invest effort was influenced by whether they preferred positive or negative feedback. Wade (1974) used a letter matching task and asked participants to confirm with a button press that they had understood the task-feedback after each trial. They either confirmed the feedback for errors, for corrects, for both the errors and corrects or neither. Selective feedback on correct responses or on the error responses led to the best performance results. Even though results suggest that external error-feedback has limited impact (Holroyd et al., 2004; Houtman et al., 2012), it may still be argued that we process error-feedback as more valuable than feedback on correct responses when errors are rare outcomes, as would be predicted from an information theoretic perspective (Shannon & Weaver, 1963). For example, if an individual makes 20% errors on a continuous performance choice task, providing external feedback on the error trials would give them more information than if external feedback was given on 20% of the correct responses. This argumentation is lined out in more detail in Information Theory section. We can compute the Mutual Information (MI) between feedback and outcome, which quantifies how informative the external feedback is about the outcome. It has been shown that external error-feedback is processed in different neural circuits than external feedback on correct responses (Ullsperger & von Cramon, 2003). These results illustrate that feedback-type, that is, erroneous and correct feedback, may matter for performance.

An interesting observation is that among the correct responses, the first correct response after an error seems to differ from other correct responses, where the correct response following an error gives rise to more activity in, for example, right dorsolateral prefrontal cortex (Kerns et al., 2004; King et al., 2010; Marco-Pallarés, Camara, Münte, & Rodríguez-Fornells, 2008). Although less explored than the post-error slowing, there are reports of the first correct response after an error also slowing RT (Laming, 1979; Marco-Pallarés et al., 2008; Rabbitt, 1969). This slowing could reflect that the individual responds more cautiously because of a recent error; in order to guard against further errors (Laming, 1968), or because a change in strategy contingent on his recognition of his mistake (Rabbitt, 1969). The impact of external feedback has not been evaluated for this phase in particular.

In the present study we investigate if three phases of performance monitoring, the error phase, the phase of the correct response after an error, and the phase of corrects following correct responses, are differentially influenced by external feedback and whether the external feedback is beneficial for performance or not. We measured accuracy and RTs on a 2-back task for letters. The 2-back task is a continuous performance task where each trial is dependent on other trials, and as such it measures a person's sustained and selective attention. This is useful when investigating interactions effects of feedback between the phases. Interactions, that is, how feedback in one phase may be influenced by feedback on previous trials, require that there is a sequential dependence between trials. This is seen for tasks such as the *n*-back task, but not for tasks where each trial is preceded by separate rules. In the present study it was important to use a task that was moderately difficult, since we are investigating error processing. The accuracy level of the *n*-back task can easily be manipulated by varying *n*. Additionally, by comparing experimental conditions with the same number of feedback events (sounds), but varying in the amount of information feedback conveys about outcome (the mutual information), we can test if information content has an effect on performance.

Because the above studies suggest that the three phases rely on different processes, we hypothesize that external feedback is processed differently for errors, correct after error, and corrects following corrects. Whereas we do not predict that external error-feedback will alter performance when compared to no external feedback on errors, we do hypothesize that error-feedback will be more informative than feedback on correct responses.

Method

Participants

Sixty-three neurologically healthy, right-handed participants took part in this study (age range 18–40 years, mean age ± *SD*: 26.8 ± 5.1, 43 females). Three participants were excluded before the data analysis because they did not complete the task. Participants were recruited from the Stockholm area and they all gave written informed consent prior to participating in the study. The study was approved by the ethics committee in Stockholm, Sweden (Dnr No. 2010/1546-31/1).

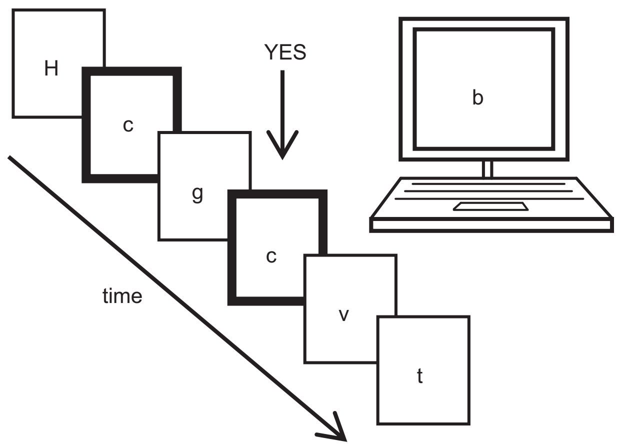

Figure 1. The 2-back task. A sequence of letters is presented on a computer screen one letter at a time. Participants are asked to make a response for each presented letter: A "yes" response on the computer keyboard if the letter also appeared two letters back, or a "no" response if it did not.

Experimental Procedure

The experimental task was performed on a PC (Latitude E5510, DELL Inc., Texas, US) with a screen resolution of 1366 × 768. We used Cogent (UCL, London, UK) for sequence presentation and data collection. Prior to data collection we conducted a pilot study where n was either 1, 2, 3, or 4 and found that $n = 2$ yielded an accuracy level of ≈ 80%. In this pilot study eight participants performed a sequence of 60 letters for each n. Accuracy was: $n = 1$ (84.0% ± 14.3), $n = 2$ (78.3% ± 21.5), $n = 3$ (63.4% ± 27.0), $n = 4$ (56.5 ± 25.4).

