Awareness of errors and feedback in human time estimation

Farah Bader[1,2] and Martin Wiener[2]

[1]Interdisciplinary Program in Neuroscience, George Mason University, Fairfax, Virginia 22032, USA; [2]Department of Psychology, George Mason University, Fairfax, Virginia 22032, USA

Behavioral and electrophysiology studies have shown that humans possess a certain self-awareness of their individual timing ability. However, conflicting reports raise concerns about whether humans can discern the direction of their timing error, calling into question the extent of this timing awareness. To understand the depth of this ability, the impact of nondirectional feedback and reinforcement learning on time perception were examined in a unique temporal reproduction paradigm that involved a mixed set of interval durations and the opportunity to repeat every trial immediately after receiving feedback, essentially allowing a "redo." Within this task, we tested two groups of participants on versions where nondirectional feedback was provided after every response, or not provided at all. Participants in both groups demonstrated reduced central tendency and exhibited significantly greater accuracy in the redo trial temporal estimates, showcasing metacognitive ability, and an inherent capacity to adjust temporal responses despite the lack of directional information or any feedback at all. Additionally, the feedback group also exhibited an increase in the precision of responses on the redo trials, an effect not observed in the no-feedback group, suggesting that feedback may specifically reduce noise when making a temporal estimate. These findings enhance our understanding of timing self-awareness and can provide insight into what may transpire when this is disrupted.

The tendency to reflect on one's mental state, or metacognition, is an essential human trait that is a central component of consciousness, memory processing, language, and decision-making, and, most importantly, time perception (Fleming and Dolan 2012; Yeung and Summerfield 2012). Accurate time perception is also a hallmark of consciousness and critical for everyday behaviors and cognitive functions ranging from speech, motor control, adaptive behavior, and survival (Meck 2005; Grondin 2010). Self-assessment of one's own timing ability without any external feedback, known as temporal metacognition, is vital for reliably determining temporal accuracy and variance despite uncertainty (Balcı et al. 2011; Lamotte et al. 2012; Akdoğan and Balcı 2017). Past research shows that humans and rodents can successfully incorporate the endogenous timing uncertainty and trial-by-trial variability associated with time perception tasks into their behavioral responses in a way that maximizes performance, updates temporal representations, and boosts reward (Balcı et al. 2011; Li and Dudman 2013). Human performance is measured via interval timing tasks that instruct participants to estimate temporal intervals of several seconds (Buhusi and Meck 2005). One particular task, temporal reproduction, involves exposure to a specific duration (encoding) and then an opportunity to recreate the interval duration via keypress (reproduction), with or without feedback. Previous research has shown that human subjects can accurately infer the distribution of intervals presented, and use this information to guide reproductions in the phase of measurement uncertainty (Jazayeri and Shadlen 2010; Acerbi et al. 2012). While the evidence already points toward an existing self-awareness of time, evaluating timing aptitude in context of feedback and learning may broaden our understanding of internal metacognitive process and its role in time perception.

Detecting and correcting errors is an important component for metacognition and self-awareness of one's cognitive state

(Fleming and Dolan 2012; Yeung and Summerfield 2012). The brain's performance monitoring system is responsible for assessing and minimizing these errors and facilitating the selection of an appropriate motor program to successfully complete the chosen task or behavior (Ullsperger et al. 2014). Error tracking mechanisms have been reported for temporal, numerosity, and spatial errors, implying that there may be a common metric error-monitoring system that underpins magnitude-based representations (Duyan and Balcı 2019). In particular, how errors related to early or late timing either with or without external feedback are managed is not fully understood. Studies on the impact of feedback on time perception have produced conflicting results, particularly due to the variability in the type of feedback delivery. Experimental timing paradigms offer a broad array of options ranging from no feedback, magnitude and directional feedback, magnitude-based feedback only, or directional-only feedback. Our study is an initial assessment of whether there is self-awareness of directional temporal information and compares the no feedback and magnitude-based conditions, serving as a launching point for further studies.

Akdoğan and Balcı (2017) used a temporal reproduction task to assess how subjects reproduced a range of suprasecond intervals; their findings demonstrated that humans are aware of both the magnitude and direction (early/late) of their timing errors despite not receiving any external feedback. Another recent behavioral study used a temporal production task, in which subjects were asked to repeatedly produce a single duration (3 sec) and compared performance during a condition when only the magnitude of the error was given (absolute) against another condition in which both the magnitude and direction (signed) were given (Riemer et al. 2019). Signed feedback delivery yielded more behavioral

Corresponding author: fbader2@masonlive.gmu.edu

adjustments in opposition to the direction of the error in subsequent trials, reduced bias in temporal estimates, and produced a more accurate and better calibrated performance when compared with absolute feedback. This study illustrated that directional information was not intrinsically accessible to the subject and that the participant's internal timing error representation failed to include that error direction. Furthermore, subjects assigned to the absolute feedback group also tended to report an overreproduction of the interval duration when in reality, they were underreproducing (Riemer et al. 2019). A key difference should be noted between these two studies, notably that feedback and retrospective self-judgments respectively were given following an entire block (Riemer et al. 2019) rather than trial by trial as in the Akdoğan and Balcı (2017) experiment. Task context also played a crucial role, as different tasks were used for the two studies; Akdoğan and Balcı (2017) used a temporal reproduction with a mixed set of intervals while Riemer et al. (2019) used a temporal production task with a singly presented interval.

In addition to simply supplying knowledge and guidance about the response (Salmoni et al. 1984), feedback has numerous other uses. It reduces response drifts over the experimental trajectory (Salmoni et al. 1984; Riemer et al. 2019) and may be erroneous or correct, but our study concentrates on correct feedback that tends to positively adjust behavioral responses (Salmoni et al. 1984). Its delivery may be absolute—after every trial or on a percentage of trials (relative). Additionally, feedback can be a motivational factor and act as an implicit reward for behavioral learning (Salmoni et al. 1984; Tsukamoto et al. 2006). Posterror feedback can also facilitate the learning of time intervals (Ryan and Robey 2002). Feedback may be processed differently depending on the quality of the learners and is reflected in a well-functioning performance monitoring system (Luft et al. 2013).

