Check for updates

RESEARCH ARTICLE

REVISED Escalating frustration - A replication attempt and extension of Yu et al. (2014)

[version 2; peer review: 2 approved, 1 approved with reservations]

Charlotte Eben[1,2*], Zhang Chen[1*], Raquel E. London[1], Frederick Verbruggen[1,3]

[1]Department of Experimental Psychology, Ghent University, Ghent, Flanders, 9000, Belgium
[2]Department of Psychology, The University of British Columbia, Vancouver, British Columbia, V6T 1Z4, Canada
[3]Department of Biology, Ghent University, Ghent, Flanders, 9000, Belgium

* Equal contributors

v2 **First published:** 09 Jul 2024, **4**:135
https://doi.org/10.12688/openreseurope.17749.1
Latest published: 10 Apr 2025, **4**:135
https://doi.org/10.12688/openreseurope.17749.2

Abstract

Background

Failures to obtain a desired reward, such as losing money in gambling, can lead to frustration. In gambling, this frustration has been shown to take the form of faster responses after losses compared with wins and non-gambling trials. In addition, reward omission or blockage can lead to more forceful responses. Yu and colleagues (2014) showed that the proximity to a reward and the effort already expended to acquire the reward increased participants' response force and their retrospective self-reported frustration when the reward was blocked.

Methods

In this study, we attempted to replicate the findings of Yu and colleagues (2014) using the same experimental procedure. In each schedule, participants (N = 32) needed to complete an arrow direction task for varying numbers of times to win a reward but could be blocked at any stage. The response time (RT) and force of confirming the outcomes were used as indicators of 'frustration'. In addition, to obtain a more real-time and objective measure of (negative) emotion, we measured facial electromyographic (EMG) activity over the corrugator supercilii (frowning muscle) and the zygomaticus (smiling muscle).

Open Peer Review

Approval Status ✓ ✓ ?

	1	2	3
version 2 (revision) 10 Apr 2025	✓ view	✓ view	? view
version 1 09 Jul 2024	? view	? view	

1. **Sean Devine**, McGill University, Montreal, Canada
2. **Sebastian Korb**, University of Essex (Ringgold ID: 2591), Colchester, UK
University of Vienna, Vienna, Austria
3. **Aijaz A. Naik**, National Institute of Mental Health (NIMH), Bethesda, USA

Any reports and responses or comments on the article can be found at the end of the article.

Results

Due to technical problems, our data on response force were invalid. In line with the original study, both goal proximity and exerted effort increased participants' self-reported motivation in the task and frustration after being blocked. An exploratory analysis showed that unexpectedly participants were slower in confirming an outcome when they were blocked closer to the reward, while exerted effort did not influence the time taken to confirm the outcome. These RT data were consistent with self-reported surprise ratings, suggesting an orienting response. In the facial EMG data, we observed no difference between wins and losses in activity over the corrugator or the zygomaticus.

Conclusion

Taken together, these data suggest that reward blockage does not necessarily lead to behavioral or psychophysiological expressions of negative emotions such as frustration.

Keywords
frustration, reward, response force, facial EMG

This article is included in the Psychology gateway.

This article is included in the Horizon 2020 gateway.

This article is included in the European Research Council (ERC) gateway.

Corresponding authors: Charlotte Eben (ebencharlotte01@gmail.com), Zhang Chen (zhang.chen@ugent.be)

Author roles: Eben C: Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Software, Visualization, Writing – Original Draft Preparation; **Chen Z**: Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Software, Visualization, Writing – Original Draft Preparation; **London RE**: Formal Analysis, Software, Writing – Review & Editing; **Verbruggen F**: Conceptualization, Funding Acquisition, Project Administration, Resources, Supervision, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: This project has received funding from the ERC Consolidator grant awarded to FV (European Union's Horizon 2020 research and innovation programme, under grant agreement No [769595] [Ctrl-ImpAct]) and a Methusalem grant (Project 01M0022, Ghent University).

Copyright: © 2025 Eben C *et al*. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Eben C, Chen Z, London RE and Verbruggen F. **Escalating frustration - A replication attempt and extension of Yu et al. (2014) [version 2; peer review: 2 approved, 1 approved with reservations]** Open Research Europe 2025, **4**:135 https://doi.org/10.12688/openreseurope.17749.2

First published: 09 Jul 2024, **4**:135 https://doi.org/10.12688/openreseurope.17749.1

REVISED **Amendments from Version 1**

In response to reviewer feedback, several key changes were made to the manuscript. First, we incorporated recent studies on goal gradients in cognitive effort (e.g., Devine *et al.*, 2024; Katzir *et al.*, 2020). Additionally, we clarified the role of confirm RT and why we focused on this measure as part of the exploration of outcome-related frustration effects.

The manuscript also now provides a clearer sample size justification, explaining the choice of 32 participants based on a medium effect size assumption (Cohen's d = 0.5) and power considerations. Furthermore, we added that the use of multilevel models (MLMs) might have provided more statistical power than the original ANOVA approach.

For better comparison of effect sizes, we include now the original study plots alongside our data plots, showing similar effects. We also clarified in our discussion, which effects were exactly replicated and which effects we were not able to replicate, for better transparency.

We also refined explanations of the coin schedule structure, payout structure, and RT exclusion criteria. Finally, various figure updates were made, including clarifications to axes, labels, and the addition of variance indicators to EMG figures.

Any further responses from the reviewers can be found at the end of the article

Introduction

Humans and non-human animals generally try to achieve desired goals. Along the way, however, they may encounter obstacles and be prevented from achieving these goals: for example, in birds, a competitor may take a valuable piece of food before the individual has the opportunity to eat; or in humans, we may not get the promotion that we think we deserve. To understand goal-directed behaviour, it is important to understand how individuals respond to such failures to achieve desired goals.

Both examples above may lead to 'frustration': When an expected reward is omitted, individuals sometimes respond with more vigor (e.g., speed, force), which is called the 'frustration' effect. For example, Amsel (1958; 1992); observed that rats that were typically rewarded in a first runway of a maze ran through the second runway faster if the reward was not given in the first runway. Furthermore, he observed more aggressive behavior, such as urinating or biting the experimenter, after reward omission. Similar effects have been observed in other species, such as fish (Vindas *et al.*, 2012), mice (de Almeida & Miczek, 2002), laying hens (Zimmerman & Koene, 1998) and pigeons (Zentall, 2011). Variations of this effect have been observed in humans as well. For example, Verbruggen *et al.* (2017) showed that participants tended to initiate a new trial more quickly after losses in gambling tasks compared to wins and non-gambles, which was interpreted as a manifestation of the 'frustration' effect (for direct and conceptual replications of this finding, see Eben *et al.*, 2020). Speeding up after losses has recently also been observed in large behavioral tracking data from a real online gambling product (Chen *et al.*, 2022).

Many goal pursuit activities require individuals to carry out multiple successive actions to obtain a reward. Accordingly, individuals may fail at any stage of the goal pursuit process, which in turn may influence the magnitude of the 'frustration' effect triggered by reward omission or blockage. As individuals make progress in their goal pursuit, their motivation to obtain the goal generally increases (i.e., the goal gradient hypothesis; Hull, 1932). Recent work has shown that people generally exert more cognitive effort when they are provided with information on their progress in a task (Devine & Otto, 2022; Katzir *et al.*, 2020), and cognitive effort exertion increases near the end of a task (i.e., near goal completion, Devine *et al.*, 2024; Emanuel *et al.*, 2022). Outside the laboratory context, the influence of perceived goal progress on motivation has also been documented. For instance, customers bought coffee more frequently the closer they were to getting a free coffee by collecting enough stamps in a reward program (Kivetz *et al.*, 2006). In addition to influencing people's motivation, goal progress can also influence the magnitude of the 'frustration' effect when they are blocked from the reward (Amsel, 1958). For instance, Haner and Brown (1955) found that children responded with more force when a failure occurred close to success, compared to when they were still far from achieving their goal.

The ubiquitous influence of goal progress on motivated behavior requires a better understanding of how such influence arises. Here we specifically focus on how goal progress may influence the magnitude of the 'frustration' effect when individuals are blocked from a reward. Two closely interrelated factors seem crucial in the effects of goal progress, namely expended effort (Pompilio *et al.*, 2006) and goal proximity (Shidara & Richmond, 2002). As individuals make progress in their goal pursuit, the effort already invested increases (i.e., expended effort), while they also get closer to the end reward (i.e., goal proximity). In many situations, these two factors are inherently confounded, but the psychological mechanisms underlying the effects of these two factors may differ. For instance, expended effort is retrospective in nature as it concerns the effort already invested, while goal proximity is prospective as it orients towards an upcoming reward (Yu *et al.*, 2014). Disentangling the effects of expended effort and goal proximity will therefore be informative in understanding successive goal-directed behaviors, and how individuals react when they fail.

Yu *et al.* (2014) developed an experimental task to disentangle these two typically confounded factors (see Method for more details). The task consisted of multiple 'schedules'. In each schedule participants could win a monetary reward, indicated by an image of a coin divided into four parts. To get the monetary reward, they needed to obtain all four parts of the coin, thus each schedule constituted one goal-pursuit attempt with multiple successive actions. In each schedule, either no part or some parts of the coin were already given to participants for 'free', so the number of 'trials' varied from one (i.e., only one quarter of the coin was missing) to four (i.e., the whole coin was missing). To obtain a missing part of a coin in a

trial, participants had to do a task in which they indicated the left versus right direction of arrows via button presses. At any stage however, they could get blocked. That meant that even when the response was correct, the participants were told that they were too slow and that they would not get the reward. By manipulating the number of trials in a schedule and at which stage participants were blocked, Yu *et al.* (2014) were able to disentangle expended effort and goal proximity. For example, when participants were blocked at the very first trial out of four trials, both goal proximity and expended effort were low; when they were blocked in the last trial out of four trials, both goal proximity and expended effort were high (i.e., the typical situation with the two factors confounded). However, when participants were blocked in the first trial in a one-trial schedule (i.e., three quarters of a coin were already available, the schedule consisted of only one trial), goal proximity was high (as participants were close to obtaining the reward) but expended effort was low (as they only did one trial). Using this procedure, Yu *et al.* (2014) showed that goal proximity and expended effort independently influenced participants' self-reported motivation and the response times (RT) in the arrow task, as well as their self-reported frustration and the force with which they confirmed an outcome when they were blocked from obtaining a reward. More concretely, participants reported being more motivated and responded faster in the arrow task the closer they were to their goal and the more effort they had expended to reach the goal. When they were blocked from obtaining a reward, both the self-reported frustration (collected retrospectively at the end of the experiment) and the response force with which people confirmed the outcome increased as goal proximity and expended effort increased. Taken together, this study shows that goal proximity and expended effort, two frequently confounded factors, independently contribute to the motivational and emotional influences of goal progress, affecting both successive goal-pursuit behaviors and how people react to outcomes (e.g., reward blockage) of their reward pursuit attempts.

The task by Yu *et al.* (2014) successfully disentangled the motivational and emotional influences of expended effort and goal proximity during reward blockage, two important yet frequently confounded factors. As this could be a very valuable tool for studying reward blockage and frustration, we aimed to replicate and extend the original study by Yu *et al.* (2014). Yu *et al.* (2014) measured experienced frustration only retrospectively at the end of the experiment with self-reports. To further explore the emotional influences of reward blockage, we extended Yu *et al.* (2014) by measuring facial muscle activities throughout the experiment with facial electromyography (fEMG). More specifically, we measured the activities of the corrugator supercilii (the muscle that is involved in frowning) and zygomaticus major (the muscle involved in smiling). When being presented with negative stimuli, normally the corrugator supercilii is active and when being presented with positive stimuli, usually the zygomaticus major is active (e.g., Mauss & Robinson, 2009). The activities of these two facial muscles thus provide a measure of the positive versus negative valence of experienced emotions *while* participants engage in the task, instead of retrospectively. This online measurement of valence has been used in previous studies to investigate the evaluation of gambling outcomes (Wu & Clark, 2015), which found heightened activity of the corrugator supercilii after losses. Furthermore, fEMG has been used to investigate outcomes in cognitive tasks, and found that errors (which are presumably negative outcomes) similarly increased the activity over the corrugator supercilii (Elkins-Brown *et al.*, 2016; Lindström *et al.*, 2013).