The 60 participants in the present study were seated in a quiet testing room and were tested on the 2-back task for letters (Figure 1), a task widely used to test the ability to maintain information across a delay (Cohen, MacWhinney, Flatt, & Provost, 1993). We used a sequence of 200 letters per condition. White letters (10 mm in height) were presented centrally on a black computer screen, one letter at the time. Each letter was presented for 230 ms with an interstimulus interval (ISI) fixed to 1,400 ms. If the letter they saw also appeared two letters back the participant made a "yes" response, otherwise they made a "no" response. The "yes" response consisted of pressing the button corresponding to the right index finger, while a "no" response was made by pressing the button corresponding to their right middle finger, on the computer keyboard. The same letter, regardless if written as capital letter or lowercase letter, was regarded a match. Both capital and lowercase letters were used in the sequences to reduce the possibility that participants solely relied on visual memory. A sequence had 30% hits ("yes" responses).

In order to study the influence of external feedback on the performance monitoring system, either an auditory signal delivered through headphones, or no sound, followed immediately after each key response. Two different sounds were used as external feedback; a 74 Hz beep (55 ms) indicating an error and a 740 Hz beep (55 ms) indicating a correct answer. The participants were not instructed to correct their errors.

We compared external and no external feedback on errors and correct responses, where the correct responses were divided into corrects after errors, and corrects following corrects. This enables us to study if the correct responses differ in their processing depending on the outcome of the preceding trial. This gives us three factors: the error phase (FE), the phase of corrects after errors (FEC), and the phase of corrects following corrects (FCC). Each of the factors had two or three levels of feedback. The error phase had two levels of feedback; either external feedback on all errors (FE_{100}) or no external feedback (FE_0). The phase "corrects after errors" had two levels of feedback; either external (FEC_{100}) or no external feedback (FEC_0). The phase "corrects following corrects" had three levels of feedback; external feedback on 100% of the correct responses (FCC_{100}), external feedback on 20% of the correct responses randomly distributed (FCC_{20}), or no external feedback (FCC_0). The reason for having three levels of feedback on FCC was because we wanted to compare external feedback with internal feedback (100% sound vs. 0% sound), as well as to investigate a parametric modulation of the amount of external feedback on performance, and thirdly, to test the information theory hypothesis suggested in the Introduction and Information Theory sections. Testing this hypothesis required that we introduce sequences with feedback on 20% of the correct following correct responses (FCC_{20}), since this would roughly correspond to the percentage of errors made. We cannot know beforehand how many errors the participants will make, so an exact correspondence in the amount of sound between the two sequences was not possible. In total, the study was made up of twelve 2-back conditions, each condition consisted of a 200-letter long 2-back sequence. These conditions fitted in a 2 × 2 × 3 factorial design (Figure 2).

The three phases of interest are denoted; FE: feedback on errors, FEC: feedback on the correct response after an error, and FCC: feedback on correct responses following corrects. When describing our 12 different feedback conditions we use the order; error, correct after errors, correct following corrects [FE; FEC; FCC]. We denote external feedback (sound) as 1 and no external feedback (silence) as 0 for the phases FE and FEC. For FCC, 0 corresponds to no external feedback (silence), 1 corresponds to external feedback on 20% of the trials, and 2 corresponds to external feedback on all of the corrects following corrects (Figure 2). For example, [101] denotes a 2-back sequence where external feedback was received on error trials as well as on 20% of the FCC trials, and [002] denotes a 2-back sequence where no external feedback is given on errors, nor the subsequent correct response, but external feedback is given on all corrects following corrects.

For each condition, instruction of the feedback characteristics was presented on the computer screen for 1,000 ms. This was followed by a sequence of 100 letters. Each feedback condition was presented twice, so in total 200 letters were presented for each condition for each

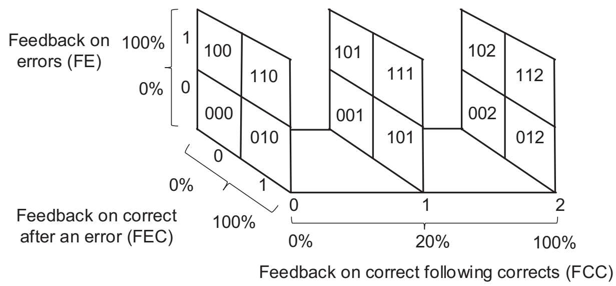

Figure 2. The 2 × 2 × 3 factorial design. We focused our analysis on three phases of performance monitoring; the error phase (FE), the phase of the correct response after an error (FEC), and the phase of the corrects following correct responses (FCC). We manipulate the performance monitoring by delivering external feedback (sounds), or no external feedback, on FE and FEC, while on FCC trials we either provide external feedback on 100%, 20%, or none of the trials. This results in 12 conditions of the 2-back task with different combinations of feedback. We denote the experimental conditions in the order [FE; FEC; FCC].

participant, apart from sequences [000] and [011] where only 96% of the letters were presented due to technical failure. There were four types of 2-back sequences of letters that were randomized between conditions. The design is a mixed design, each participant performed on average 3.5 ± 1.5 conditions. The order of conditions between participants was pseudorandomized, and the subject effect was taken into account in the statistical analysis.

Prior to data collection, the participants practiced each of the sequences they were to perform, for 25 letters per condition, and were at the same time becoming familiar with the two sounds representing errors and corrects respectively. They were verbally instructed on the task rules with the aid of a cartoon. They were carefully instructed on the characteristics of each sequence and its corresponding computer instruction label.

Statistical Analysis

We measured percent correct responses (accuracy) and RT as dependent variables (Table 1). Prior to data analysis, we excluded nonresponse trials and removed the first two trials of each 100-letter sequence because of the nature of the 2-back task, that is, only from the third letter presented can a response be a match or a mismatch. When computing RT, we excluded error-trials that were followed by another error trial. When computing the RTs we extracted the time between the stimulus presentation and key press. Accuracy was computed on all trials included in the analysis. In total 31,103 trials were entered into the analysis. On average 173.4 ± 10.0 trials/condition/participant were entered into the analysis.