Numerous timing studies demonstrate that participants are aware of their errors prior to or independent of the administration of feedback (Akdoğan and Balcı 2017; Brocas et al. 2018; Kononowicz et al. 2018). In a recent M/EEG study using a temporal production task with the objective of repeatedly producing the same single interval duration, Kononowicz et al. (2018) asked respondents to first judge their own performance. Afterwards, they were provided with 100% directional feedback in two blocks out of the six and 15% feedback on the remaining blocks. The initial self-assessment of their performance prior to feedback matched the true interval duration and β power operated as an index of the actual duration and the self-evaluative ability to track timing errors (Kononowicz et al. 2018). In yet another single duration temporal production study, Brocas et al. (2018) tasked participants to generate interval durations of 30+ sec repeatedly for 10 trials and introduced a reward scheme to incentivize making accurate temporal estimates. Similarly to Riemer's paradigm, participants accurately self-evaluated their performance after a block of trials and correctly identified what proportion of trials were above or below the target interval duration (Brocas et al. 2018). Although accurate in their assessments of bias or tendency to overestimate or underestimate, the participants were less successful in prediction beforehand or correction of their responses (Brocas et al. 2018).

Self-knowledge about your internal timing behavior also incorporates reinforcement learning, another adaptive behavioral operation that integrates previous behavioral experiences and applies them to future scenarios to improve outcomes and maximize future rewards (Lee et al. 2012). Predicting the value of a set action along with its outcome and determining whether it will be awarded involves precise timing mechanisms (Petter et al. 2018; Mikhael and Gershman 2019). Errors are prone to occur in this process, particularly when there is a mismatch between the expected and actual outcomes, manifesting itself in the form of a reward prediction error (RPE) (Hollerman and Schultz 1998).

Notwithstanding errors, the process of learning interval durations transpires fairly quickly and is achieved with only one trial (Simen et al. 2011). To measure the speed of temporal learning, Simen et al. (2011) devised the "beat the clock task," a paradigm where a green square is displayed on a computer screen for an unknown amount of time. It remains onscreen until the appearance of a red square outline signals that the interval is ending. Participants must respond with a keypress before the stimulus terminates and are rewarded for on-time responses. Larger rewards are delivered the closer a participant makes an on-time response (Simen et al. 2011). Although the task comprises mixed interval durations that transition rapidly, early responding is minimized and on-time responding becomes the norm as participants learn the structure of the task. It is noteworthy that despite the endogenous timing uncertainty stemming from the presentation of rapidly changing durations, performance improvement was not impeded (Simen et al. 2011). In fact, all beat the clock participants improved in response times after only a single trial of beat-the-clock task, quickly reduced their timing errors, and were able to implement a strategy that facilitated learning interval durations (Simen et al. 2011).

Using an appropriate feedback technique is pivotal to understanding self-timing awareness. The traditional feedback structure in psychophysics studies is best suited for single interval reproductions because the same interval is successively reproduced; therefore, the majority of the studies described above are single interval experiments. Challenges arise when reproducing a random set of a mixed range of intervals because corrective guidance is given on one interval duration, yet the next interval may be of a different duration (Ryan 2016). This is problematic since there is no opportunity to use the feedback from the previous trial to the new trial, so the feedback is frequently misapplied to an entirely different duration (Ryan 2016). To rectify this issue, we introduced a "redo" trial, which allows the subject to use the original feedback from the first trial in a second trial of the same duration. We hypothesized that the redo trials will be beneficial to subjects for improving performance and in minimizing the Vierordt effect (underestimation of long intervals and overestimation of short intervals) (Ryan 2016). It is for this reason our study also incorporated absolute (nondirectional) feedback: to assess whether subjects possess awareness of the direction of the timing errors. If there is a substantial directional awareness of timing error, then the central tendency would be reduced.

Results

Experiment I

Subjects administered the feedback version of the temporal reproduction task exhibited central tendency with overestimation of short durations and underestimations of long durations. The accuracy of the reproduced times for the first and second trials was compared and a significant main effect of duration ($F_{(2.12, 33.92)} = 291.655$, $P < 0.001$, $\eta = 0.948$), of the trial ($F_{(1,16)} = 12.134$, $P = 0.03$, $\eta = 0.431$) and a trial × duration interaction ($F_{(2.297, 36.758)} = 3.527$, $P = 0.034$, $\eta = 0.181$) were observed. Second trial estimates were more accurate and closer to the target durations, and there was less uncertainty in time estimation, as revealed by nearer estimations to the reference line or identity line (Fig. 1A).

The second trial time estimates were also more precise ($F_{(1,16)} = 9.398$, $P = 0.007$, $\eta = 0.370$), as measured by the lower coefficient of variation in the second trial estimates (Fig. 1B). Similar to accuracy, precision also exhibited a main effect of duration ($F_{(1.553, 24.841)} = 8.862$, $P = 0.002$, $\eta = 0.356$) (Fig. 1B). Individual slope changes by trial type were compared and shown to be significantly higher $t_{(16)} = -2.692$, $P = 0.016$ in the second trial and above the identity
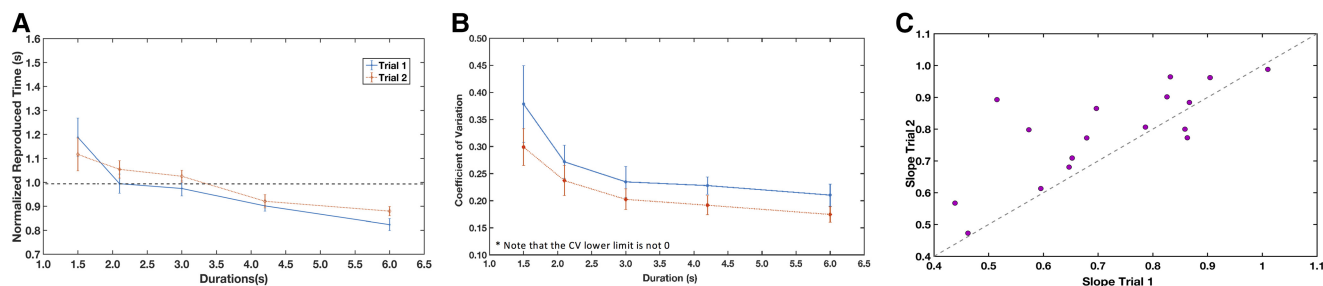
Figure 1. (*A*) Reproduction accuracy for on/off feedback condition: Participants exhibited significantly more accurate reproductions in the second trial compared with the first trial. Trial 1 is represented by the blue lines while trial 2 is represented by the dotted red lines. The dotted black horizontal line represents a perfect reproduction. Data are normalized and expressed in reproduced time/target duration ± SEM. (*B*) Temporal precision for on/off feedback condition participants were significantly more precise (lower coefficient of variations) in the second trial compared with the first trial. Trial 1 is represented by the blue lines while trial 2 is represented by the dotted red lines. Data are expressed in the coefficient of variation ± SEM. (*C*) Temporal certainty for on/off feedback condition. The slope of the second trial was significantly higher than the first trial and above the identity line. Data expressed in scatter plot with *x* and *y* coordinates corresponding to the first and second trial slopes, respectively of each individual participant.