Taken together, the current study had two aims. The first aim was to replicate the original findings by Yu *et al.* (2014), that expended effort and goal proximity would lead to shorter response times and higher self-reported motivation in the task, and more response force and higher self-reported frustration when participants were blocked from a reward. For this, we used the same task as Yu *et al.* (2014) and measured RT and response force with a custom response box, and self-reported frustration, surprise and motivation at the end of the experiment. Furthermore, we measured activity over the corrugator supercilii and the zygomaticus major in an exploratory manner. Note that some unexpected issues with the response box made our measurement of response force invalid. We reported, however, the results of the force measurement for consistency and transparency in the online supplementary materials (https://osf.io/6v9ak) and discussed the issue of the response box in the discussion.

Methods

Transparency and openness

All raw and processed data, code, and materials can be found on OSF (Eben *et al.*, 2024, https://osf.io/rmh2d/). We did not preregister the analyses as we did not have experience with measurements of force or facial EMG. We did however decide on a data analysis approach before looking at the data. In this manuscript we clearly report which analyses were planned and which analyses were added as exploratory analyses. We report how we determined our sample size, all data exclusions, all manipulations, and all measures in the study. The research was approved by the Ethics committee of the Faculty of Psychology and Educational Sciences of Ghent University (ethical approval number 2022-135; approved 17/01/2023). Thus, the study adhered to the Declaration of Helsinki. Participants provided written informed consent before starting the experiment.

Sample size

In the original analysis, Yu and colleagues examined the effects of proximity and expended effort by combining the p values from multiple repeated-measures ANOVAs using the Stouffer's method. As far as we know, no analytical method exists to compute the statistical power for the Stouffer's combined p values. We therefore decided to go beyond the sample sizes used in Yu *et al.* (between 20 and 27), but still within our resource constraints. We were additionally interested in potential differences in the response force after wins and losses, for which Yu *et al.* did not report the results. We therefore assumed a medium effect size (Cohen's d = 0.5), for which 32 participants would provide approximately 80% power to detect in a within-subject design. Our response force data, however,

was compromised and we never followed up on that specific effect. The eventual sample size of 32 was therefore chosen given all these considerations, to (1) exceed the original sample sizes, (2) be within our resource constraints, and (3) provide sufficient power to detect potential response force differences between wins and losses.

Participants

In total 32 participants (recruited via the local Sona system from Ghent University) completed the experiment between March and July 2023 and remained in the final analysis (23 females, 1 non-binary; 3 left-handed; age M = 24.53 years, SD = 3.38 years, *range* = 18–33). Participants were paid at a rate of 10€/hour plus a bonus (see below) of 10€. As this experiment took approximately 90 minutes (30 minutes of preparation plus 60 minutes on the task), they received 25€ (1.5*10€ +10€) in total.

Apparatus and stimuli of experimental task

The experiment was programmed in PsychoPy (version 2022.2.4; Peirce, 2007). A custom response box was used to register response latency and force with a sampling frequency of 200Hz. The box had two response keys (i.e., mechanical key switches in keyboards), which registered responses and RTs when pressed. Flexiforce sensors on top of the keys were attached to measure the response force. The Flexiforce measuring range had been fixed at 2500g, with 2.45g measuring steps (via Arduino ADC). Even though we tested the set-up before the experiment, post-hoc inspection of the force data occasionally revealed force value of 0 for some of the participants, even though a response (and the corresponding response latency) was registered by the box. As responses were registered using keys similar to those of a keyboard and force was measured with sensors on these keys, it is possible that participants pressed the button without fully placing their index fingers on the sensors, despite being instructed to do so. Therefore, we cannot exclude that all our force data is underestimated and therefore invalid. For consistency and transparency reasons, we report the force data in the online supplementary materials (https://osf.io/6v9ak) but we will not discuss their theoretical implications.

Procedure

We used the same multi-trial reward schedule task developed by Yu *et al.* (2014). At the beginning of each schedule, participants were presented with a cue for 2000 ms (see video on Open Science Framework: https://osf.io/z28b4). The cue consisted of four boxes that were filled with either green stripes (light green boxes in Figure 1) or white color. Depending on how many trials needed to be completed, either no part or some parts of the coin were visible (e.g., if participants had to complete four trials, no part of the coin was visible; if they had to complete three trials, a quarter of the coin was already visible etc.). The number of boxes with green stripes indicated how many parts of a coin were given to participants 'for free'. The number of white boxes indicated the amount of trials

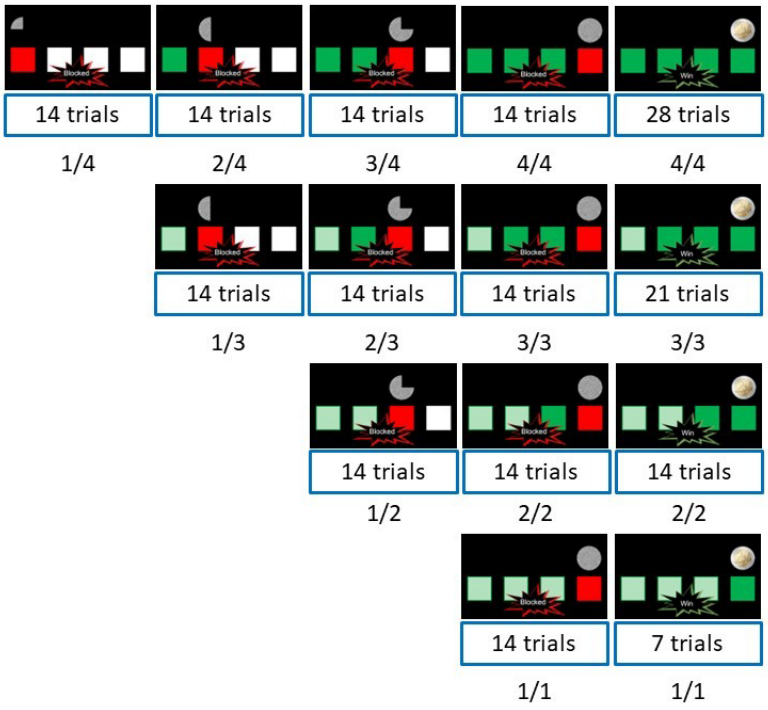

Figure 1. The different schedules used in the experiment. The outcome screens of the different schedules used in the experiment. Note that the screenshots here do not represent the progression across trials within a schedule, but rather only the outcome screen at the end of a schedule. Blocked 1/4 means that the schedule consists of 4 trials, and participants were blocked at trial 1. The numbers within the boxes indicate the number of each specific schedule included in the experiment.

to complete in order to obtain the reward (varying from 1 to 4 trials across schedules). After the cue screen, a blank screen was shown for 750 ms, followed by the arrow task. In the arrow task, participants were presented with three white arrows either all pointing to the left or to the right, and needed to indicate the direction by pressing the left or right button. The arrows were presented for 1 second. Participants were instructed to respond as fast and accurately as possible. We told participants (as in the original study) that on each trial, the computer randomly determined a RT criterion that they would have to meet, otherwise they would be blocked and not receive the reward. In fact, whether participants would be blocked or not on a particular trial was pre-determined (see Figure 1 for the predetermined schedules). Furthermore, if participants made a mistake or were too slow (i.e., RT > 500 ms), they would be blocked regardless of the pre-determined outcome of a trial.

If participants were not blocked on a trial, one white box became green and an additional quarter of the 2 euro coin image became visible. The cue was again shown for 2 seconds, and participants received a new arrow task after a blank screen of 750 ms. As long as participants were not blocked, this sequence of events was repeated until the end of a schedule. In that case, all white boxes were filled green, and the whole 2 euro coin was visible. A text message "Win" was superimposed on the cue, to indicate the outcome of the current schedule. If participants were blocked at any stage, the square for the current trial became red, and the coin image was replaced by an image of white noise (see Figure 1). A text message "Blocked" was superimposed on the cue and the current schedule ended immediately (that is, participants did not need to do the remaining trials in a blocked schedule). In both cases (win and blocked), participants then needed to confirm the current outcome. In the confirm task, they were either presented with 'win? blocked?' with the left button press to confirm a win trial and the right button press to confirm a blocked trial, or with 'blocked? win?' with a left button press to confirm a blocked trial and a right button press to confirm a win trial. As in the original study, participants were instructed to take their time to give the confirm response to avoid having response speed confounding the response force. The confirmed screen was presented for four seconds, regardless of the participants' confirmation RT.

The experiment started with a practice block of 5 schedules, followed by seven experimental blocks of 30 schedules (210 experimental schedules in total). We used the same hidden schedule structure as in Yu *et al.* (2014) (see Figure 1). Participants were told that at the end of the experiment, the program would randomly pick 10 schedules. For each win outcome, they would receive 2 euros. For each blocked outcome, they would receive no extra money. We told them that the results of these 10 selected schedules would be added up, to determine their final bonus (with 10 euros as the maximum) when in fact, all participants received the 10€ bonus. During the self-paced break after each block, participants were asked to tell the experimenter when they wanted to continue, as this could only be done with the computer keyboard.

At the end of the experiment, participants indicated how frustrated they felt at different schedule stages using a 10-point Likert scale (1 = not at all, 10 = very intensely). For this, we presented the different schedules as images and asked participants to click on the Likert scale to indicate their frustration level. They also indicated how surprised they felt after being blocked at each schedule stage and how motivated they were to complete a schedule after seeing the different cues. They then filled in a handedness questionnaire (which can be found in the supporting docs folder on Open Science Framework). Participants were debriefed, and all received 25 euro as compensation.

Electrophysiological recording and data reduction

According to common guidelines and practices (Fridlund & Cacioppo, 1986; Korb *et al.*, 2017), we attached shielded EMG electrodes over the right zygomaticus major and corrugator supercilii areas and placed a ground electrode at the top of the central forehead. Using the Biopac EMG100C amplifier with the MP150 module and the Acqknowledge software (www.biopac.com), we acquired continuous EMG at 1000 Hz sampling rate, with a 100 Hz high-pass and a 500 Hz low-pass.

The data reduction and preprocessing pipeline was based on Korb *et al.* (2017), van Boxtel (2010) and Whelan (2003), and adjusted for our design. We applied a 20-200 Hz band pass filter and full-wave rectified the data by flipping all negative values of the signal to positive values. To exclude high-frequency noise, we then applied a 40 Hz low-pass filter. We epoched the data time-locked to the outcome presentation (win or block) from -750 ms to 2000 ms. Before cleaning the data, we compared the behavioral and EMG data to identify extra "ghost" triggers (triggers that Biopac picked up but were never sent) and deleted these. Then, we excluded error trials and trials where the baseline was more than 2 standard deviations from the participant's baseline mean. The signal was normalized by expressing it as a percentage of the baseline mean per trial. This was done to ensure comparability between individuals and muscles, as well as to control for changes in muscle tone over time. Finally, we excluded the epochs with responses two standard deviations higher than the participant's condition mean.

Planned analyses

Confirm RT measures and force data. All behavioral data processing and analyses were completed with R (R Core Team, 2013, version 4.2.0) using the packages ez (Lawrence, 2016, version 4.4-0), poolr (Cinar & Viechtbauer, 2022, version 1.1-1), lmer (Kuznetsova *et al.*, 2020, version 3.1-3), report (Makowski *et al.*, 2023, version 0.5.7), extraDistr (Wolodzko, 2023, version 1.10.0), loo (Vehtari *et al.*, 2023, version 2.6.0), bridgesampling (Gronau *et al.*, 2020, version 1.1-2), brms (Bürkner, 2017; Bürkner, 2018; Bürkner, 2021, version 2.20.4), bayesplot (Gabry & Mahr, 2024; Gabry *et al.*, 2019, version 1.11.0), bayestestR (Makowski *et al.*, 2019, version 0.13.1), tidaybayes

(Kay, 2023, version 3.0.6) tidyverse (Wickham, 2021, version 1.3.2) and PerformanceAnalytics (Peterson & Carl, 2020, version 2.0.4).