We performed a 3-way ANOVA based on summary statistics for each subject and feedback combination (df = 164) using Matlab (r2010a, The Math Works, Natick, MA) and the spm_ancova function from the SPM software library (Friston, Ashburner, Kiebel, Nichols, & Penny, 2007) compatible with Matlab, to make a between-subjects design after correcting for subject effects. We investigated the main effect of FE, FEC, and FCC for accuracy and RTs using the 12 different conditions, as well as the interaction effects among them. The main effects show us the average effect of a factor when this factor is "high" versus "low," that is, to compute the main effects (RT and accuracy) of external feedback on errors we subtract the average response of all experimental runs for which FE was low (no external feedback, conditions [000] [001] [002] [010] [011] [012]) from the average responses of all experimental runs for which FE was high (external feedback on errors [100] [101] [102] [110] [111] [112]).

Table 1. Descriptive statistics for accuracy, RT and double errors are shown for each of the 12 conditions.

Conditions [FE;FEC;FCC]	Accuracy (%) ±*SEM*	RT (ms) ± *SEM*	Double errors ±*SEM*
[000]	87.4 ± 2.2	580.0 ± 38.7	1.4 ± 0.8
[001]	88.1 ± 0.8	558.0 ± 12.3	2.5 ± 0.3
[002]	85.2 ± 0.8	545.2 ± 13.7	4.9 ± 1.0
[010]	81.2 ± 0.9	561.2 ± 6.9	8.1 ± 1.1
[011]	85.2 ± 0.7	578.1 ± 9.31	3.8 ± 0.8
[012]	87.2 ± 0.8	539.6 ± 6.7	4.3 ± 0.8
[100]	88.9 ± 1.3	570.9 ± 18.5	1.0 ± 0.3
[101]	86.8 ± 0.6	552.5 ± 7.8	3.4 ± 0.9
[102]	86.0 ± 1.5	566.7 ± 14.7	4.4 ± 1.3
[110]	85.0 ± 0.7	575.3 ± 8.7	3.0 ± 1.2
[111]	85.0 ± 1.2	581.4 ± 15.4	3.9 ± 1.2
[112]	83.9 ± 0.8	555.6 ± 9.4	4.6 ± 0.8

We then counted the committed double-errors for each participant and condition. The number of double-errors is sometimes used to study how readily participants are monitoring and adjusting their errors (Hajcak & Simons, 2008; Houtman et al., 2012; Notebaert et al., 2009). This measure will give us an indication of: (i) if in the absence of external error-feedback the participants make more double-errors because they monitor their error less readily or (ii) if when external error-feedback is present, the feedback disturbs the participants' internal error-monitoring hence resulting in more double-errors. We compared the number of double-errors between conditions where external error-feedback was presented with those without external error-feedback using a one-way ANOVA. We also correlated double-errors with performance using Pearson's correlations (SPSS Statistics 17.0, Chicago, IL) for each condition to study possible individual differences in response to external feedback.

Additionally, for each subject and sequence, we compute the Mutual Information (MI) between feedback and outcome. Details of this computation are provided in Information Theory section. MI quantifies how informative the external feedback is about the outcome. For each sequence, we then regress subject RT's onto subject MI's to see if, over the group, more informative feedback significantly increases or decreases RT. Here we could compare MI for the sequences [100] and [001] to test our information theoretic hypothesis (see Introduction). We also compare their accuracy levels with a two-sided Student's *t*-test.

We supplemented our hypothesis testing concerning "feedback on errors" with Bayesian statistics in order to quantify how much evidence there is in favor of the null hypothesis. This approach is now becoming widely adopted in experimental psychology (Dienes, 2011). Our analysis was based on mean-corrected average accuracy and average RT for each condition. We used a custom written Matlab script for Bayesian ANOVAs where computations were based on Equation 1 in Wetzels and Wagenmakers (2012). The output of this analysis is a Bayes Factor which quantifies the strength of evidence for the alternative versus the null hypotheses, with values larger than 1 favoring the alternative and less than 1 favoring the null. These values are grouped in ranges (Jeffreys, 1961) quantifying "weak" (1/3–1), "substantial" (1/10–1/3), and "strong" (1/30–1/10) evidence for the null. The equivalent Log Bayes Factors are −1.1 to 0 for weak, −2.3 to −1.1 for substantial, and −3 to −2.3 for strong.

Information Theory

If there is an outcome $o = \{c,e\}$ where c is correct and e is error with probabilities pc and pe with $pc = 1 - pe$ then Shannon defines the "surprise" of an outcome as measuring the improbability of that event (Shannon & Weaver, 1963). Mathematically, surprise is defined as $\log_2(1/p)$ where p is the probability of an event and use of base-2 logarithms means that surprise is measured in bits. Thus the surprise associated with an error is $\log_2(1/pe)$ and with a correct is $\log_2(1/pc)$. The information content of a variable, also known as the entropy, is then the average surprise. The entropy of the outcome is $H(o) = pe \times \log_2(1/pe) + pc \times \log_2(1/pc)$. The entropy measures the information content of a variable, in bits. The more uncertain we are about the value of a variable the greater the information conveyed when it is observed. For $pe = 0.2$, we have $H(o) = 0.72$ bits. Note that $H(o)$ would reach a maximal possible value of one bit if $pe = 0.5$.

If there is feedback (f) in the form of a sound $f = \{s,n\}$ where s is sound and n is no sound with probabilities ps and pn, with $pn = 1 - ps$, then the entropy of the feedback is $H(f) = ps \times \log_2(1/ps) + pn \times \log_2(1/pn)$.

Importantly we can also quantify the information one variable contains about another. This is given by the mutual information (MI). For example, the mutual information between feedback and outcome is the reduction in uncertainty about outcome after experiencing feedback. Mathematically this is given by the uncertainty in the outcome, $H(o)$, minus the uncertainty in the outcome after having received feedback, $H(o|f)$. That is, $\text{MI} = H(o) - H(o|f)$. The mutual information is a strictly positive quantity.