line, demonstrating more certainty in the estimates and less bias (Fig. 1C).

Further dissection of the second trial reproduction estimates based on first trial feedback (on/off) was performed by separating the second trial reproduction time estimates based on whether the first trial was an off-target or on-target estimate. The redo trial estimates following positive feedback (on-target) time estimates trended closer to the target durations and identity line than negative (off-target feedback) but were not significantly more accurate $(F_{(1,16)} = 1.369, P = 0.259, \eta = 0.079)$, unlike duration $(F_{(2.623,41.964)} = 163.236, P < 0.001, \eta = 0.91)$ and trial × duration $(F_{(2.833,45.322)} = 5.434, P = 0.003, \eta = 0.254)$ for the second trial, which did significantly differ between feedback types (Fig. 2A).

Positive feedback in the first trial resulted in significantly more precise estimates (lower CVs) in the second trial $(F_{(1,16)} = 8.106, P = 0.012, \eta = 0.336)$ performance than negative feedback (Fig. 2B). The individual slopes of the redo trials $t_{(16)} = 4.699, P < 0.001$ between the first trial positive (on) and negative feedback (off) also varied significantly, demonstrating more certainty in temporal estimation following positive feedback (Fig. 2C).

Experiment II

Similar to experiment I, experiment II subjects exhibited central tendency with subjects overestimating short time intervals and un-

derestimating long time intervals. More accurate reproduced temporal estimates in the second trial for the no-feedback group signaled that they were reproducing the time closer to the target duration with significant differences between the original and redo trials $(F_{(1,18)} = 6.2, P = 0.023, \eta = 0.256)$ (Fig. 3A).

Duration $(F_{(1.49,26.92)} = 123.03, P < 0.01, \eta = 0.872)$ and the duration × trial interaction $(F_{(3.353,60.346)} = 2.786, P = 0.043, \eta = 0.134)$ also varied significantly between the trials despite the lack of feedback (Fig. 3A). Although more temporally accurate, subjects failed to show greater temporal precision, as evidenced by CVs in the redo trials of the no-feedback version of the task $F_{(1,18)} = 1.672, P = 0.212, \eta = 0.085$ (Fig. 3B). No significant differences in precision were observed in the duration $(F_{(4,72)} = 1.921, P = 0.116, \eta = 0.096)$ or duration × trial $(F_{(2.93,52.73)} = 2.583, P = 0.064, \eta = 0.125)$ (Fig. 3B).

The individual slopes of the redo trials displayed significantly greater certainty in the second trial temporal reproductions $t_{(18)} = 2.33, P = 0.021$ despite the lack of feedback (Fig. 3C).

Experiment I and II comparisons

For each subject participating in the on/off feedback and the no-feedback experiments, individual slope values for the second trial reproduced times were subtracted from the first trial reproduced interval estimates to compute Δ slope values
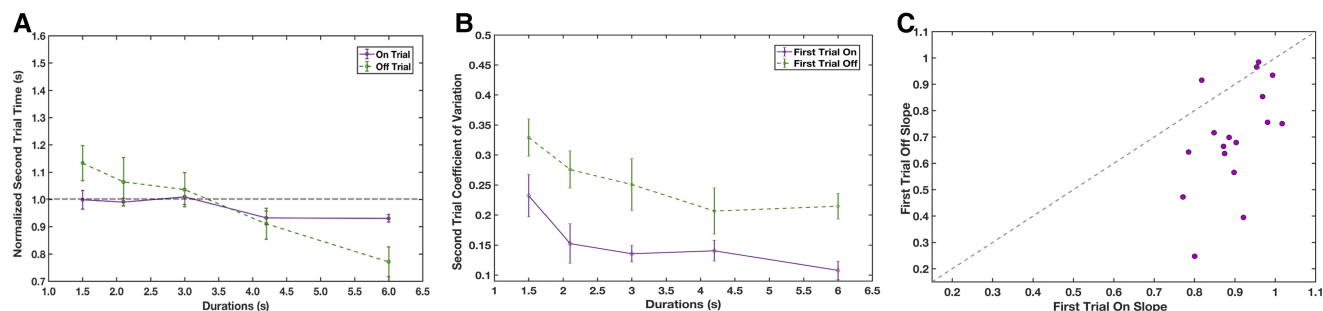