We excluded and replaced two participants where we had problems with data recording, one participant where a memory error in PsychoPy prevented us from finishing the testing, two participants where the electrodes fell off during testing, one participant who seemed to have not understood the task and told us they adopted a guessing strategy, and four participants with fewer than 7 observations in at least one cell after excluding error trials.

For the analyses, we excluded all practice trials, trials in which participants made an error or the RT was above 500 ms, and schedules in which the confirm response (win vs. block) was incorrect. We also excluded trials with an RT of 0 ms as this was an impossible value. The analyses focused on the effect of getting blocked on RT; thus, we only analyzed trials in which participants got blocked. In line with Yu *et al.* (2014), we tested a proximity effect and an effort effect. For the proximity effect, we conducted three univariate ANOVAs: one ANOVA compared schedules on which participants got blocked in the first trial (i.e., 1/1, 1/2, 1/3, 1/4). For these schedules, the effort was held constant at a level of 1 but the proximity could vary between 1 and 4. The second ANOVA compared all schedules on which the participants got blocked in the second trial (i.e., 2/2, 2/3, 2/4). Here the effort level was constant at 2 but the proximity could vary between 1 and 3. The last ANOVA for the proximity effect compared all schedules on which participants got blocked on the third trial (i.e., 3/3, 3/4). Here, the effort level was constant at 3 and the proximity could vary between 1 and 2. To calculate the statistical significance level for the overall proximity effect, Stouffer's combined p (Whitlock, 2005) was computed as it had been done in the original study. This procedure was applied to all dependent variables, including RT to the arrows, confirm RT, frustration ratings, surprise ratings and motivation ratings. The analyses on the confirm response force can be found in the online supplementary materials.

For the effort effect, we also conducted three ANOVAs per dependent variable: keeping the proximity constant at a level of four, we compared all trials on which the participants got blocked on the last trial of a schedule (i.e., 4/4, 3/3, 2/2, 1/1). Here the effort level varied between 4 and 1. The second ANOVA compared all schedules on which the participants got blocked on the penultimate trial (i.e., 3/4, 2/3, 1/2). Here the proximity level was constant at 3 but the effort level varied between 3 and 1. Lastly, we compared all schedules on which the participants got blocked on the antepenultimate trial (2/4, 1/3). Here, the proximity level was 2 and the effort level varied between 2 and 1. Again, we used Stouffer's p to combine all three p-values into one index for the effort effect for each dependent variable.

Furthermore, as in the original study we calculated Pearson's correlations between the mean RT in the arrow task, the mean RT of confirming an outcome (after being blocked), and the self-reported motivation, frustration, and surprise ratings in different schedules. For this analysis, we computed the mean score of each variable per schedule and collapsed the data across all participants. In other words, schedules rather than participants were the unit of analysis. All correlations including the response force can be found in the online supplementary materials.

We report Stouffer's combined p-values and p-values for statistical inferences. The alpha level was .05.

facial EMG data. The facial EMG analyses were conducted with python (Python Software Foundation, 2000, version 3.9.18) using the libraries numpy (Harris *et al.*, 2020, version 1.24.3), matplotlib.pyplot (Hunter, 2007), bioread (Vack, 2023, version 3.0.1), mne (Gramfort *et al.*, 2013, version 1.5.1), scipy (Virtanen *et al.*, 2020, version 1.11.1), and pandas (pandas development team, 2020, version 2.0.3). The individual event related potentials (ERP's) were created by averaging the win trials and all conditions of the blocked trials for each muscle and individual. This resulted in 4 ERP's per individual (2 types of outcomes, and two muscles). Note that we did not distinguish among the different win and blocked schedules, to maximize the number of trials per condition. As depicted in Figure 1, we did not have enough trials per individual schedule for fEMG analyses. The Grand Average ERP's were created by averaging over individual ERP's, resulting in 4 ERP's total. To test for significance, for both muscles we repeated a 2-sided t-test at each time-point comparing win to block conditions using an alpha of 0.05. We controlled for multiple comparisons with a cluster-based permutation test (Maris & Oostenveld, 2007). All data-points with a t-value exceeding the threshold (corresponding to a probability < 0.05) were clustered based on adjacency in the temporal domain. For each cluster the sum of t-values was then calculated and the maximum of these cluster-level statistics was taken. To create a reference distribution against which to test the value of this cluster-level statistic, 5000 permutations of the data were conducted. Each iteration yielded a maximum cluster level statistic and over iterations a null distribution of maximum cluster level values was constructed. The p-value of the effect was then estimated as the proportion of elements in the null distribution exceeding the observed maximum cluster-level test statistic.

Results

Original analyses - Stouffer's p

The descriptive results can be found in Figure 2. In line with the original study, the RT to the arrows showed a significant decrease as a function of goal proximity (Stouffer's $p < 0.01$) and increased effort (Stouffer's combined $p < 0.01$). Similarly, in line with the original study, the confirm RT did not show any significant increase as a function of goal proximity (see Figure 2; Stouffer's combined $p = 0.13$) or increased effort (Stouffer's combined $p = 0.28$). Similarly, subjective emotional ratings completed after the experiment showed that participants' frustration increased with increased goal proximity (Stouffer's combined $p < 0.01$) and increased effort (Stouffer's combined $p < 0.01$). Motivation ratings also increased with goal proximity (Stouffer's combined $p < 0.01$) and increased effort

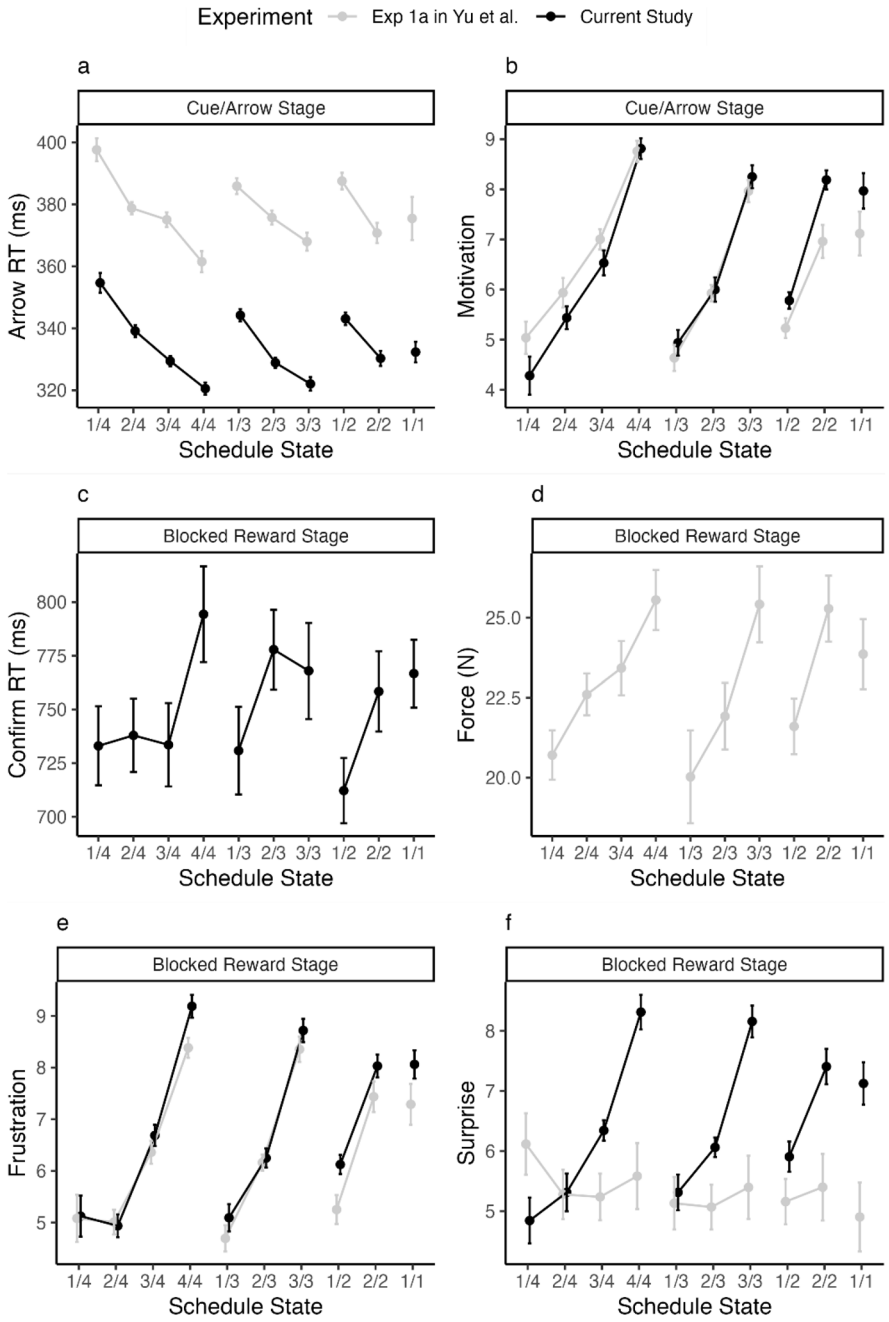

Figure 2. Behavioral and self-report data. Behavioral and self-report data. (**a**) RTs of correct responses to arrows, (**b**) self-reported motivation for each cue during the arrow task, (**c**) confirm RTs and (**d**) response force after being blocked, (**e**) self-reported frustration and (**f**) self-reported surprise at being blocked. Data for plotting the results of Yu *et al.* were extracted from their Figure 3 using plotdigitizer.com. Note that for (**c**) confirmRTs, Yu *et al.* did not plot the results. For (**d**) response force, our force data was invalid and thus not shown. Error bars indicate within subject standard errors.

(Stouffer's combined $p < 0.01$). In contrast to the original study, the surprise ratings also increased with goal proximity (Stouffer's combined $p < 0.01$) but not as a function of increased effort (Stouffer's combined $p = 0.99$). All results from the single ANOVAs can be found in Table 1.

Original analyses - correlational analyses

For completeness, we present the whole correlation matrix in Figure 3 but only describe the correlations that were reported in the original study here. In line with Yu *et al.* (2014), we found a significant correlation between mean RT on the

Table 1. The *F* values and *p* values from the ANOVAs on the effects of proximity and effort across 4, 3, and 2 trials. Stouffer's combined *p* values are reported as an index of the significance of combined effects of proximity or expended effort across the three ANOVAs.