Calculating the mutual information of our two fictive sequences (see next section for details of this calculation) gives the following result: Sequence 1 (20% errors, auditory feedback on all errors) gives MI = 0.722; Note that this is the same as $H(o)$ because there is no uncertainty in the outcome after feedback (i.e., $H(o|f)$ is zero). This is because feedback is always provided after an error so, upon hearing a sound we can be sure we made an error. Sequence 2 (20% errors, no feedback on errors, auditory feedback on 20% of correct responses) gives MI = 0.057. That is, Sequence 2 feedback provides less information about outcome than does Sequence 1.

Note that we cannot match the number of sounds between the two sequences perfectly since the error rate varies between participants. We have used an estimation based on previous data that participants perform between 80% and 90% correct and therefore set the amount of feedback received on the correct trials to 20%, which corresponds to approximately 16%–18% of the total amount of trials. However, the potential difference between the two sequences is small.

Computing the Mutual Information Between Outcome and Feedback

For many of the sequences we have used, the type of feedback (sound or no sound) depends on the outcome of the current trial and the previous trial. The levels of the three experimental factors FE, FEC, and FC determine the values of the following probabilities:

$$
\begin{aligned}
p(\text{FE}) &= p(s_t|e_t)\\
p(\text{FEC}) &= p(s_t|c_t,\, e_{t-1})\\
p(\text{FC}) &= p(s_t|c_t,\, c_{t-1})
\end{aligned}
$$

where t indexes the trial. $p(\text{FE})$ can be 0 or 1, $p(\text{FEC})$ can be 0 or 1, and $p(\text{FC})$ can be 0, 0.2, or 1. The experimental condition specifies these probabilities. Given these, and the error probabilities $p(e_t) = 1\text{-}p(c_t)$, we have the quantities we need to compute the entropies and mutual information. First we compute the joint probability of the eight possible three-way events:

$$
\begin{aligned}
p(s_t, e_t, e_{t-1}) &= p(s_t|e_t)p(e_t)p(e_{t-1})\\
p(n_t, e_t, e_{t-1}) &= [1 - p(s_t|e_t)]p(e_t)p(e_{t-1})\\
p(s_t, e_t, c_{t-1}) &= p(s_t|e_t)p(e_t)p(c_{t-1})\\
p(n_t, e_t, c_{t-1}) &= [1 - p(s_t|e_t)]p(e_t)p(c_{t-1})\\
p(s_t, c_t, e_{t-1}) &= p(s_t|c_t, e_{t-1})p(c_t)p(e_{t-1})\\
p(n_t, c_t, e_{t-1}) &= [1 - p(s_t|c_t, e_{t-1})]p(c_t)p(e_{t-1})\\
p(s_t, c_t, c_{t-1}) &= p(s_t|c_t, c_{t-1})p(c_t)p(c_{t-1})\\
p(n_t, c_t, c_{t-1}) &= [1 - p(s_t|c_t, c_{t-1})]p(c_t)p(c_{t-1})
\end{aligned}
$$

where we have assumed $p(e_t,\ e_{t-1}) = p(e_t)p(e_{t-1})$. We also assume $p(e_t) = p(e_{t-1})$. We then compute the probabilities of the four possible two-way events.

And then the probabilities of sound and no sound:

$$
\begin{aligned}
p(s_t) &= p(s_t, e_t) + p(s_t, c_t)\\
p(n_t) &= p(n_t, e_t) + p(n_t, c_t)
\end{aligned}.
$$

The mutual information between feedback and outcome is then given by

$$MI = p(s_t, e_t)\log_2 \frac{p(s_t, e_t)}{p(s_t)p(e_t)} + p(n_t, e_t)\log_2 \frac{p(n_t, e_t)}{p(n_t)p(e_t)} + p(s_t, c_t)\log_2 \frac{p(s_t, c_t)}{p(s_t)p(c_t)} + p(n_t, c_t)\log_2 \frac{p(n_t, c_t)}{p(n_t)p(c_t)}.$$

Our calculation of the mutual information assumes that subjects have no knowledge of the outcome prior to receiving external feedback. However, it may be the case that subjects are able to assess whether their response was correct or incorrect using their internal monitoring system. Evidence against the information theoretic hypothesis (as characterized using the MI equation derived above) is therefore evidence in favor of an internal monitoring system. We return to this topic in the discussion.

Results

Accuracy

Effect of Gender, Age, and Order

There was no effect of gender, age, or condition order on the accuracy level.

Main Effects

FE: External feedback on errors showed no significant difference compared to no external feedback on errors in accuracy level, $F(1, 164) = 0.02$, $p > 0.89$, log Bayes Factor = −2.55 (Figure 3A). The Bayes factor provides strong evidence for the null hypothesis.

FEC: External feedback on corrects after errors revealed a significant effect, compared to no external feedback on corrects after errors, $F(1, 164) = 9.94$, mean effect size 0.11%, $p < 0.001$. As seen in Figure 3B, there was a reduction in performance when participants were presented with external feedback.

FCC: External feedback on correct following corrects revealed a significant effect, $F(2, 164) = 4.74$, mean effect size 0.6%, $p < 0.0001$ (Figure 3C). A post hoc pairwise analysis showed a significant difference in performance between FCC_{100} and FCC_{20} ($p < 0.04$). The comparison between FCC_{100} and FCC_0 did not reach significance ($p > 0.16$). There was no significant change in accuracy between FCC_{20} and FCC_0 ($p > 0.55$).

Interactions

FE-FCC: There was a significant interaction between errors and corrects following corrects, where external feedback on error, together with FCC_{100}, that is, the two conditions [102] [112], resulted in reduced performance compared to other FE and FCC combinations, $F(2, 164) = 71.8$, $p < 0.0001$.