Figure 2. (*A*) Second trial accuracy based on first trial performance for on/off feedback condition positive feedback (on-target) trended toward the target durations but was not significantly different from the negative (off-target) feedback. Positive feedback is represented by the solid purple and the negative feedback by the dotted green lines. The horizonal black dotted line represents a perfect reproduction. Data are normalized and expressed in reproduced time/target duration ± SEM. (*B*) Second trial precision depends on first trial performance for on/off feedback condition positive (on-target) feedback in the first trial led to significantly higher precision (lower CVs) in the second trial. Positive feedback is represented by the solid purple and the negative feedback by the dotted green lines. Data are expressed in CV ± SEM. (*C*) Second trial slopes depend on first trial performance for on/off feedback condition slopes associated with positive feedback (on-target) in the first trial were significantly different from negative feedback (off-target). Data expressed in scatter plot with *x* and *y* coordinates corresponding to on-target and off-target slopes, respectively of each individual. The reference line represents a slope = 1.
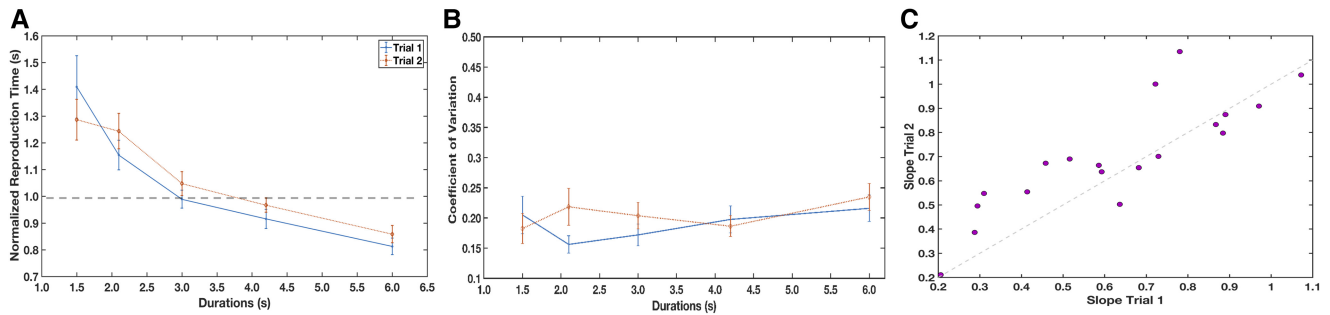
Figure 3. (*A*) Reproduction accuracy for no feedback condition second trial duration reproductions were significantly more accurate than the first trial reproductions. Trial 1 is represented by the blue lines while trial 2 is represented by the dotted red lines. The dotted horizontal black line represents a perfect reproduction with each reproduced duration matching the target durations. Data are normalized and expressed in reproduced time/target duration ± SEM. (*B*) Temporal precision for no feedback condition participants were not significantly more precise in the second trial compared with the first trial. Trial 1 is represented by the blue lines while trial 2 is represented by the dotted red lines. Data are expressed in the coefficient of variation ± SEM. (*C*) Slopes by trial for no feedback condition the second trial slopes were significantly higher from the first trial slopes with more participants above the identity line. Data expressed in a scatter plot with *X* and *Y* coordinates corresponding to the first and second trial slopes, respectively of each individual. The reference line represents a slope = 1.

(slope$_{trial2}$ – slope$_{trial1}$). The Δ precision was also calculated for each subject (CV$_{trial2}$ – CV$_{trial1}$) in experiments I and II. Notably, the on/off feedback group was significantly more precise (lower CV) than the no-feedback condition $t_{(34)} = 3.234$, $P = 0.003$ (CI: 0.0223–0.09784) (Fig. 4A).

Due to the presence of outliers in the CVs, a Mann–Whitney *U* was performed ($U = 62$, $z = -3.15$, $P = 0.002$), further confirming the previous independent *t*-test results of significantly greater precision in the on/off feedback group.

An independent sample *t*-test also revealed no significant differences in the Δ slope values between experiments, $t_{(34)} = 0.021$, $P = 0.983$ (CI: −0.085–0.866) (Fig. 4B).

Discussion

Our study demonstrated the beneficial impacts of feedback when generating temporal estimates of short visual durations in the temporal reproduction task. Estimates tended to be more accurate and precise and reflected greater certainty when post-trial feedback indicated whether the response was on- or off-target. Furthermore, initial on-target feedback improved the precision and certainty of the second trial time estimation, as reflected by a reduced central tendency effect. The significant interaction effect of trial × duration further showed decreased central tendency and enhanced certainty that participants had in their second trial temporal estimates, particularly when they were provided with on-target feedback in the first trial. Crucially, these benefits persisted despite the fact that the feedback lacked informational content on direction indicating whether it was an underreproduction or overreproduction of the interval duration.

Even without feedback, participants improved in the second redo trials and demonstrated a smaller central tendency

effect, suggesting that subjects are indeed metacognitively aware of the direction of their errors. Again, the trial × duration interaction for the no-feedback group was significant, indicative of less uncertainty in the temporal judgements and lowered central tendency producing a less noticeable, muted expression of Vierordt's law. However, precision did not improve when subjects did not receive feedback, signaling a more noisy estimate in the absence of feedback. Further comparison between the two groups
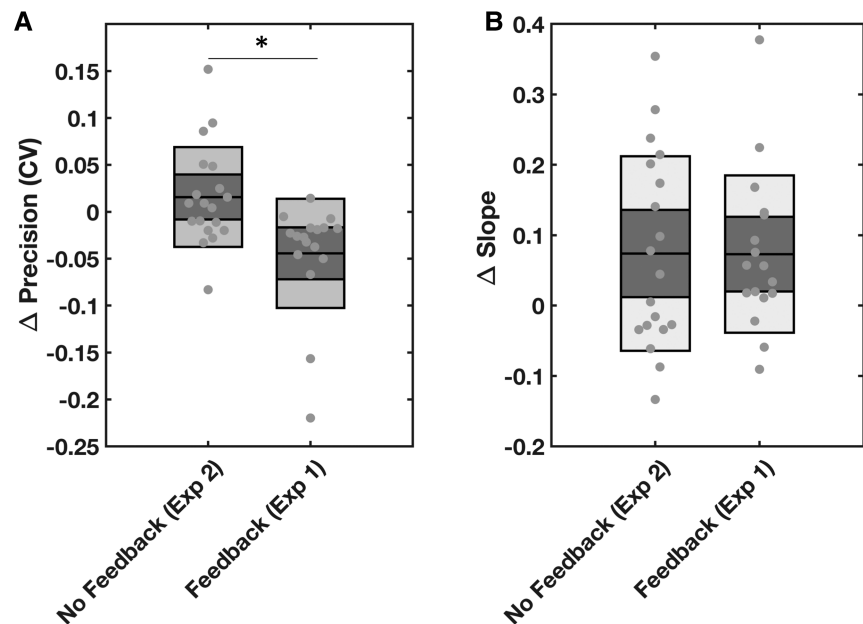

Figure 4. (*A*) Δ Precision between on/off and no feedback conditions. Participants in the on/off feedback condition exhibited significantly reduced CVs (higher precision) when compared with the no-feedback condition. The solid black horizontal lines represent the averages of the Δ precision values, 95% confidence intervals are represented by the outer boxes, and ±one standard deviation are represented by the inner boxes. Individual dots represent both sets of participants from the two conditions. The asterisk indicates a significant difference at *P* < 0.05. (*B*) Δ Slopes between on/off and no feedback conditions. The differences in slopes between the second and first trials (Δ slope) were not significantly different when the on/off and no feedback conditions were compared. The solid black horizontal lines represent the averages of the Δ slopes, 95% confidence intervals are represented by the outer boxes, and ±one standard deviation are represented by the inner boxes. Individual dots represent both sets of participants from the two conditions.