Conditions	arrow RT	Motivation	confirm RT	Frustration	Surprise
Proximity effect 1/4, 1/3, 1/2, 1/1	14.32 ($p < 0.001$)	27.62 ($p < 0.001$)	1.78 ($p = 0.156$)	24.42 ($p < 0.001$)	10.15 ($p < 0.001$)
2/4, 2/3, 2/2	9.29 (p < 0.001)	37.31 ($p < 0.001$)	1.12 ($p = 0.332$)	49.99 ($p < 0.001$)	14.69 ($p < 0.001$)
3/4, 3/3	1.04 (p = 0.315)	31.02 ($p < 0.001$)	1.04 ($p = 0.315$)	53.18 ($p < 0.001$)	57.29 ($p < 0.001$)
combined effect	*p < 0.001*	$p < 0.001$	$p = 0.133$	$p < 0.001$	$p < 0.001$
Effort effect 4/4, 3/3, 2/2, 1/1	6.35 (p < 0.001)	3.64 ($p = 0.015$)	0.68 ($p = 0.564$)	9.46 ($p < 0.001$)	5.39 ($p < 0.001$)
3/4, 2/3, 1/2	19.07 (p < 0.001)	4.35 ($p = 0.017$)	3.88 ($p = 0.026$)	2.63 ($p = 0.080$)	1.33 ($p = 0.272$)
2/4, 1/3	3.47 (p = 0.072)	7.29 ($p = 0.011$)	0.07 ($p = 0.792$)	0.42 ($p = 0.523$)	< 0.01 ($p > 0.999$)
combined effect	*p < 0.001*	$p < 0.001$	$p = 0.287$	$p < 0.001$	$p = 0.993$

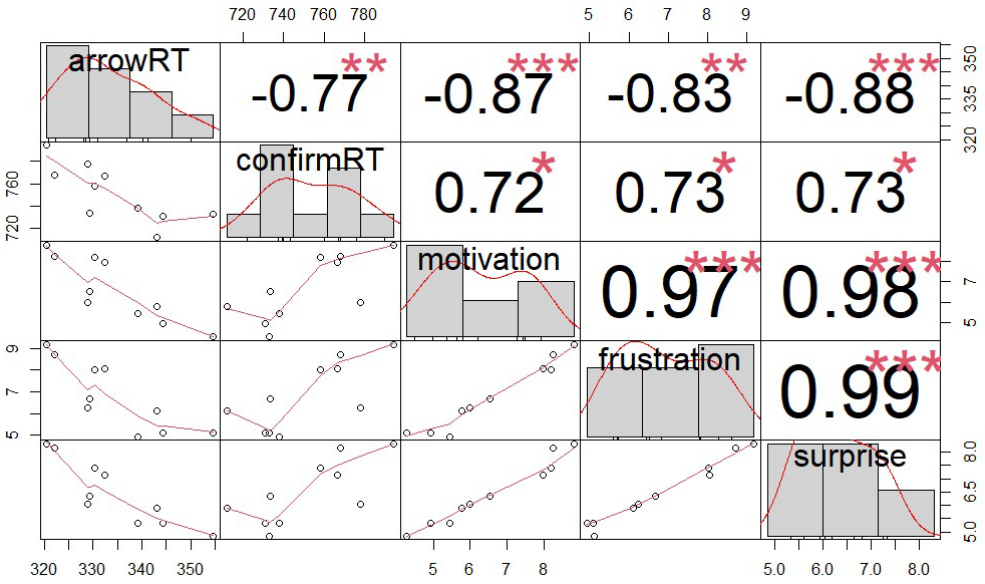

Figure 3. Correlations between the mean RT in the arrow task, the mean RT of confirming a blocked outcome, and self-reported motivation, frustration and surprise in different schedules. The numbers above the diagonal show the correlation coefficients. ***p < .001, **p < .01, *p < .05.

arrow task in the 10 schedule stages and mean self-reported motivation (r = -0.87, $p < .001$). Participants reported a higher motivation to do the task in schedules where they responded more quickly to the arrow, suggesting a high consistency between these two measures. Additionally, motivation prior to blocking in the 10 schedule stages was significantly correlated with mean frustration (r = 0.97, $p < .001$). In schedules where participants reported a higher motivation to complete the task, they also reported more frustration when they were blocked from the reward. For more details, see Figure 3.

Exploratory analyses on the arrow RTs and confirm RTs

In the analyses above, we assessed the effects of proximity and effort separately, using the same data analysis approach adopted by Yu *et al.* (2014). As an exploratory analysis to assess both effects simultaneously, we analyzed both arrow RTs and confirm RTs using Bayesian mixed-effects models in brms (Bürkner, 2021). For both analyses, the RT data were first log transformed. We simultaneously included proximity level (from 1 to 4, with a larger value standing for more proximity to a reward) and effort level (from 1 to 4, with a larger value

standing for more expended effort) as predictors, with random intercepts and random slopes per participant. Both proximity and effort were modelled as monotonic predictors (Bürkner & Charpentier, 2018). The Student's t distribution was used as the likelihood function rather than the default normal distribution, as the latter did not fit the arrow RT data well. Default priors in brms were used. For MCMC sampling, we ran 4 parallel chains with 3000 iterations in the warm-up phase and 5000 iterations in the sampling phase per chain. We inspected the trace plots, R hat values, effective sample sizes and posterior predictive checks to ensure that there were no convergence issues and that the estimates were stable.

For the arrow RT, we observed a negative coefficient for proximity (estimate = -0.017, 95% CI = [-0.024, -0.011]). This indicates that participants responded more quickly in an arrow task when they were closer to the end of a schedule. The coefficient for effort was also negative, estimate = -0.008, 95% CI = [-0.013, -0.004], indicating that participants responded more quickly in an arrow task when they had exerted more effort (i.e., completed more trials) in a schedule so far. For the confirm RT, we observed a different pattern: the coefficient for proximity was positive, estimate = 0.017, 95% CI = [0.001, 0.350], while the coefficient for effort did not differ from 0 credibly, estimate = 0.003, 95% CI = [-0.019, 0.027]. The exploratory analysis thus showed that contrary to the effect of proximity on arrow RTs, participants confirmed an outcome more slowly when the reward blockage was closer to the end of a schedule. This was in contrast to the original study, which did not find any effect on the confirm RT. Note that when we used the same data analysis approach above as in the original study, we also did not observe statistically significant effects of proximity or effort. This discrepancy in results might arise because the exploratory analysis here included proximity and effort simultaneously in one model and analyzed data on each trial (rather than aggregated means in each condition), and thus might have more statistical power to detect an effect. Yu *et al.* did not report or visualize the results on confirm RTs, so it is unclear whether a similar pattern was also present in their data. These exploratory results on confirm RTs were also in line with our finding that participants were more surprised by reward blockage when proximity was high, but expended effort did not influence self-reported surprise. The positive correlation between self-reported surprise and the mean RT of confirming a blocked outcome (Figure 3) further corroborated the idea that being blocked when getting close to a reward might be a surprising event that slowed responses down.

facial EMG analyses

We tested whether the zygomaticus major and corrugator supercilii muscles were differentially activated in win versus blocked trials. Being essential for frowning, we expected the corrugator supercilii to be more active in block trials. We expected the zygomaticus major, responsible for smiling, to be more active in win trials. None of our expectations were borne out by the data. Figure 4 shows the grand average activation of the zygomaticus major muscle for win and block trials. While on average, the zygomaticus major is more active on win trials, this difference was not significant (largest cluster, which is the sum of all *t*-values of all time points in the cluster was *t*(sum) = -34 with a *p* = .84). The individual data shown in Figure 5 reveals that this difference was driven by only a small proportion of participants. The corrugator supercilii also did not reveal any differential activity between win and block trials (largest cluster, which is the sum of all *t*-values of all time points in the cluster was was *t*(sum) = 164 with a *p* = .15; see Figure 6 and Figure 7).

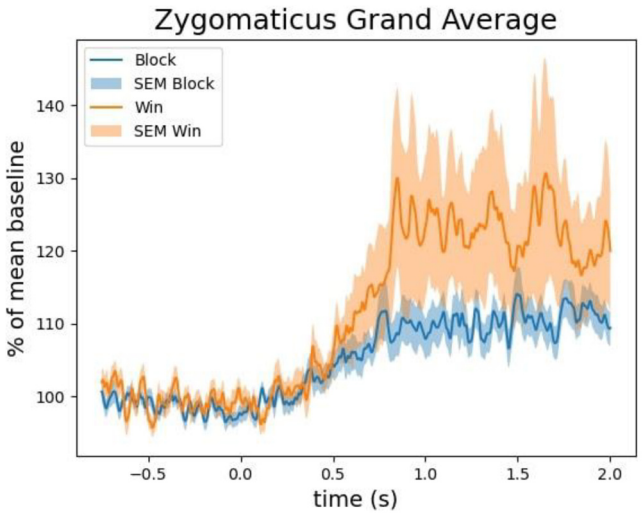

Figure 4. The grand average of the activity over the zygomaticus major as a function of time after the outcome presentation and trial type (win vs. blocked).

Discussion

Humans and non-human animals often face obstacles that block them from achieving their goals, which can lead to frustration. In the current study, we investigated the influence of goal proximity and expended effort on frustration by trying to replicate and extend the study by Yu *et al.* (2014). For this, participants completed multi-trial schedules in order to obtain a reward. On each trial they had to indicate the direction of arrows by a left or right button press, and had to complete one to four arrow trials successfully to obtain a reward, depending on the schedule. Importantly, they could get blocked at any stage of the schedule, resulting in different levels of goal proximity (i.e., the number of trials that still needed to be completed to obtain the reward) and expended effort (i.e., the number of trials that were already completed). In line with Yu *et al.* (2014), we measured the RT in the arrow task, the RT and force with which participants confirmed the outcome of the current schedule, and self-reported surprise, motivation and frustration for each schedule stage at the end of the experiment. As an extension, we also measured the activity over the corrugator supercilii and the zygomaticus major as online measures of emotional valence. Note that our force measure

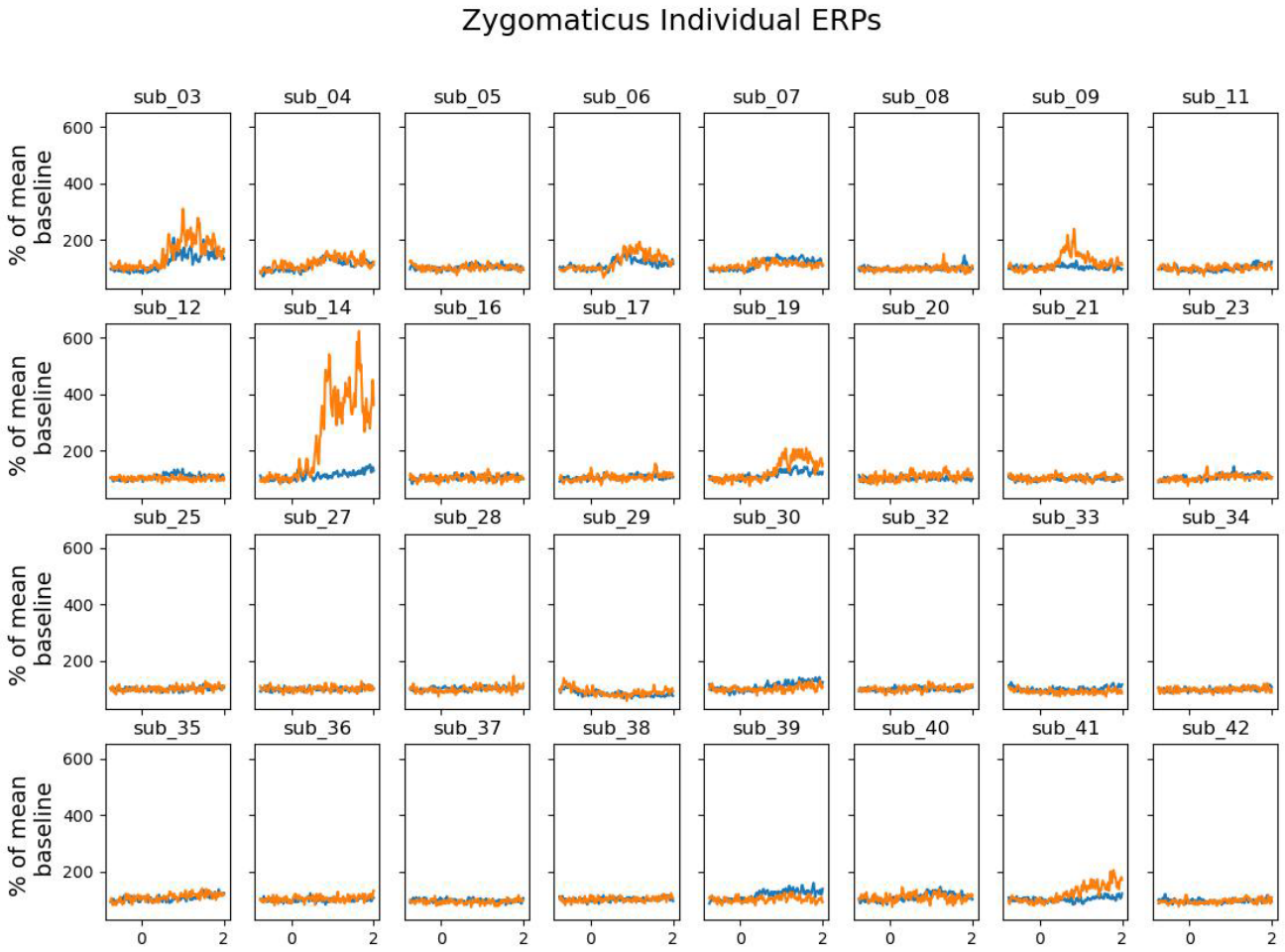

Figure 5. The grand average of the activity over the zygomaticus major as a function of time after the outcome presentation and trial type (win vs. blocked) for each individual.