FEC-FCC: The interaction analysis between corrects after errors and corrects following corrects also revealed a significant effect $F(2, 164) = 75.2$, $p < 0.0001$. Performance was significantly improved when no external feedback was presented on FEC (FEC_0) in combination with either no external feedback on the corrects following corrects or when there is external feedback on only 20% of the corrects following corrects.

The interaction between FE and FEC and the three-way interaction FE-FEC-FCC did not reveal any significant differences.

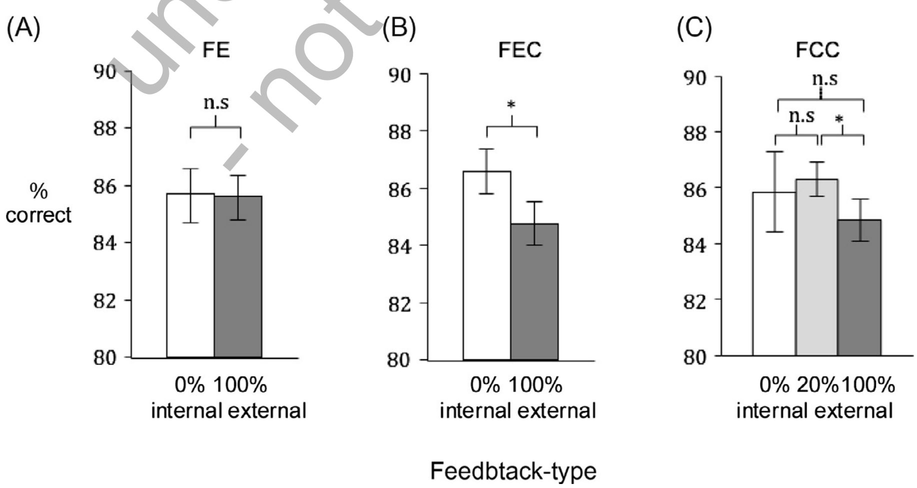

Figure 3. Accuracy; main effects of feedback. Each bar corresponds to the average ± *SEM* of the mean accuracy of each of the conditions within the main effects. (A) Errors: The main effect of FE showed no significant difference to whether external or internal (no external) feedback was presented ($p > 0.89$). (B) The correct response after an error: Main effect of FEC showed a significant reduction in performance when external feedback was presented during this period ($p < 0.002$). (C) Correct following corrects: Main effect of FCC showed a significant effect ($p < 0.01$).

Feedback and Double-Errors

To investigate if there may be any sign of reduced error detection in the six conditions without external error-feedback we compared the number of double-errors between the conditions with and without external error-feedback. There was no significant difference between the two groups $t(10) = 0.31$, $p > 0.76$, log Bayes Factor = −2.31 (mean double-errors external error-feedback: 4.17 ± 0.86; no external error-feedback: 3.39 ± 0.49), nor between the 12 conditions, $F(11) = 1.6$, $p > 0.10$. See Table 1 for individual data. The Bayes factor provides substantial evidence for the null hypothesis.

Correlations between accuracy and double-errors showed that in the four sequences where external feedback was given on some random general corrects (FCC_{20}), participants who performed worse made more double-errors; [001] $r^2 = 0.495$, $p < 0.01$; [101] $r^2 = 0.61$ $p > 0.001$; [011] $r^2 = 0.42$, $p < 0.001$; [111] $r^2 = 0.60$ $p < 0.01$. Also in two of the conditions where external feedback was presented on all "corrects following corrects" (FCC_{100}) the participants that performed the worse made more double-errors [112] $r^2 = 0.34$, $p < 0.05$; [002] $r^2 = 0.52$ $p < 0.05$. There was a marginal significance in the condition [012] $r = 0.211$ $p < 0.11$.

Reaction Time

Main Effects

FE: External feedback on errors revealed no significant effect compared to no external feedback on errors $F(1, 164) = 0.69$, $p > 0.41$, log Bayes Factor = −2.23 (Figure 4A). The Bayes factor provides substantial evidence for the null hypothesis.

FEC: External feedback on corrects after errors did not show any significant difference in RTs compared to no external feedback on corrects after errors $F(1, 164) = 0.14$, $p > 0.71$ (Figure 4B).

FCC: There was a significant main effect of external feedback on corrects following corrects, $F(2, 164) = 4.88$, mean effect size 17.41 ms, $p < 0.008$. A significant shortening in RT was observed for FCC_{100} when compared to FCC_0 ($p < 0.05$). There was a marginal significance ($p < 0.11$), in a shortening of RT for FCC_{100} when compared to FCC_{20}. No significant difference was observed between no external feedback and 20% external feedback on corrects following corrects, FCC_0 versus FCC_{20} ($p > 0.66$; Figure 4C).

Interactions

FEC-FCC: The interaction analysis regarding RT between external feedback on corrects after errors and corrects following corrects revealed a significant effect $F(2, 164) = 3.3$, $p < 0.04$ meaning that RT was significantly faster in the conditions where external feedback was received on corrects after errors together with external feedback on all corrects following corrects, that is, the [012] and [112].

No other interactions were found to be significant.