(experiments I and II) revealed that only precision differed significantly between the two conditions of feedback and no feedback, highlighting that feedback is integral for making us more precise by potentially reducing measurement noise.

Notably, other studies confirm the role of feedback in enhancing the precision and essentially reducing the behavioral variability of temporal estimates. Recently, researchers showed that the rewarding or nonrewarding nature of the previous trial outcome causally controls the current trial behavioral variability in a ready-set-go paradigm, a motor context-dependent timing task where subjects were required to flexibly produce timing intervals using either a keypress or eye movement (Wang et al. 2020). When given probabilistic feedback, the participants' timing variability was higher for incorrect trials than correct trials regardless of the error size (Wang et al. 2020). This finding aligned well with our own study's results demonstrating that the second trial performance following previous on-target feedback improved temporal precision. Further evidence of feedback's beneficial effect on precision comes from a study that examined bias and variance changes for three time estimation tasks with a single interval design—motor reproduction, auditory comparison (duration discrimination), and auditory reproduction—and found that in all three tasks, participants overestimated time durations (Shi et al. 2013). Auditory feedback, included as a component of the auditory reproduction task, produced a lower overestimation bias and variance than the motor reproduction task, although the bias still surpassed the auditory comparison task. In this same study, the signal to noise (SNR) ratios were determined by varying the decibel levels of the comparison and feedback tones thus altering the comparison/feedback ratio and simultaneously introducing pink noise to produce a varied SNR. This was later manipulated to yield low and high SNR conditions, and the variances on the reproduced interval, measured by standard deviations, were reduced in the high SNR condition compared with the low SNR in the same auditory reproduction task (Shi et al. 2013). Similar benefits of auditory feedback have also been observed in the experiment performed by Mitani and Kashino (2018), which required participants to reproduce the duration of a single tone after hearing it twice successively. Feedback was then delivered and participants indicated whether they were early or late in responding. Bias and variability improved, self-judgement of timing error matched the actual temporal reproductions, and serial dependency was dampened but only for subsecond rather than suprasecond intervals (Mitani and Kashino 2018). Our study extends the findings of these two studies to the visual modality and demonstrates a benefit in reducing timing variability in the longer suprasecond range.

The type of feedback that is delivered plays a pivotal role in outcome. From our experiment I, the presence of positive on-target feedback specifically led to more precise temporal reproductions. A paradigm that can explain this behavior is the explore and exploit trade-off, an algorithm used in foraging related decision-making, which calculates the trade-off between selecting a choice with a known reward value rather than an alternative uncertain reward option with unknown value (Addicott et al. 2017). Applied to our study, with positive feedback, participants exploited the known reward and temporal structure of the task to improve their temporal precision in their second redo trials (Navarro et al. 2016). Conversely, when given negative feedback, subjects elected to "explore" other options and thus produced temporally variable reproductions when allowed to redo (Navarro et al. 2016). Dopamine has been linked to both choices in the explore/exploit paradigm; at the subcortical level, genes controlling striatal DA function have been connected to the "exploit" selection and prefrontal DA function in the *catechol-O-methyltransferase* (COMT) gene has been associated with the "explore" selection (Frank et al. 2009). Dopaminergic transmission has also been linked to timing and re-

inforcement learning due to sharing of striatal circuitry and association with reward (Jahanshahi et al. 2006; Coull and Nobre 2008). Reward prediction errors (RPE) rely on the same dopaminergic circuitry; therefore, an increase or decrease in DA activation is reflective of positive and negative predictive errors, respectively (Schultz et al. 1997; d'Ardenne et al. 2008) and can influence perceived duration (Soares et al. 2016) and by extension, it may impact our self-awareness of timing behavior. The RPE itself may operate as a temporal feedback signal designed to modulate the internal clock and generate subsequent compensatory timing behavior. Related to this connection, Toren et al. (2020) recently performed a fMRI study with the time discrimination paradigm to explore how dopamine-mediated prediction error distorts time, revealing that negative prediction errors compress perceived duration while positive prediction errors dilate the perceived duration. This discovery was linked to activity in the putamen (Toren et al. 2020). Dopamine is not the only neurotransmitter involved in meting out the careful balance of exploration and exploitation. Human gaze pattern and pupillometry studies focused on attentional task shifting have also demonstrated a role for the tonic firing of noradrenaline in the locus coeruleus during exploration (Pajkossy et al. 2017). Specifically, human pharmacological studies where humans were administered norepinephrine transporter blockers showed a reduction of random rather than targeted exploration in gambling tasks (Warren et al. 2017).

The findings of our study have implications for a more extensive understanding of clinical conditions when there is a disruption of the self-awareness of timing behavior. Doenyas et al. (2019) reported on autistic spectrum disorder (ASD) children who performed a no-feedback version of the auditory temporal reproduction task that also used confidence ratings and self-assessment of underreproductions or overreproductions. Performance-wise, ASD children achieved the task objectives and the mean reproduction accuracy and coefficient of variation demonstrated similarity to neurotypical children (Doenyas et al. 2019). However, there was a minimal match between the level of subjective self-confidence and actual objective performance in ASD children, suggesting that ASD children were unaware of temporal errors (Doenyas et al. 2019).