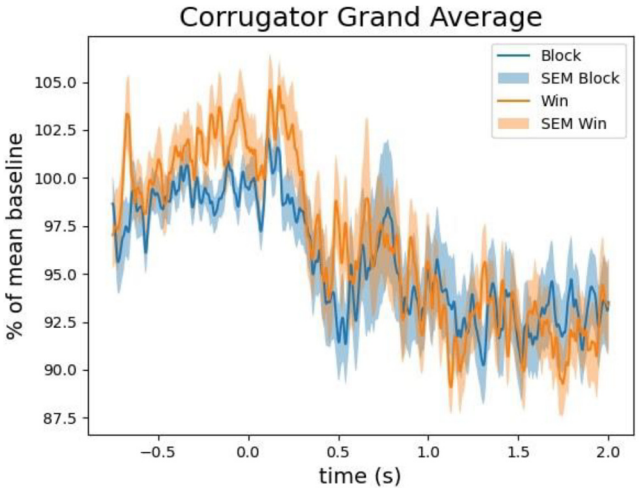

Figure 6. The grand average of the activity over the corrugator supercilii as a function of time after the outcome presentation and trial type (win vs. blocked).

was most likely underestimated. Thus, we will not discuss any theoretical implications related to these force results and will instead focus on the remaining measures.

In the self-report measures, we were able to replicate Yu *et al.* (2014)'s finding that the closer participants were to the reward and the more effort they had exerted, the more motivation they felt in the arrow task. In line with their self-reported motivation, participants responded faster to the arrows the closer they got to the expected reward and the more effort they had exerted. When they were eventually blocked from a reward, they reported more frustration when getting blocked closer to the reward and with more effort exerted. Using the same data analysis approach as in the original research (i.e., combining the p values from multiple ANOVAs), we similarly observed no effect of proximity and effort on the RT of confirming a blocked outcome. In sum, we were able to replicate the effects (or the absence of the effects) of proximity and effort on arrow RTs and confirm RTs, and self-reported motivation and frustration at the end of the experiment.

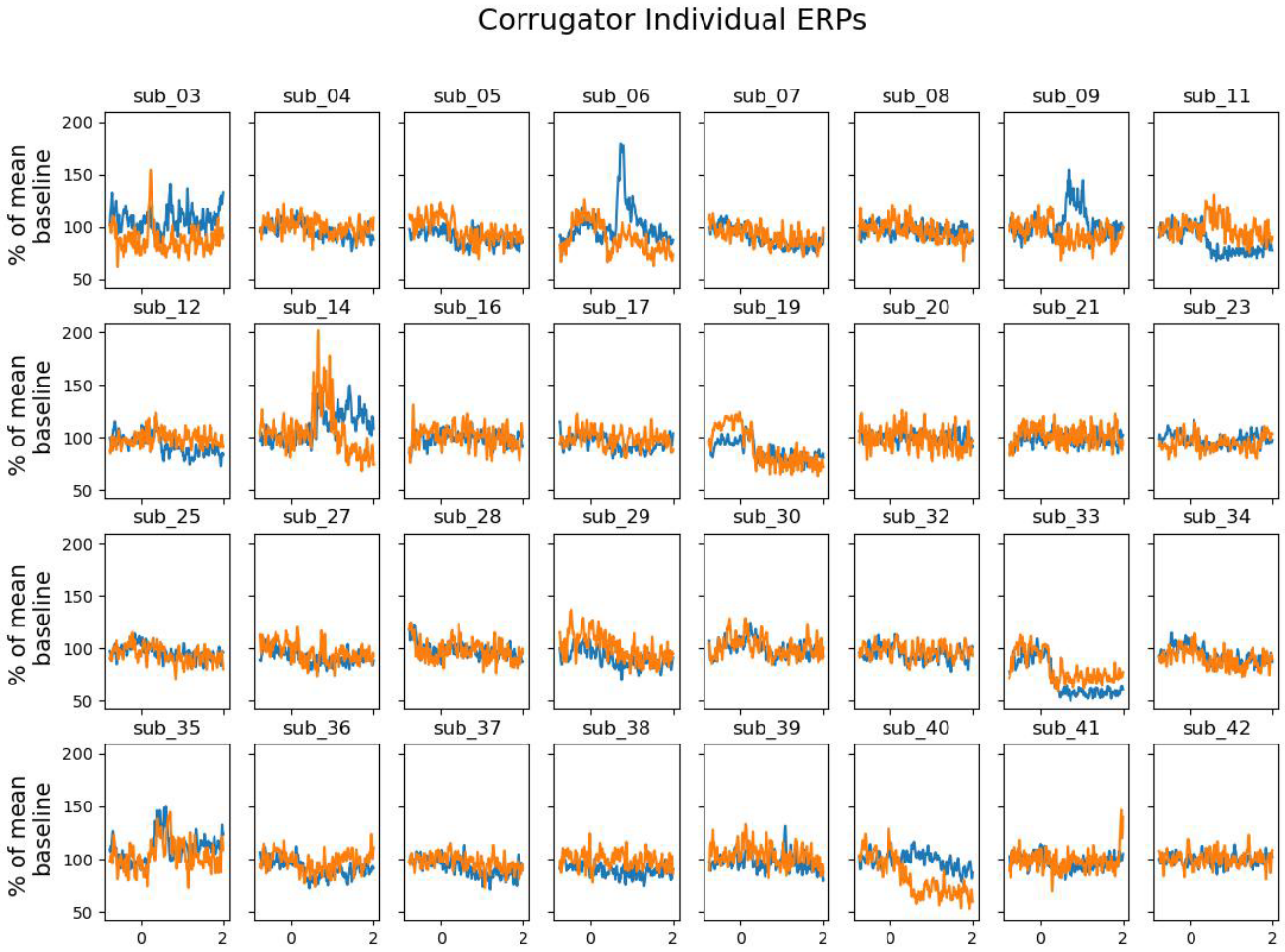

Figure 7. The grand average of the activity over the corrugator supercilii as a function of time after the outcome presentation and trial type (win vs. blocked) for each individual.

Some results differed from those of Yu *et al.* (2014) though. First, participants were more surprised by a blocked outcome when they were getting closer to the end reward, while exerted effort did not influence surprise. This pattern of results was corroborated by the results of an exploratory analysis on confirm RT, in which we found that higher proximity (but not higher effort) increased the RT of confirming an outcome. Furthermore, self-reported surprise also correlated with confirm RT, so that the more surprised participants felt in a schedule, the slower they were in confirming the outcome. Lastly, in the facial EMG data, we found no difference in activation when winning compared to getting blocked, neither in the corrugator supercilii nor in the zygomaticus major.

In general, our replication results seem to be valid despite the underestimation of the force data: the self-report measures show that we were able to induce frustration in participants and that they were generally more motivated to successfully finish the task in later schedule stages. Furthermore, our RT data are likely valid. The mean arrow RTs in the 10 schedules were negatively correlated with the mean self-reported motivation, suggesting that participants indeed responded more quickly when they were more motivated to complete a schedule. The finding that participants got slower in confirming their outcome after they had been blocked at later schedule stages is in contrast to the original study, but in line with the self-reported surprise ratings. In the original study, Yu *et al.* (2014) did not find an effect of surprise. Here however, we found that participants are more surprised the later they got blocked. In the outcome phase, blocking trials were more likely than win trials. However, each schedule contained varying number of trials, and the participants did experience many correct trials on those schedules where they were eventually blocked. From Figure 1, one can compute that the whole design contains 350 correct arrow trials, and 140 'blocked' arrow trials (the probability of being blocked after an arrow trial was therefore around 28%). Those correct responses potentially influenced the participants expectation to win on the next trial. Thus, this surprise might have induced an orienting response (Notebaert *et al.*, 2009; Wessel, 2018) leading to

slower response times in the confirm task after getting blocked at a later schedule stage. In summary, it seems that participants were highly motivated, which was reflected in the response speed as long as they thought they had control over the outcome. Once they got blocked, participants were surprised and took longer to confirm this outcome (Notebaert *et al.*, 2009). Note that other other research using gambling tasks found speeding after reward blockage (see Eben *et al.*, 2020; Verbruggen *et al.*, 2017). Since earlier research however suggests that perceived control over the outcome plays an important role in whether participants speed up or slow down after sub-optimal outcomes (Eben *et al.*, 2023), we assume that here, participants might have felt in control over the outcome and hence showed an orienting response after being blocked.

Our facial EMG results did not show the expected results. While some participants showed the expected effects most of the participants did not show much activation difference over the corrugator supercilii or the zygomaticus major for wins and blocked trials. However, given that some participants showed more activation of the zygomaticus major than over the corrugator supercilii in the win trials, we assume that fEMG in general seems to be a suitable online measure of emotional valence (Mauss & Robinson, 2009). Here we assume that the sample size might have been too small to detect reliable differences in psychophysiological measurements. After all, the sample size was calculated based on a behavioral medium effect size.

As already mentioned, we ran into post-hoc problems with our force measurements, presumably because not all participants placed their fingers directly and constantly on top of the sensors. Unfortunately, we have to assume that all force data are underestimated in our experimental set-up, even in participants who do not have 0 values in their force measurement. Note that similar set-ups with sensors on top of keys have been used successfully in previous studies to measure force (e.g., Pickering *et al.*, 2020), albeit not in the context of examining 'frustration'. In hindsight, the requirement for participants to keep their fingers constantly on the sensors (and not moving their fingers) may be inappropriate in the current study, as the 'frustration' triggered by blocked reward may lead participants to move their fingers despite the instructions to not do so. This may be the reason why the force measures with our custom box were invalid. Alternative ways of measuring force exist, such as by attaching buttons to a leaf spring, and measuring the force applied via strain gauges (e.g., Ko *et al.*, 2012). In such a set-up, force is determined by measuring how much the leaf spring is bent (which in turn is measured by strain gauges). There will therefore be no requirement for participants to keep their finger locations constant on the keys. While we initially experimented with this idea, we eventually decided to not use such a set-up, because the tactile feedback of keys being pressed down was much less clear in this set-up than in our adopted set-up. Since we used rigged feedback to introduce blocked reward, such uncertainty in whether a key had indeed been pressed down or not might lead participants to adjust their subsequent responses, such as pressing keys harder to make sure that a response was registered. To reduce this potential influence, we therefore did not use this alternative set-up in our study. Yu *et al.* (2014) did not provide technical details on the custom box they used, so it is unclear whether these differences in the boxes may explain the different results.

Despite the force measurements being invalid, in light of a continued publication bias in psychological research (Scheel *et al.*, 2021), we would like to follow the call of Lakens and Ensinck (2024) and make this research available, hoping it might be helpful for others who want to use the measure of response force or try to replicate the original study (see Pennington *et al.*, 2023, for a similar take on publishing results that might not answer the research question).

In conclusion, although the self-report measures suggest that frustration was induced in our participants by blocking them from an expected reward, we were unable to replicate all the behavioral findings of Yu *et al.* (2014). As in the original study, our RT data showed participants sped up in the arrow task the closer they got to the reward. Once they were blocked, however participants seemed to be surprised about the outcome and confirmed the outcome with a slowed response. These results were in contrast to the original findings, which did not find any effects of surprise or confirm RT. Given that our current results are not fully in line with the original study, yet plausible, we call for further investigations of frustration and the effects of goal proximity and exerted effort on frustration.

Ethics and consent

The research was approved by the Ethics committee of the Faculty of Psychology and Educational Sciences of Ghent University (ethical approval number 2022–135 approved 17/01/2023). Participants provided written informed consent before starting the experiment.

Data availability statement

Open Science Framework: "Escalating frustration - A replication attempt and extension of Yu *et al.* (2014)", DOI: 10.17605/OSF.IO/RMH2D (Eben *et al.*, 2024).