Testing the Use of Feedback With Information Theory

To evaluate the hypothesis that external feedback on an error would be of more information value to participants

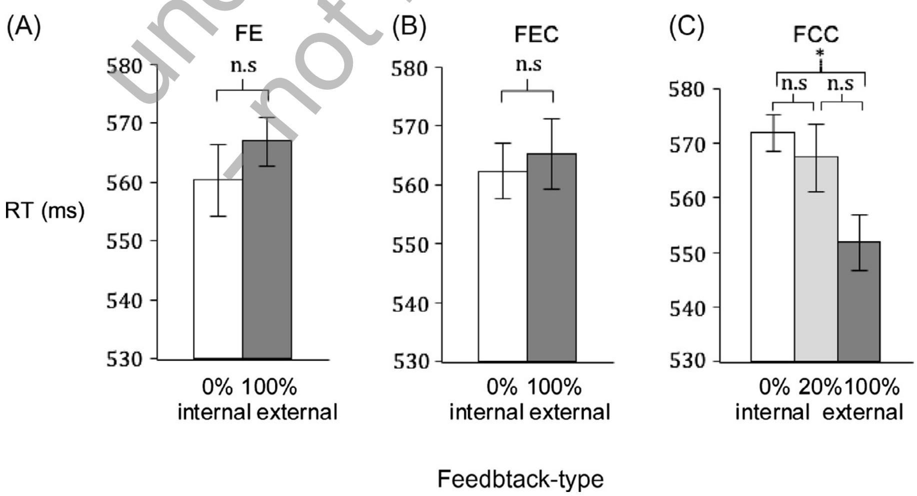

Figure 4. RT; Main effect of feedback. Each bar corresponds to the average ± *SEM* of the mean RT of each of the conditions within the main effects. (A) Errors: Main effect of FE showed no significant effect. (B) The correct response after an error: Main effect of FEC showed no significant effect. (C) Correct following corrects: There was a significant main effect of FCC. RT was significantly faster when external feedback was provided on FCC_{100} compared to no external feedback ($p < 0.05$).

than external feedback on correct responses, we compared conditions [100] and [001]. There was no significant difference in performance between the conditions [100] and [001], $t(22) = 0.27$, $p > 0.6$, log Bayes Factor = −3.1, nor were these conditions influenced by MI, that is, the amount of information the feedback signal provides about the outcome, [100] $r = 0.2$, $r^2 = 0.04$, $p > 0.52$; [001] $r = 0.02$, $r^2 = 0.0004$, $p > 0.94$. The Bayes factor provides strong evidence for the null hypothesis.

The instances when the external feedback signal was used by the participants as sufficient information about the outcome to influence RT were for the two sequences that contained the largest amount of sound: [012] and [102] ([012] $r = -0.64$, $r^2 = 0.41$, $p < 0.02$; [102] $r = -0.52$, $r^2 = 0.27$, $p < 0.04$). These significant correlations mean that the more information the participant extracts from the feedback signal about the outcome the shorter the RT. Note that the analysis could not be performed on the sequences [000] and [112], as MI did not vary over participants (this is because external feedback is provided on none or every outcome).

Discussion

Our results indicate a differential effect of feedback on performance depending on in which phase the feedback is presented. Accuracy and RTs vary depending on feedback-type and phase. We find that error-monitoring differs from the subsequent correct response, in the sense that the phase on the correct after an error (FEC) is sensitive to external feedback, whereas errors (FE) are not. FEC appears to differ from FCC responses as well. There was a reduction in performance for both the main effects (FEC and FCC) when external feedback was provided, however a closer look on the FCC conditions revealed that FCC_{100} was responsible for this effect. Moreover, the feedback did not influence RTs on FEC, but did so significantly for FCC_{100}. This finding shows that the FEC in particular is a phase sensitive to external disturbance.

We do not seem to care about whether we are externally informed about errors or not, since there is no difference in how people perform with and without error-feedback, as revealed by our main effects analyses. To quantify how much evidence there is in favor of no difference in performance between external and no external feedback on errors, we computed the log Bayes Factor (logBF). We found that for accuracy logBF was −2.55 and for RT logBF was −2.23. This tells us that it is about (exp(2.5) = 12.2) 12 times more likely that the data have occurred under the null hypothesis than the alternative hypothesis. In other words, this is a strong support for the null hypothesis (Jeffreys, 1961). When investigating the effect of external feedback on errors with an information theoretic model, again we found no evidence for the hypothesis that the brain utilizes external error information more readily than external information about other outcomes in a cognitively demanding sequential response task. Looking at the two sequences with the highest performance scores [100] and [001], one of which had external feedback on errors (approximately 20% errors), the other which had sounds delivered on approximately 20% of the correct responses randomly distributed, there was no significant difference in accuracy scores. Supplementary Bayesian statistical analysis gave a Log Bayes Factor of −3.10, which gave strong support for the null hypothesis. The finding is in line with a brain imaging study by Holroyd et al. (2004) showing that ACC responds in a similar magnitude to errors independent of external or internal feedback. It therefore seems unlikely that the participants are unaware of their errors in the conditions without external error-feedback, or that the external error-feedback would interfere with performance monitoring. Nevertheless, we looked into this issue by counting double-errors arguing that there would be more of these if the participants lacked coherent error-monitoring. We found no support for more double-errors being committed in either the internal or the external error-feedback conditions. The estimated Log Bayes Factor was −2.31, which gives us a substantial support in favor of the null hypothesis. This supports the claim that feedback-type on errors, on a task where the accuracy level is around 80%, has no impact on error-monitoring.

When we computed the MI, that is, the reduction in uncertainty before versus after hearing the feedback, we assumed that the participant thinks they got it right with a probability of 80% (average performance level) before hearing the tone. This however, turned out to be wrong. This is most likely due to the fact that the brain has already worked out the outcome (error or correct) prior to the feedback signal. The real uncertainty before hearing the feedback is much less and so the MI is much less. Thus, we can infer from the results given from the information theory that the participants are not ignorant about the outcome before hearing the feedback because the internal monitoring system is doing a good job. This is consistent with our other analyses, which show that external feedback does not help. We argue that this is due to the efficiency of our error-monitoring system, which has developed through evolution to assist progress and survival without having to rely on external sources.