A few limitations of our study should be noted. Our study had a limited number of feedback conditions; therefore, future studies could include an additional feedback condition with directional (signed) feedback that informs the participant whether it is an underreproduction or overreproduction or whether the response is early or late. Comparing directional feedback with our existing nondirectional and no-feedback groups would further solidify and strengthen previous findings by Akdoğan and Balcı (2017) and satisfy the question of whether the internal representation of error includes directional information in addition to magnitude-based representation. Riemer et al. (2019) also tackled this query but used a different task (temporal production) and with a single interval (only 3 sec) rather than a mixed set. Participants in their study were given both absolute (magnitude of error only) and signed (magnitude and direction) feedback delivery; however, the signed feedback surpassed the absolute feedback group with regards to accuracy, bias, and behavioral adjustments in opposition to the directional error (Riemer et al. 2019).

Assessing participants' self-confidence in their responses is also a critical item to examine in order to fully evaluate the degree of metacognition. Previous studies have commonly included a post-trial self-assessment of participant performance prior to the provision of external feedback, the purpose of which is to correlate the level of confidence to the magnitude and the direction of the error in timing (e.g., low confidence and huge error). The next iteration of our study could integrate an assessment of confidence to reinforce our claim of self-awareness of timing errors. Allowing

subjects to use a sliding scale to mark how close they are to the target duration has been successfully used in prior studies, such as the M/EEG study of self-awareness during temporal production (Kononowicz et al. 2018) or a more recent experiment investigating the influence of social modulation of being observed while performing the timing reproduction task (Öztel et al. 2020). Finally, as the auditory modality has greater temporal resolution than the visual domain (Kanabus et al. 2002) and is the privileged, dominant sensory domain in the area of time perception (Guttman et al. 2005; Burr et al. 2009; Kanai et al. 2011), future studies may entail an auditory version of the temporal reproduction task.

Overall, our study deepens our understanding of self-timing awareness and shows that if corrective nondirectional feedback is delivered on and tested on the same duration, it can be used to improve time perception and reduce the central tendency in mixed interval duration sets. Timing performance, excluding timing variability, remains uncompromised even in the absence of any feedback. This is supported by human studies that have confirmed the ability to ascertain timing errors with confidence (Akdoğan and Balcı 2017) and animal studies that reveal an innate awareness of temporal accuracy and prediction of success or failure in a uncertain environments (Balcı et al. 2009). Studying the impact of feedback and reinforcement learning on time contributes to a more nuanced appreciation of temporal processing and error monitoring in neurotypical populations, which can then potentially inform us about what transpires when timing mechanisms are disrupted.

Materials and Methods

Subjects
George Mason University undergraduates were recruited via flyers and through a posting in the George Mason research studies database. Eligible participants who completed the research study received undergraduate psychology course credit for their participation. Researchers gave each subject a questionnaire to determine eligibility prior to beginning the experiment. Participants with any neurological and psychological disorders, hospitalization for a psychological disorder, or diagnosis or treatment for substance abuse were excluded.

Twenty subjects each were originally recruited for experiment I, which provided on- and off-target feedback. For the final analysis, data from only 17 right-handed neurologically healthy subjects (average age 20.8 yr, five males, 3.84 = SD) in experiment I were analyzed. Two subjects did not fully understand the task parameters and one recruited subject later informed us that she had a concussion so these three were excluded.

Twenty right-handed subjects were also recruited for experiment II (no feedback); however, for the final data analysis, 19 right-handed subjects (average age 19.6 yr, six males, 1.53 SD) were analyzed because one subject had outlier responses that varied three standard deviations from the average mean responses.

The difference in ages between experiments I and II was also not significant $t_{(34)} = 1.196$, $P = 0.245$. Additionally, a Pearson χ^2 revealed that the gender ratio was not significantly different between the on/off and no feedback groups $\chi(1) = 0.020$, $P = 0.888$.

Task paradigm
The temporal reproduction task was delivered via Psychopy2 on a 27-in Mac desktop while subjects sat ~60 cm from the computer screen (Dell S2716DGR, 120 Hz refresh rate). The task structure was comprised of three phases: estimation, reproduction, and feedback. These three phases were performed twice for each duration. Each trial initiated with a centrally presented fixation cross for a randomly presented duration of 2–6 sec. In the estimation phase, a blue square was visually shown to the participant for one of five logarithmically spaced, randomly presented intervals (1.5–6 sec). Until the square was on-screen, the participant was instructed

to encode the duration in memory and to not use counting as a method to do so, which has been demonstrated as an effective means of eliminating counting strategies (Rattat and Droit-Volet 2012). Following the estimation phase, there was a 4- to 8-sec gap prior to the reproduction phase. Then, the blue square reappeared on-screen in the reproduction phase and the participant was asked to press any number key when the blue square has remained on-screen for the same time duration as the time elapsed in the estimation phase. This keypress caused the square to disappear, signaling interval termination (Fig. 5).

After every trial, adaptive feedback (duration = 1 sec) was delivered 2–4 sec after the disappearance of the square and informed the participant whether the response was on-target or off-target; notably, this feedback provided no index of direction. On each trial, a feedback constant (k), starting with an initial value of 3.5 was adjusted such that that the reproduced interval had to be within the window [interval/k] and was updated according to the 1-up/1-down rule with a step size of 0.015 (Jazayeri and Shadlen 2010). If the participant's reproduced interval was either 15% above or below the target duration, an on-target feedback would be delivered; otherwise, an off-target feedback was delivered. Critically, after each trial, participants had a second opportunity (essentially a redo trial) to perform the entire sequence of phases (estimation, reproduction, and feedback) again, ensuring feedback was applied to the appropriate duration.

In total, there are 120 trials (10 durations/block × s blocks × two trials). Participants were given a break after each block. We performed two versions of this task with one experimental condition with on/off feedback delivery and another experimental condition with no feedback at all. A different set of participants were used for each condition. In the first experiment, on and off-feedback was provided for both trials while the second experiment had two trials per duration but lacked any feedback.