This project contains the following underlying data:

Code – analysis code and experimental code including supporting documents

Data – raw and processed data. Raw: One data file per participant for demographics, force, behavioral data, self-report ratings, raw force, event trigger journals and fEMG raw data. Processed: For details please see data ducmentation in the documentation folder.

Documentation - Three .md documents containing detailed documentation about data, analysis and the general experimental procedure and conditions. A folder with results overview files including extended analyses and a folder with supporting

documents such as the lab notes, the handedness questionnaire and the checklist used for our fEMG testibg procedure.

Checklist for "Escalating frustration - A replication attempt and extension of Yu *et al*. (2014)", https://osf.io/vsfe2

Creative Commons Attribution 4.0 International Public License

Authors' contributions

CE, ZC and FV conceptualised the study. CE and ZC wrote the experimental code, did the data acquisition and the behavioral analyses. REL did the facial EMG analyses.

CE and ZC wrote the first draft of the manuscript. FV and REL provided feedback.

References

Amsel A: **The role of frustrative nonreward in noncontinuous reward situations.** *Psychol Bull.* 1958; **55**(2): 102–119.
PubMed Abstract | **Publisher Full Text**

Amsel A: **Frustration theory: an analysis of dispositional learning and memory.** Cambridge University Press, 1992.
Publisher Full Text

Bürkner PC: **brms: an R package for Bayesian multilevel models using Stan.** *J Stat Softw.* 2017; **80**(1): 1–28.
Publisher Full Text

Bürkner PC: **Advanced Bayesian multilevel modeling with the R package brms.** *The R Journal.* 2018; **10**(1): 395–411.
Publisher Full Text

Bürkner PC: **Bayesian item response modeling in R with brms and Stan.** *J Stat Softw.* 2021; **100**(5): 1–54.
Publisher Full Text

Bürkner PC, Charpentier E: **Modeling monotonic effects of ordinal predictors in Bayesian regression models.** 2018.
Publisher Full Text

Chen Z, Doekemeijer RA, Noël X, *et al.*: **Winning and losing in online gambling: effects on within-session chasing.** *PLoS One.* 2022; **17**(8): e0273359.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Cinar O, Viechtbauer W: **The poolr package for combining independent and dependent *p* values.** *J Stat Softw.* 2022; **101**(1): 1–42.
Publisher Full Text

de Almeida RMM, Miczek KA: **Aggression escalated by social instigation or by discontinuation of reinforcement ("frustration") in mice: inhibition by anpirtoline: a 5-HT$_{1B}$ receptor agonist.** *Neuropsychopharmacology.* 2002; **27**(2): 171–181.
PubMed Abstract | **Publisher Full Text**

Devine S, Otto AR: **Information about task progress modulates cognitive demand avoidance.** *Cognition.* 2022; **225**: 105107.
PubMed Abstract | **Publisher Full Text**

Devine S, Roy M, Beierholm U, *et al.*: **Proximity to rewards modulates parameters of effortful control exertion.** *J Exp Psychol Gen.* 2024; **153**(5): 1257–1267.
PubMed Abstract | **Publisher Full Text**

Eben C, Chen Z, London RE, *et al.*: **Escalating frustration - a replication attempt and extension of Yu *et al.*** (2014). [dataset]. 2024.
http://www.doi.org/10.17605/OSF.IO/RMH2D

Eben C, Chen Z, Vermeylen L, *et al.*: **A direct and conceptual replication of post-loss speeding when gambling.** *R Soc Open Sci.* 2020; **7**(5): 200090.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Eben C, Vermeylen L, Chen Z, *et al.*: **When response selection becomes gambling: post-error slowing and speeding in self-paced colour discrimination tasks.** *Collabra: Psychology.* 2023; **9**(1): 73052.
Publisher Full Text

Elkins-Brown N, Saunders B, Inzlicht M: **Error-related electromyographic activity over the corrugator supercilii is associated with neural performance monitoring**. Facial EMG relates to PE. *Psychophysiology.* 2016; **53**(2): 159–170.
PubMed Abstract | **Publisher Full Text**

Emanuel A, Katzir M, Liberman N: **Why do people increase effort near a deadline? An opportunity-cost model of goal gradients.** *J Exp Psychol Gen.* 2022; **151**(11): 2910–2926.
PubMed Abstract | **Publisher Full Text**

Fridlund AJ, Cacioppo JT: **Guidelines for human electromyographic research.** *Psychophysiology.* 1986; **23**(5): 567–589.
PubMed Abstract | **Publisher Full Text**

Gabry J, Mahr T: **bayesplot: plotting for Bayesian models.** [R package version 1.11.0]. 2024.
Reference Source

Gabry J, Simpson D, Vehtari A, *et al.*: **Visualization in Bayesian workflow.** *J R Stat Soc A.* 2019; **182**(2): 389–402.
Publisher Full Text

Gramfort A, Luessi M, Larson E, *et al.*: **MEG and EEG data analysis with MNE-Python.** *Front Neurosci.* 2013; **7**: 267.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Gronau QF, Singmann H, Wagenmakers EJ: **bridgesampling: an R package for estimating normalizing constants.** *J Stat Softw.* 2020; **92**(10): 1–29.
Publisher Full Text

Haner CF, Brown PA: **Clarification of the instigation to action concept in the frustration-aggression hypothesis.** *J Abnorm Psychol.* 1955; **51**(2): 204–206.
PubMed Abstract | **Publisher Full Text**

Harris CR, Millman KJ, van der Walt SJ, *et al.*: **Array programming with NumPy.** *Nature.* 2020; **585**(7825): 357–362.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Hull CL: **The goal-gradient hypothesis and maze learning.** *Psychol Rev.* 1932; **39**(1): 25–43.
Publisher Full Text

Hunter JD: **Matplotlib: a 2D graphics environment.** *Comput Sci Eng.* 2007; **9**(3): 90–95.
Publisher Full Text

Katzir M, Emanuel A, Liberman N: **Cognitive performance is enhanced if one knows when the task will end.** *Cognition.* 2020; **197**: 104189.
PubMed Abstract | **Publisher Full Text**

Kay M: **tidybayes: tidy data and geoms for Bayesian models.** [R package version 3.0.6], 2023.
Publisher Full Text

Kivetz R, Urminsky O, Zheng Y: **The goal-gradient hypothesis resurrected: purchase acceleration, illusionary goal progress, and customer retention.** *J Marketing Res.* 2006; **43**(1): 39–58.
Publisher Full Text

Ko YT, Alsford T, Miller J: **Inhibitory effects on response force in the stop-signal paradigm.** *J Exp Psychol Hum Percept Perform.* 2012; **38**(2): 465–477.
PubMed Abstract | **Publisher Full Text**

Korb S, Osimo SA, Suran T, *et al.*: **Face proprioception does not modulate access to visual awareness of emotional faces in a continuous flash suppression paradigm.** *Conscious Cogn.* 2017; **51**: 166–180.
PubMed Abstract | **Publisher Full Text**

Kuznetsova A, Bruun Brockhoff P, Haubo Bojesen Christensen R: **lmerTest: tests in linear mixed effects models.** [R package version 3.1-3], 2020.
Reference Source

Lakens D, Ensinck ENF: **Make abandoned research publicly available.** *Nat Hum Behav.* 2024; **8**(4): 609–610.
PubMed Abstract | **Publisher Full Text**

Lawrence MA: **ez: easy analysis and visualization of factorial experiments.** [R package version 4.4-0], 2016.
Reference Source

Lindström BR, Mattsson-Mårn IB, Golkar A, *et al.*: **In your face: risk of punishment enhances cognitive control and error-related activity in the corrugator supercilii muscle.** *PLoS One.* 2013; **8**(6): e65692.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Makowski D, Ben-Shachar MS, Lüdecke D: **bayestestr: describing effects and their uncertainty, Existence and significance within the Bayesian framework.** *J Open Source Soft.* 2019; **4**(40): 1541.
Publisher Full Text

Makowski D, Lüdecke D, Patil I, *et al.*: **Automated results reporting as a practical tool to improve reproducibility and methodological best practices adoption.** *CRAN.* 2023.
Reference Source

Maris E, Oostenveld R: **Nonparametric statistical testing of EEG- and MEG-data.** *J Neurosci Methods.* 2007; **164**(1): 177–190.
PubMed Abstract | **Publisher Full Text**

Mauss IB, Robinson MD: **Measures of emotion: a review.** *Cogn Emot.* 2009; **23**(2): 209–237.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Notebaert W, Houtman F, Opstal FV, *et al.*: **Post-error slowing: an orienting account.** *Cognition.* 2009; **111**(2): 275–279.
PubMed Abstract | **Publisher Full Text**

pandas development team: **pandas-dev/pandas: pandas.** (Version latest). *Zenodo*. 2020.
Publisher Full Text

Peirce JW: **PsychoPy—Psychophysics software in Python.** *J Neurosci Methods.* 2007; **162**(1–2): 8–13.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Pennington CR, Ploszajski M, Mistry P, *et al.*: **Relationships between the race Implicit Association Test and other measures of implicit and explicit social cognition.** *Front Psychol.* 2023; **14**: 1197298.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Peterson BG, Carl P: **PerformanceAnalytics: econometric tools for performance and risk analysis.** [R package version 2.0.4], 2020.
Reference Source

Pickering JS, Leroi I, McBride J, *et al.*: **Continuous force measurements reveal no inhibitory control deficits in Parkinson's disease.** *Exp Brain Res.* 2020; **238**(5): 1119–1132.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Pompilio L, Kacelnik A, Behmer ST: **State-dependent learned valuation drives choice in an invertebrate.** *Science.* 2006; **311**(5767): 1613–1615.
PubMed Abstract | **Publisher Full Text**

Python Software Foundation: **Python: a dynamic, open source programming language.** Python Software Foundation, Dover, NH, 2000.
Reference Source

R Core Team: **R: a language and environment for statistical computing.** R Foundation for Statistical Computing. Vienna, Austria, 2013.
Reference Source

Scheel AM, Schijen MRMJ, Lakens D: **An excess of positive results: comparing the standard psychology literature with registered reports.** *Advan Methods Pract Psychol Sci.* 2021; **4**(2): 25152459211007467.
Publisher Full Text

Shidara M, Richmond BJ: **Anterior cingulate: single neuronal signals related to degree of reward expectancy.** *Science.* 2002; **296**(5573): 1709–1711.
PubMed Abstract | **Publisher Full Text**

Vack N: **Bioread.** 2023.
Reference Source

van Boxtel A: **Facial EMG as a tool for inferring affective states.** *Proceedings of Measuring Behavior 2010.* 2010; 104–108.
Reference Source

Vehtari A, Gabry J, Magnusson M, *et al.*: **loo: efficient leave-one-out cross-validation and waic for bayesian models**. [R package version 2.6.0]. 2023.
Reference Source

Verbruggen F, Chambers CD, Lawrence NS, *et al.*: **Winning and losing: effects on impulsive action.** *J Exp Psychol Hum Percept Perform.* 2017; **43**(1): 147–168.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Vindas MA, Folkedal O, Kristiansen TS, *et al.*: **Omission of Expected Reward agitates Atlantic salmon (*Salmo salar*).** *Anim Cogn.* 2012; **15**(5): 903–911.
PubMed Abstract | **Publisher Full Text**

Virtanen P, Gommers R, Oliphant TE, *et al.*: **SciPy 1.0: fundamental algorithms for scientific computing in Python.** *Nat Methods.* 2020; **17**(3): 261–272.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Wessel JR: **An adaptive orienting theory of error processing.** *Psychophysiology.* 2018; **55**(3): e13041.
PubMed Abstract | **Publisher Full Text**

Whelan PJ: **Electromyogram recordings from freely moving animals.** *Methods.* 2003; **30**(2): 127–141.
PubMed Abstract | **Publisher Full Text**

Whitlock MC: **Combining probability from independent tests: the weighted Z-method is superior to Fisher's approach.** *J Evol Biol.* 2005; **18**(5): 1368–1373.
PubMed Abstract | **Publisher Full Text**

Wickham H: **tidyverse: easily install and load the tidyverse.** [R package version 1.3.1], 2021.
Reference Source

Wolodzko T: **extraDistr: additional univariate and multivariate distributions.** [R package version 1.10.0], 2023.
Reference Source

Wu Y, Clark L: **Disappointment and regret enhance corrugator reactivity in a gambling task.** *Psychophysiology.* 2015; **52**(4): 518–523.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Yu R, Mobbs D, Seymour B, *et al.*: **The neural signature of escalating frustration in humans.** *Cortex.* 2014; **54**: 165–178.
PubMed Abstract | **Publisher Full Text**

Zentall TR: **Maladaptive "gambling" by pigeons.** *Behav Processes.* 2011; **87**(1): 50–56.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

Zimmerman PH, Koene P: **The effect of frustrative nonreward on vocalisations and behaviour in the laying hen, *Gallus gallus domesticus*.** *Behav Processes.* 1998; **44**(1): 73–79.
PubMed Abstract | **Publisher Full Text**

Open Peer Review

Current Peer Review Status: ✔ ✔ ?