Only when external feedback was given on each of the correct responses following corrects was there a significant reduction in both accuracy and RT. Reduced RT with increased amount of external feedback has previously been observed by Houtman et al. (2012). The correlation between MI and RT for these conditions supported the above finding in showing that RT is influenced by the information from the external feedback when the sequences consist of a large amount of external feedback (>80%) and is influenced in such a way that RT is being shortened. Our finding of the information theory that the participants most likely register their outcome before the feedback signal is delivered suggests that the effect that feedback on many correct responses has is preparatory, or confirmatory, rather than reactive. We know from a previous study that predictable auditory signals automatically activate pre- and primary motor cortices and suggestively lower the execution threshold (Bengtsson et al., 2009). In order to generate a response, according to the Evidence Accumulation type models (Gold & Shadlen, 2001), the motor system triggers a response signal when enough information has been accumulated to reach decision threshold. In the present study, it seems as if the feedback signal is incorporated into preparing a response that lowers the threshold. For about 80% of the trials the participants are doing fine, they are in a "standard/automatic response mode," perhaps gradually losing task control exercised on the motor system by the prefrontal cortex. Alternatively, the effect that large amount of external feedback leads to reduced performance accuracy could be due to superfluous external information taking up attentional resources (MacLeod & MacDonald, 2000). A third possibility is that the phonological loop used during working memory (Baddeley, Gathercole, & Papagno, 1998) is active during the *n*-back task for letters, and that the auditory feedback interferes with this loop. However, we have unpublished pilot data showing that also visual feedback, in the form of a flash of light, disturbs performance, which would speak against an interaction between the external feedback and the *n*-back task within the phonological loop. Future brain imaging data will shed light on which of these mechanisms is operating.

From our results we conclude that processes active during FEC are different from those active during FE. It is therefore unlikely that the phase FEC would display simply more "cautious" behavior as a consequence of the error as suggested by Laming (1968). Instead, we suggest that this period contains an additional process unique for this phase, which may be one of consolidation, stating that the change of strategy was accurate. This finding is in line with brain imaging studies showing a different activity pattern in this phase when compared to errors as well as other correct responses (Marco-Pallarés et al., 2008). Delivering external feedback on 20% of corrects following corrects did not significantly change performance. When participants make an error they need to reset their response mode and the outcome of the trial after an error is therefore crucial for evaluating whether the response mode is reset correctly. While they are assessing this it seems particularly deleterious to also process external feedback signals, while on a correct response after a correct response they have already established that their response mode has been appropriately reset.

We found that the participants who were the weaker performers made significantly more double-errors in the conditions where they were presented with feedback on random correct responses and FCC_{100}. This shows that not only are there individual differences in how people handle external feedback, but that the sequential structure of the feedback matters as well. In fact, we find that certain combinations of feedback between the different phases matter for accuracy. For example, external error-feedback together with external feedback on corrects gave the poorest accuracy, whereas no external feedback on the first correct after an errors together with less than 20% feedback on other corrects, regardless of error-feedback, led to the best performance. This suggests that the participants, to a certain degree, process an outcome in relation to the character of previous trials.

Conclusion

In summary, our finding that external error-feedback does not influence performance is in line with the theories that outline ACC as a generic error-monitoring system (Botvinick et al., 2001; Holroyd et al., 2005) and resonates with the finding of Houtman et al. (2012). Thus, our finding supports the notion that the internal error-monitoring system is sufficient in cognitive tasks where accuracy is around 80%. We find that external feedback on correct responses leads to deteriorating accuracy, which suggests that external signals are diverting attention away from the task when present on correct responses. An interesting novel finding is that the correct response after an error is particularly sensitive to external signals, which suggests that important internal consolidation of strategy implementation takes place here. We propose that feedback manipulations of three different phases can be used in future studies to investigate individual characteristics and deviations in performance monitoring.

Acknowledgments

This study was supported by VINNMER, Vinnova – Swedish Governmental Agency for Innovation Systems, the Swedish Research Council (VR), and Cornells Stiftelse. We thank Martin Ingvar, Mats Olsson, Johan Eriksson, and Yvonne Brehmer for valuable discussions and comments on the manuscript.

References

Baddeley, A., Gathercole, S., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review, 105*, 158–173. doi: 10.1037/0033-295X.105.1.158

Bengtsson, S. L., Ullén, F., Ehrsson, H. H., Hashimoto, T., Kito, T., Naito, E., Forssberg, H., et al. (2009). Listening to rhythms activates motor and premotor cortices. *Cortex, 45*, 62–71. doi: 10.1016/j.cortex.2008.07.002 Q3

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review, 108*, 624–652.

Cohen, J. D., MacWhinney, B., Flatt, M. R., & Provost, J. (1993). Psy-Scope: A new graphic interactive environment for designing psychology experiments. *Behavior Research Methods, Instruments, & Computers, 25*, 257–271.

Danielmeier, C., Eichele, T., Forstmann, B. U., Tittgemeyer, M., & Ullsperger, M. (2011). Posterior medial frontal cortex activity predicts post-error adaptations in task-related visual and motor areas. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience, 31*, 1780–1789. doi: 10.1523/JNEUROSCI.4299-10.2011

Dienes, Z. (2011). Bayesian versus Orthodox statistics: Which side are you on? *Perspectives on Psychological Science, 6*, 274–290. doi: 10.1177/1745691611406920

Friston, K. J., Ashburner, J. T., Kiebel, S. J., Nichols, T. E., & Penny, W. D. (2007). Statistical parametric mapping: The analysis of functional brain images. In K. Friston, J. Ashburner, S. Kiebel, T. Nichols, & W. Penny (Eds.), *Statistical parametric mapping the analysis of functional brain images* (Vol. 8, p. 647). Academic Press. Q4

Gold, J. I., & Shadlen, M. N. (2001). Neural computations that underlie decisions about sensory stimuli. *Trends in Cognitive Sciences, 5*, 10–16. http://dx.doi.org/10.1016/S1364-6613(00)01567-9

Goodman, J. (1998). The interactive effects of task and external feedback on practice performance and learning. *Organizational Behavior and Human Decision Processes, 76*, 223–252.