Statistical analysis
SPSS 19.0 (IBM, SPSS) with α set to 0.05 was used to analyze the behavioral data from the temporal reproduction task in the two experimental groups independently and together. For both experiments I and II, the majority of the durations were normally distributed and passed the Shapiro–Wilks test of normality for both accuracy and precision (as measured by the coefficient of variation [CV]). The individual slopes and at each duration were also normally distributed for both experiment I and II. To measure accuracy, we calculated the mean reproduced duration/base duration to generate a normalized value. Underreproduction was graphically represented as less than one while values exceeding a value of one were displayed as overreproductions. The CVs was calculated as the standard deviation of the mean reproduced durations/participant's mean reproduced durations and two separate CV values were generated for the first and second trials. The slope was calculated from a regression line of the mean reproduced durations (y-values)/aggregated average of the sample target durations
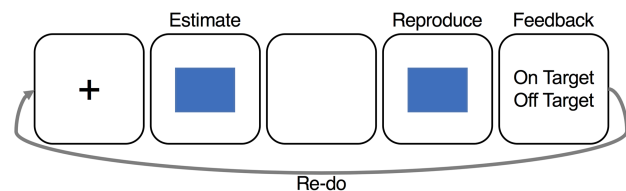

Figure 5. Task schematic for the stages of the task (encoding, reproduction, feedback). A blue square is displayed for a randomized duration from 1–6 sec and disappears on its own. A second blue square then appears and the participant must press a number key to make it disappear once the on-screen duration time matches. Feedback indicating whether the participant was on-target or off-target is then delivered. The participant then has a second opportunity to redo the trial in its entirety, with a second recurrence of the estimation and reproduction stimulus, another keypress response to terminate the reproduction stimuli, and a postresponse feedback.

(*x*-values). Similarly to the CVs, two slope values representing the first and second trials were generated. Repeated measures ANOVAs were performed for accuracy and precision and paired *t*-tests for comparisons within the first and second trials. Separate analysis were conducted for each experiment. Then, comparisons between experiments I and II were performed and repeated-measures ANOVA was conducted again and the independent t-test was used for the change (Δ) in slopes and precision values between the first and second trials.

References

Acerbi L, Wolpert DM, Vijayakumar S. 2012. Internal representations of temporal statistics and feedback calibrate motor-sensory interval timing. *PLoS Comput Biol* **8:** e1002771. doi:10.1371/journal.pcbi.1002771

Addicott MA, Pearson JM, Sweitzer MM, Barack DL, Platt ML. 2017. A primer on foraging and the explore/exploit trade-off for psychiatry research. *Neuropsychopharmacology* **42:** 1931–1939. doi:10.1038/npp.2017.108

Akdoğan B, Balcı F. 2017. Are you early or late? Temporal error monitoring. *J Exp Psychol Gen* **146:** 347–361. doi:10.1037/xge0000265

Balcı F, Freestone D, Gallistel CR. 2009. Risk assessment in man and mouse. *Proc Natl Acad Sci* **106:** 2459–2463. doi:10.1073/pnas.0812709106

Balcı F, Freestone D, Simen P, DeSouza L, Cohen JD, Holmes P. 2011. Optimal temporal risk assessment. *Front Integr Neurosci* **5:** 56. doi:10.3389/fnint.2011.00056

Brocas I, Carrillo JD, Tarrasó J. 2018. Self-awareness of biases in time perception. *J Econ Behav Organ* **148:** 1–19. doi:10.1016/j.jebo.2018.02.001

Buhusi CV, Meck WH. 2005. What makes us tick? Functional and neural mechanisms of interval timing. *Nat Rev Neurosci* **6:** 755–765. doi:10.1038/nrn1764

Burr D, Banks MS, Morrone MC. 2009. Auditory dominance over vision in the perception of interval duration. *Exp Brain Res* **198:** 49–57. doi:10.1007/s00221-009-1933-z

Coull J, Nobre A. 2008. Dissociating explicit timing from temporal expectation with fMRI. *Curr Opin Neurobiol* **18:** 137–144. doi:10.1016/j.conb.2008.07.011

D'ardenne K, McClure SM, Nystrom LE, Cohen JD. 2008. BOLD responses reflecting dopaminergic signals in the human ventral tegmental area. *Science* **319:** 1264–1267. doi:10.1126/science.1150605

Doenyas C, Mutluer T, Genç E, Balcı F. 2019. Error monitoring in decision-making and timing is disrupted in autism spectrum disorder. *Autism Res* **12:** 239–248. doi:10.1002/aur.2041

Duyan YA, Balcı F. 2019. Metric error monitoring in the numerical estimates. *Conscious Cogn* **67:** 69–76. doi:10.1016/j.concog.2018.11.011

Fleming SM, Dolan RJ. 2012. The neural basis of metacognitive ability. *Philos Trans R Soc B Biol Sci* **367:** 1338–1349. doi:10.1098/rstb.2011.0417

Frank MJ, Doll BB, Oas-terpstra J, Moreno F. 2009. The neurogenetics of exploration and exploitation: prefrontal and striatal dopaminergic components. *Nat Neurosci* **12:** 1062–1068. doi:10.1038/nn.2342

Grondin S. 2010. Timing and time perception: a review of recent behavioral and neuroscience findings and theoretical directions. *Atten Percept Psychophys* **72:** 561–582. doi:10.3758/APP.72.3.561

Guttman SE, Gilroy LA, Blake R. 2005. Hearing what the eyes see: auditory encoding of visual temporal sequences. *Psychol Sci* **16:** 228–235. doi:10.1111/j.0956-7976.2005.00808.x

Hollerman JR, Schultz W. 1998. Dopamine neurons report an error in the temporal prediction of reward during learning. *Nat Neurosci* **1:** 304–309. doi:10.1038/1124

Jahanshahi M, Jones CR, Dirnberger G, Frith CD. 2006. The substantia nigra pars compacta and temporal processing. *J Neurosci* **26:** 12266–12273. doi:10.1523/JNEUROSCI.2540-06.2006

Jazayeri M, Shadlen MN. 2010. Temporal context calibrates interval timing. *Nat Neurosci* **13:** 1020–1026. doi:10.1038/nn.2590

Kanabus M, Szelag E, Rojek E, Pöppel E. 2002. Temporal order judgment for auditory and visual stimuli. *Acta Neurobiol Exp* **62:** 263–270.