Version 2

Reviewer Report 10 May 2025

https://doi.org/10.21956/openreseurope.21800.r53210

© **2025 Naik A.** This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

?

Aijaz A. Naik
National Institute of Mental Health (NIMH), Bethesda, USA

Major Comments

1. Incomplete Theoretical Framework and Literature Integration

The manuscript appears to present only a truncated introduction, referencing amendments from a previous version without a full narrative arc. While key citations such as Devine et al. (2024), Katzir et al. (2020), and Amsel (1958) are mentioned, the broader theoretical context of frustration, emotional regulation, and facial EMG remains underdeveloped.

- A complete introduction is essential to establish the theoretical relevance of facial EMG in the context of frustration research.
- The authors should clearly articulate the conceptual gaps, particularly regarding the multimodal relationship between behavioral responses, self-reported affect, and physiological data.
- The manuscript would benefit from integrating relevant models of emotion (e.g., the valence-arousal framework or frustration-aggression theory) to ground their hypotheses.

Suggestions:

- Situate the current study within classic and contemporary frustration paradigms.
- Discuss how the EMG findings extend or replicate known effects (e.g., zygomaticus inhibition or corrugator activation in negative affect).
- Explicitly define the novelty of the current approach—what does this study reveal that has not been addressed by prior work?

2. Unclear Research Questions and Hypotheses

While the methods and results are outlined, the absence of clearly stated research questions or a hypothesis framework makes it difficult to evaluate the logical flow of the study.

- Are the authors predicting stronger corrugator activation and zygomaticus suppression in blocked (vs. win) trials?
- How are these patterns expected to correlate with self-report measures?
- Do the authors expect individual differences in these patterns, and if so, are they theory-driven?

Recommendations:

- Clearly state specific, testable hypotheses.
- Justify predictions based on past literature, especially within affective neuroscience or psychophysiology.

3. Methodological Gaps in Experimental Design

While the general paradigm appears to involve alternating win and blocked reward trials, the manuscript lacks critical details:

- **Operationalization of frustration:** How is the blocked outcome manipulated to induce frustration?
- **Force data omission:** The authors state the force data was invalid but fail to explain why or how this impacts interpretation.

Concerns:

- Was the experimental task pre-registered?
- How was the trial order counterbalanced, if at all?

4. Bayesian Modeling and Statistical Transparency

The manuscript indicates a shift from ANOVA to multilevel Bayesian models using brms, yet reporting is vague and lacking transparency.

Required clarifications:

- Justification for choosing Bayesian multilevel modeling beyond general statements.
- Clear specification of model structure: priors, fixed and random effects, convergence diagnostics (e.g., R-hat, ESS).
- Report of model fit indices (e.g., WAIC, LOO) if applicable.
- Presentation of posterior distributions, credible intervals, and interpretation of effect sizes.

5. Missing or Severely Truncated Discussion

The manuscript lacks a robust discussion section. As a result:

- The theoretical implications of the findings remain unclear.
- There is no reflection on how the results support or challenge existing frustration theories.
- Alignment (or misalignment) between self-report and EMG data is not addressed.
- Limitations and future directions are not discussed.

Suggestions:

- Reconnect the findings with the introduction and hypotheses.
- Discuss whether EMG data serve as valid indices of affective frustration.
- Address real-world implications or extensions to related domains (e.g., decision-making, cognitive control, psychopathology).

Minor Comments

1. Figures and Captions

- **Figure 1** does not clearly explain what is being shown; a timeline or schematic of the task would be more informative.
- **Figure 2** has panels a–f but lacks sufficient explanation of each panel and the relevant results.
- **Figures 5–7** need:
  - Y-axis units
  - Directionality of effect explained (e.g., does higher corrugator activity mean more negative affect?)
  - Participant identifiers (or consistent ordering) for better cross-referencing.

2. Statistical Reporting Inconsistencies
- Table 1 includes inconsistently reported p-values (some with <, others exact).
- The statistical tests corresponding to the reported F-values are not described.
- Effect sizes (e.g., η^2, Cohen's d) are missing.
- Multilevel model results are not clearly summarized with parameter estimates.

4. Terminology and Accessibility
- Briefly define terms like "zygomaticus major" and "corrugator supercilii" for broader audiences.
- Avoid jargon like "post-loss speeding" without first defining it.

Is the work clearly and accurately presented and does it engage with the current literature?
Partly

Is the study design appropriate and is the work technically sound?
Partly

Are sufficient details of methods and analysis provided to allow replication by others?
Partly

Are all the source data and materials underlying the results available?
Partly

If applicable, is the statistical analysis and its interpretation appropriate?
Partly

Are the conclusions drawn adequately supported by the results?
Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Neuroscience, Plasticity, Engrams, Frustration-non-reward, circuits,

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Reviewer Report 09 May 2025

https://doi.org/10.21956/openreseurope.21800.r53044

© 2025 Devine S. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Sean Devine
Department of Psychology, McGill University, Montreal, Québec, Canada

I am satisfied with the authors' response to my comments and commend them on their contribution to the field.

Is the work clearly and accurately presented and does it engage with the current literature?
Partly

Is the study design appropriate and is the work technically sound?
Partly

Are sufficient details of methods and analysis provided to allow replication by others?
Partly

Are all the source data and materials underlying the results available?
Partly

If applicable, is the statistical analysis and its interpretation appropriate?
Partly

Are the conclusions drawn adequately supported by the results?
Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Cognitive psychology, cognitive effort, goal-gradient effects, EMG analysis, computational modeling, multilevel modeling

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Reviewer Report 07 May 2025

https://doi.org/10.21956/openreseurope.21800.r53045

© 2025 Korb S. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Sebastian Korb
[1] Psychology, University of Essex (Ringgold ID: 2591), Colchester, England, UK
[2] Psychology, University of Vienna, Vienna, Austria

I am pleased with the changes made to the manuscript, which in its latest version has much improved. I especially found useful the new figure 2 and discussion on which of the original

findings had been replicated and which ones had not. I have no further comments and commend the authors for a nice piece of work.

Is the work clearly and accurately presented and does it engage with the current literature?
Partly

Is the study design appropriate and is the work technically sound?
Partly

Are sufficient details of methods and analysis provided to allow replication by others?
Partly

Are all the source data and materials underlying the results available?
Partly

If applicable, is the statistical analysis and its interpretation appropriate?
Partly

Are the conclusions drawn adequately supported by the results?
Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: emotion, facial expressions, embodiment, facial EMG, EEG

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Version 1

Reviewer Report 21 February 2025

https://doi.org/10.21956/openreseurope.19188.r50620

© 2025 Korb S. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

?

Sebastian Korb
[1] Psychology, University of Essex (Ringgold ID: 2591), Colchester, England, UK
[2] Psychology, University of Vienna, Vienna, Austria
[3] Psychology, University of Essex (Ringgold ID: 2591), Colchester, England, UK
[4] Psychology, University of Vienna, Vienna, Austria

The manuscript describes a replication attempt of a study by Yu et al. (2014), who had shown that

the more participants have already put effort into receiving a reward, and the closer they are to getting it, the more they become frustrated when blocked from receiving the announced reward, resulting in increased response force in the next trials. The same methods were used as in the original study, but in addition facial EMG was acquired from the main smiling (zygomaticus major) and frowning (corrugator supercilii) muscles. The force measurement was faulty and therefore not included (but see supplementary materials), and the facial EMG did not result in significant effects. Other than that, the main findings by Yu et al. (2014) were replicated, although participants reported higher levels of surprise about blocked trials, compared to the original study.
The methods are sound, and the manuscript is well written. I mainly have minor comments (see below), but also struggle with the authors' conclusion that participants were surprised by the 'blocking' trials. This interpretation seems unlikely to me, based on the fact that blocking trials were 2 times more likely (n = 140) than win trials (n = 70). Also, surprise was only measured retrospectively, at the end of the experiment (it is thus memory about feeling surprised).
Abstract
_clarify that the following results were unexpected: "participants were slower in confirming an outcome when they were blocked closer to the reward"
_the phrase "confirm RT" is unclear to the naïve reader
Introduction & Methods
_p. 4: "In each schedule, some parts of the coin were already given to participants for 'free', so the number of 'trials' varied from one (i.e., only one quarter of the coin was missing) to four (i.e., the whole coin was missing)." - this sentence suggests that the experiment included schedules in which all parts of the coin were missing, but Figure 1 suggests otherwise. Similarly, text in the first paragraph on page 6 is not compatible with this scenario ("Depending on how many trials needed to be completed, parts of the coin were visible as well (e.g., if participants had to complete three trials, a quarter of the coin was already visible)." – how can parts of the coin be visible in schedules with all coin quarters missing?). Please clarify or correct if necessary.
_p. 6-7 explains that participants were told that at the end of the experiment "the program would randomly pick 10 schedules. For each win outcome, they would receive 2 euros. For each blocked outcome, they would receive no extra money. [...] the results of these 10 selected schedules would be added up, to determine their final bonus (with 10 euros as the maximum)". If I understand this correctly, participants were first told they could potentially earn 20 euros (10 win schedules x 2 euros), but then that they would maximally receive half of that? If that is correct, I would find it rather disappointing (and confusing) for participants.
_p. 7: the word "filter" is missing after "We applied a 20-200 Hz band pass"
_p.7: trials with an RT of 0 ms were removed from analyses, but how come trials with very fast (i.e. < 100 or 200 ms) RTs were kept in the analyses? It would also seem implausible to answer that fast.

Results
_Figure 2: please number the panels (also given that the caption refers to the 5 panels by number), improve the y-axis label in panel 1 (change to "confirm RT (ms)"), either make all x-axis ticks the same or explain why the differ (some include blocked trials, some do not), and clarify which of these analyses are on win/block trials only or on both types of trials. Finally, is it possible to indicate significant effects (e.g. with a star)?
_Figure 4: please add a ribbon around the EMG lines to show some measure of variance (e.g. SEM or 95% CI)
_Figures 5 and 7 are interesting but not necessary in the main text, I would move them to the supplementary materials

_no significant effects were found in the facial EMG. However, this could be due to the somewhat unusual analysis, which included all time points and then applied a correction for multiple comparisons. Instead, you could average over time windows of 250 or 500 ms, and then include the factor time (with 4 or 8 levels).
_p10: missing 's' after 'participant': "For this, participant completed multi-trial schedules"

Is the work clearly and accurately presented and does it engage with the current literature?
Yes

Is the study design appropriate and is the work technically sound?
Yes

Are sufficient details of methods and analysis provided to allow replication by others?
Yes

Are all the source data and materials underlying the results available?
Yes

If applicable, is the statistical analysis and its interpretation appropriate?
Yes

Are the conclusions drawn adequately supported by the results?
Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: emotion, facial expressions, embodiment, facial EMG, EEG

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 02 Apr 2025

Charlotte Eben

1. I mainly have minor comments (see below), but also struggle with the authors' conclusion that participants were surprised by the 'blocking' trials. This interpretation seems unlikely to me, based on the fact that blocking trials were 2 times more likely (n = 140) than win trials (n = 70). Also, surprise was only measured retrospectively, at the end of the experiment (it is thus memory about feeling surprised).