Hajcak, G., McDonald, N., & Simons, R. F. (2003). To err is autonomic: Error-related brain potentials, ANS activity, and post-error compensatory behavior. *Psychophysiology, 40*, 895–903. doi: 10.1111/1469-8986.00107

Hajcak, G., & Simons, R. F. (2008). Oops!.. I did it again: An ERP and behavioral study of double-errors. *Brain and Cognition, 68*, 15–21. doi: 10.1016/j.bandc.2008.02.118

Holroyd, C. B., Nieuwenhuis, S., Yeung, N., Nystrom, L., Mars, R. B., Coles, M. G. H., & Cohen, J. D. (2004). Dorsal anterior cingulate cortex shows fMRI response to internal and external error signals. *Nature Neuroscience, 7*, 497–498.

Holroyd, C. B., Yeung, N., Coles, M. G. H., & Cohen, J. D. (2005). A mechanism for error detection in speeded response time tasks. *Journal of Experimental Psychology: General, 134*, 163–191. doi: 10.1037/0096-3445.134.2.163

Houtman, F., Castellar, E. N., Notebaert, W., & Nu, E. (2012). Orienting to errors with and without immediate feedback. *Journal of Cognitive Psychology, 24*, 37–41.

Huettel, S. A., & McCarthy, G. (2004). What is odd in the oddball task? *Neuropsychologia, 42*, 379–386. doi: 10.1016/j.neuropsychologia.2003.07.009

Jeffreys, H. (1961). *The theory of probability*. Oxford, UK: Oxford University Press.

Kerns, J. G., Cohen, J. D., MacDonald, A. W., Cho, R. Y., Stenger, V. A., & Carter, C. S. (2004). Anterior cingulate conflict monitoring and adjustments in control. *Science, 303*, 1023–1026. doi: 10.1126/science.1089910

King, J. A., Korb, F. M., Von Cramon, D. Y., & Ullsperger, M. (2010). Post-error behavioral adjustments are facilitated by activation and suppression of task-relevant and task-irrelevant information processing. *Journal of Neuroscience, 30*, 12759–12769.

Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin, 119*, 254–284. doi: 10.1037/0033-2909.119.2.254

Laming, D. (1979). Choice reaction performance following an error. *Acta Psychologica, 43*, 199–224. doi: 10.1016/0001-6918(79)90026-X

Laming, D. R. J. (1968). *Information theory of choice-reaction times*. London, UK: Academic Press.

MacLeod, C., & MacDonald, P. (2000). Interdimensional interference in the Stroop effect: Uncovering the cognitive and neural anatomy of attention. *Trends in Cognitive Sciences, 4*, 383–391.

Marco-Pallarés, J., Camara, E., Münte, T. F., & Rodríguez-Fornells, A. (2008). Neural mechanisms underlying adaptive actions after slips. *Journal of Cognitive Neuroscience, 20*, 1595–1610. doi: 10.1162/jocn.2008.20117

Notebaert, W., Houtman, F., Opstal, F. V., Gevers, W., Fias, W., & Verguts, T. (2009). Post-error slowing: An orienting account. *Cognition, 111*, 275–279. doi: 10.1016/j.cognition.2009.02.002

Rabbitt, P. (1969). Psychological refractory delay and response-stimulus interval duration in serial, choice-response tasks. *Acta Psychologica, 30*, 195–219.

Rabbitt, P., & Rodgers, B. (1977). What does a man do after he makes an error? An analysis of response programming. *The Quarterly Journal of Experimental Psychology, 29*, 727–743. Psychology Press.

Ridderinkhof, K. R., Ullsperger, M., Crone, E. A., & Nieuwenhuis, S. (2004). The role of the medial frontal cortex in cognitive control. *Science, 306*, 443–447. doi: 10.1126/science.1100301

Schmidt, R. A., Young, D. E., Swinnen, S., & Shapiro, D. C. (1989). Summary knowledge of results for skill acquisition: Support for the guidance hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*, 352–359. doi: 10.1037/0278-7393.15.2.352

Shannon, C. E., & Weaver, W. (1963). *The mathematical theory of communication (first published in 1949)*. Champaign, IL: University of Illinois Press.

Ullsperger, M., Nittono, H., & von Cramon, D. Y. (2007). When goals are missed: Dealing with self-generated and externally induced failure. *NeuroImage, 35*, 1356–1364. doi: 10.1016/j.neuroimage.2007.01.026

Ullsperger, M., & Von Cramon, D. Y. (2003). Error monitoring using external feedback: Specific roles of the habenular complex, the reward system, and the cingulate motor area revealed by functional magnetic resonance imaging. *Journal of Neuroscience, 23*, 4308–4314.

Van-Dijk, D., & Kluger, A. N. (2004). Feedback sign effect on motivation: Is it moderated by regulatory focus? *Applied Psychology, 53*, 113–135. doi: 10.1111/j.1464-0597.2004.00163.x

Wade, T. C. (1974). Relative effects on performance and motivation of self-monitoring correct and incorrect responses. *Experimental Psychology, 103*, 245–248.

Wetzels, R., & Wagenmakers, E. (2012). A Default Bayesian Hypothesis Test for correlations and partial correlations. *Psychonomic Bulletin & Review, 66*, 104–111. doi: 10.3758/s13423-012-0295-x

Received March 1, 2013
Revision received July 8, 2013
Accepted August 12, 2013

Sara Bengtsson

Department of Clinical Neuroscience
Karolinska Institutet
Retziusv 8, A2:A3
171 65 Stockholm
Sweden
E-mail Sara.Bengtsson@ki.se