Kanai R, Lloyd H, Bueti D, Walsh V. 2011. Modality-independent role of the primary auditory cortex in time estimation. *Exp Brain Res* **209:** 465–471. doi:10.1007/s00221-011-2577-3

Kononowicz TW, Roger C, van Wassenhove V. 2018. Temporal metacognition as the decoding of self-generated brain dynamics. *Cereb Cortex* **29:** 4366–4380. doi:10.1093/cercor/bhy318

Lamotte M, Izaute M, Droit-Volet S. 2012. Awareness of time distortions and its relation with time judgment: a metacognitive approach. *Conscious Cogn* **21:** 835–842. doi:10.1016/j.concog.2012.02.012

Lee D, Seo H, Jung MW. 2012. Neural basis of reinforcement learning and decision making. *Annu Rev Neurosci* **35:** 287–308. doi:10.1146/annurev-neuro-062111-150512

Li Y, Dudman JT. 2013. Mice infer probabilistic models for timing. *Proc Natl Acad Sci* **110:** 17154–17159. doi:10.1073/pnas.1310666110

Luft CDB, Nolte G, Bhattacharya J. 2013. High-learners present larger mid-frontal theta power and connectivity in response to incorrect performance feedback. *J Neurosci* **33:** 2029–2038. doi:10.1523/JNEUROSCI.2565-12.2013

Meck WH. 2005. Neuropsychology of timing and time perception. *Brain Cogn* **58:** 1–8. doi:10.1016/j.bandc.2004.09.004

Mikhael JG, Gershman SJ. 2019. Adapting the flow of time with dopamine. *J Neurophysiol* **121:** 1748–1760. doi:10.1152/jn.00817.2018

Mitani K, Kashino M. 2018. Auditory feedback assists post hoc error correction of temporal reproduction, and perception of self-produced time intervals in subsecond range. *Front Psychol* **8:** 2325. doi:10.3389/fpsyg.2017.02325

Navarro DJ, Newell BR, Schulze C. 2016. Learning and choosing in an uncertain world: an investigation of the explore–exploit dilemma in static and dynamic environments. *Cogn Psychol* **85:** 43–77. doi:10.1016/j.cogpsych.2016.01.001

Öztel T, Eskenazi T, Balcı F. 2020. Temporal error monitoring with directional error magnitude judgments: a robust phenomenon with no effect of being watched. *Psychol Res* doi:10.1007/s00426-020-01379-0

Pajkossy P, Szőllősi Á, Demeter G, Racsmány M. 2017. Tonic noradrenergic activity modulates explorative behavior and attentional set shifting: evidence from pupillometry and gaze pattern analysis. *Psychophysiology* **54:** 1839–1854. doi:10.1111/psyp.12964

Petter EA, Gershman SJ, Meck WH. 2018. Integrating models of interval timing and reinforcement learning. *Trends Cogn Sci* **22:** 911–922. doi:10.1016/j.tics.2018.08.004

Rattat AC, Droit-Volet S. 2012. What is the best and easiest method of preventing counting in different temporal tasks? *Behav Res Methods* **44:** 67–80. doi:10.3758/s13428-011-0135-3

Riemer M, Kubik V, Wolbers T. 2019. The effect of feedback on temporal error monitoring and timing behavior. *Behav Brain Res* **369:** 111929. doi:10.1016/j.bbr.2019.111929

Ryan L. 2016. Why doesn't feedback correct Vierordt's law? *J Cogn Psychol* **28:** 948–964. doi:10.1080/20445911.2016.1221829

Ryan LJ, Robey TB. 2002. Learning and performance effects of accurate and erroneous knowledge of results on time perception. *Acta Psychol* **111:** 83–100. doi:10.1016/S0001-6918(02)00044-6

Salmoni A, Schmidt RA, Walter CB. 1984. Knowledge of results and motor learning: a review and critical reappraisal. *Psychol Bull* **95:** 355–386. doi:10.1037/0033-2909.95.3.355

Schultz W, Dayan P, Montague PR. 1997. A neural substrate of prediction and reward. *Science* **275:** 1593–1599. doi:10.1126/science.275.5306.1593

Shi Z, Ganzenmüller S, Müller HJ. 2013. Reducing bias in auditory duration reproduction by integrating the reproduced signal. *PLoS One* **8:** e62065. doi:10.1371/journal.pone.0062065

Simen P, Balcı F, DeSouza L, Cohen JD, Holmes P. 2011. A model of interval timing by neural integration. *J Neurosci* **31:** 9238–9253. doi:10.1523/JNEUROSCI.3121-10.2011

Soares S, Atallah BV, Paton JJ. 2016. Midbrain dopamine neurons control judgment of time. *Science* **354:** 1273–1277. doi:10.1126/science.aah5234

Toren I, Aberg KC, Paz R. 2020. Prediction errors bidirectionally bias time perception. *Nat Neurosci* **23:** 1198–1202. doi:10.1038/s41593-020-0698-3

Tsukamoto T, Kotani Y, Ohgami Y, Omura K, Inoue Y, Aihara Y. 2006. Activation of insular cortex and subcortical regions related to feedback stimuli in a time estimation task: an fMRI study. *Neurosci Lett* **399:** 39–44. doi:10.1016/j.neulet.2006.01.061

Ullsperger M, Fischer AG, Nigbur R, Endrass T. 2014. Neural mechanisms and temporal dynamics of performance monitoring. *Trends Cogn Sci* **18:** 259–267. doi:10.1016/j.tics.2014.02.009

Wang J, Hosseini E, Meirhaeghe N, Akkad A, Jazayeri M. 2020. Reinforcement regulates timing variability in thalamus. *Elife* **9:** e55872. doi:10.7554/eLife.55872

Warren CM, Wilson RC, van der Wee NJ, Giltay EJ, van Noorden MS, Cohen JD, Nieuwenhuis S. 2017. The effect of atomoxetine on random and directed exploration in humans. *PLoS One* **12:** e0176034. doi:10.1371/journal.pone.0176034

Yeung N, Summerfield C. 2012. Metacognition in human decision-making: confidence and error monitoring. *Philos Trans R Soc B Biol Sci* **367:** 1310–1321. doi:10.1098/rstb.2011.0416