Reply: *We have added this to the discussion about the results on the surprise ratings (p.18):* "In the outcome phase, blocking trials were more likely than win trials. However, each schedule contained varying number of trials, and the participants did experience many correct trials on those schedules where they were eventually blocked. From Figure 1, one can compute that the whole design contains 350 correct arrow trials, and 140 'blocked' arrow trials (the probability of being blocked after an arrow trial was therefore around 28%). This probability

of correct responses potentially influenced the participants expectation to win on the next trial. Thus, this surprise might have induced an orienting response (Notebaert et al., 2009; Wessel, 2018) leading to slower response times in the confirm task after getting blocked at a later schedule stage."

1. _clarify that the following results were unexpected: "participants were slower in confirming an outcome when they were blocked closer to the reward"

Reply: *We have done that.*

1. _the phrase "confirm RT" is unclear to the naïve reader

Reply: *We have changed the wording to 'the time taken to confirm the outcome'.*

1. _p. 4: "In each schedule, some parts of the coin were already given to participants for 'free', so the number of 'trials' varied from one (i.e., only one quarter of the coin was missing) to four (i.e., the whole coin was missing)." - this sentence suggests that the experiment included schedules in which all parts of the coin were missing, but Figure 1 suggests otherwise. Similarly, text in the first paragraph on page 6 is not compatible with this scenario ("Depending on how many trials needed to be completed, parts of the coin were visible as well (e.g., if participants had to complete three trials, a quarter of the coin was already visible)." – how can parts of the coin be visible in schedules with all coin quarters missing?). Please clarify or correct if necessary.

Reply: *All schedules in which participants had to complete four trials to receive the coin started without any part of the coin given. We have now revised the texts throughout the manuscript to make this clearer. Figure 1 only indicates the outcome trials, but not the progression of trials within a schedule. We have now also clarified this in the figure caption. For further information, see the video of the trial procedure on OSF: https://osf.io/z28b4.*

1. _p. 6-7 explains that participants were told that at the end of the experiment "the program would randomly pick 10 schedules. For each win outcome, they would receive 2 euros. For each blocked outcome, they would receive no extra money. [...] the results of these 10 selected schedules would be added up, to determine their final bonus (with 10 euros as the maximum)". If I understand this correctly, participants were first told they could potentially earn 20 euros (10 win schedules x 2 euros), but then that they would maximally receive half of that? If that is correct, I would find it rather disappointing (and confusing) for participants.

Reply: *Indeed, participants were told that they could receive extra bonus and that this bonus would be the added results of the 10 chosen trials with 2€ per win. At the same time, they were informed that the maximum payout was 10€. They received this information before the experiment started and it is a common way of keeping the payout somewhat equal in gambling like tasks where the experimenter does not have control over the outcome. In this case, all participants received the 10€ maximum payout in addition to the 15€ reimbursement for their participation. None of the participants complained about this payout structure and in fact, most of them were pleasantly surprised that they would leave the experiment with 10€ extra.*

1. _p. 7: the word "filter" is missing after "We applied a 20-200 Hz band pass"

Reply: *We have corrected this.*

1. _p.7: trials with an RT of 0 ms were removed from analyses, but how come trials with very fast (i.e. < 100 or 200 ms) RTs were kept in the analyses? It would also seem implausible to answer that fast.

Reply: *Yu et al. (2014) did not report any exclusion criterion for RT analyses. To be consistent*

with their analysis approach, we therefore also did not implement any exclusion criterion, except for RTs of 0 which might indicate premature responses that were initiated before the stimulus presentation. However, considering this comment, we repeated the ANOVA analysis as in the original analysis, but with RTs below 200 ms excluded. For the confirmation RT, this led to the exclusion of 0.03% of the trials. For the arrow RT, this led to the exclusion of 0.18% of the trials. Importantly, all results reported in the manuscript remain the same. Since (1) only a small number of trials had an RT below 200ms, (2) the results remain the same with and without excluding RTs below 200ms, and (3) Yu and colleagues did not use such an exclusion criterion, we decide to report the results without excluding RTs below 200 ms, as they currently are, in the manuscript.

1. _Figure 2: please number the panels (also given that the caption refers to the 5 panels by number), improve the y-axis label in panel 1 (change to "confirm RT (ms)"), either make all x-axis ticks the same or explain why the differ (some include blocked trials, some do not), and clarify which of these analyses are on win/block trials only or on both types of trials. Finally, is it possible to indicate significant effects (e.g. with a star)?

Reply: *We have added letters to the panels, improved the y-axis of the confirm RT and added explanations to the schedule stages. As explained on p. 9, we only included blocked trials in the analyses (and correct trials for the arrow task). We are however not able to add indications of significance since the effects we looked at are not reflected in the descriptives (effort vs. proximity effect p. 9/10 for further explanation). We have added the original plots for better comparison.*

1. _Figure 4: please add a ribbon around the EMG lines to show some measure of variance (e.g. SEM or 95% CI)

Reply: *We have added 95% CI as a ribbon around the EMG lines in Figure 4 and Figure 6.*

1. _Figures 5 and 7 are interesting but not necessary in the main text, I would move them to the supplementary materials

Reply: *We think it is important to highlight the substantial differences across participants, which is often not shown when only group averages are presented. We also refer to these individual differences in the manuscript when discussing the fEMG results. For these reasons, we think it is more convenient to have Figures 5 and 7 in the manuscript.*

1. _no significant effects were found in the facial EMG. However, this could be due to the somewhat unusual analysis, which included all time points and then applied a correction for multiple comparisons. Instead, you could average over time windows of 250 or 500 ms, and then include the factor time (with 4 or 8 levels).

Reply: *Thank you for this suggestion. When analyzing the fEMG data, we followed an analysis approach that is often used to analyze EEG data. Although not pre-registered, we decided on this approach before looking at the data. While the proposed analysis can be a viable alternative, the selection of time windows (250, 500, or other durations) can be somewhat arbitrary. More importantly, we are concerned that trying out an alternative analysis because our original analysis did not yield significant results, may come close to "p-hacking". Furthermore, Figures 5 and 7 show large variations across participants. Even if the proposed analysis shows significant results, these results should still be treated with caution. For all these reasons, we would like to transparently report the results from our originally planned analysis.*

1. _p10: missing 's' after 'participant': "For this, participant completed multi-trial schedules"

Reply: *We have corrected this.*

Competing Interests: No competing interests were disclosed.

Reviewer Report 19 July 2024

https://doi.org/10.21956/openreseurope.19188.r42232

© 2024 Devine S. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

?

Sean Devine
[1] Department of Psychology, McGill University, Montreal, Québec, Canada
[2] Department of Psychology, McGill University, Montreal, Québec, Canada

Overall appraisal ----------

In this paper, Eben and colleagues attempt a replication and extension of Yu et al. (2014). Originally, Yu and colleagues found that self-reported ratings of frustration, motivation, response time, and pressing force varied as a function of goal proximity in a rewarded computerized "blocking" experiment. Eben et al. successfully replicate Yu et al.'s procedure and find in some cases similar results and in others non-consistent results.

Overall, this paper was well-written and approached the replication with good methods and a clear rationale. I especially appreciated the transparency with which the authors reported their findings and described the technical limitations they ran into. As a replication of a study that's been cited over 100 times, I think this paper succeeds in providing additional information about the reliability and size of the original effects.

Below I detail some comments that I think can help further improve this paper.

Main comments ----------

1. I hate to be the reviewer who requests to cite his own work, but in this case I think it is appropriate. In the introduction, the authors discuss both recent and classical work on the goal-gradient hypothesis but neglect the burgeoning literature on goal-gradients in cognitive effort expenditure (Katzir et al., 2020; Emanuel et al., 2022; Devine et al., 2020; 2024). This work seems especially relevant since the current task leverages an attentional task (the arrow task) not too dissimilar from the one in Devine et al. (2024), where we found similar speeding proximity effects. So, at the least, it might be useful to put Yu et al.'s (2014) results (and the current replication) in conversation with some more recent work exploring these proximity effects in psychology. Moreover, the authors also discuss past work which has used EMG to measure affect, but do not discuss more recent work specifically linking EMG activity to cognitive effort exertion (Devine et al., 2024; Berger et al., 2020). This seems like something that would be directly relevant to the current discussion and justification of

EMG in this context. I leave it up to the authors if they choose to include this work and they should not feel compelled to do so just because some of this work is mine, as the reviewer's. I just think some of it is particularly relevant for the topic at hand.

2. One key element which I found unclear throughout the paper was *which* effects Eben and colleagues were aiming to replicate. In their Discussion they state: "Our RT data showed participants sped up in the arrow task the closer they got to the reward. Once they were blocked, participants seemed to be surprised about the outcome and confirmed the outcome with a slowed response [...O]ur current results are not in line with the original study" (p. 14). Yet, in reading Yu et al. (2014), they specifically de-emphasize the confirm-RTs in their analysis: "[Participants] were told that "Confirming the outcome is not a speeded task, and that you have plenty of time (4 sec) to make this response" (p. 167). Further, only RTs to arrows (not confirm-RTs) are reported in Yu et al.'s main text figures and tables. In fact, apart from listing average confirm-RTs, I am not sure where Yu et al. report confirm-RT results whatsoever. Thus, I am not entirely sure what to make of the discussed discrepancy in confirm-RTs. In any case, I am not sure it is correct to position the current replication as strongly inconsistent with the original findings—especially when the majority of behavioral and self-report measures do seem consistent. Perhaps the authors could further clarify their view on the importance of confirm-RTs to the original findings.

3. Regarding statistical power, I think there is a bit more work to be done here. Firstly, I was not clear how the authors computed power for their analyses. They report a benchmark Cohen's d (0.5) which they are able to detect with 80% power. But both their replication analyses (ANOVAs) and exploratory analyses (Bayesian MLMs) do not have effect sizes on this metrics. So perhaps a conversion was done?

   Leaving this question aside, the authors' rightfully highlight that using Yu et al.'s ANOVA-based analysis, they too fail to find a significant confirm-RT effect (replicating the original null finding). However, when using a multilevel approach, this proximity effect on confirm-RTs emerges. I wholly agree that the multilevel approach is the superior one to the ANOVA-based analysis. However, calculating power in an MLM context is quite different than doing so for ANOVA and usually requires simulation. If indeed the MLM approach afforded increased power over the inferior ANOVA approach, then it could be that the original Yu et al. study was simply underpowered to detect the proximity effect on confirm-RTs. Again, it is hard to judge since, from what I can see, Yu et al. do not report or visualize any confirm-RT data in their original publication. But if this is indeed just a power issue (e.g., Yu et al also found numerical increases in confirm-RTs as a function of proximity, but which failed to achieve significance under an ANOVA approach), this would be a helpful piece of information for explaining the discrepancy in results.

4. Going beyond replication-by-significance, it would be interesting to see a brief discussion of relative effect sizes. It is difficult to compare right now because arrow-RT effects are reported by ANOVA in Yu et al. but by MLM in Eben et al. Nevertheless, eyeballing Figure 2 in Eben et al. against Figure 3 in Yu et al., the RT and self-report effects seem remarkably similar (e.g., a roughly ~40ms difference from 1/4 to 4/4 schedules in both samples). At the most, something like a small-telescopes approach could be used (Simonsohn, 2015), but even just a qualitative comparison of the relative effect sizes—particularly for the confirm-RTs (see power issue from comment #3 above)—could be very informative.

5. Some more details about the fEMG analysis would be helpful. What preprocessing steps were taken? Were signals bandpassed filtered, and if so at what frequency range? Were they rectified and low-pass filtered? (see van Boxtel, 2010). Some of these cleaning steps may help more subtle effects emerge, especially under conditions of potentially low power as the authors highlight on p. 13.

   Additionally, here the authors adopted a grand-mean ERP approach for computing fEMG signals (p. 8). Instead, the authors could consider a single-trial approach, in which a summary value for each muscle is computed per trial per subject (e.g., the percent change in maximal constriction from baseline). This approach yields a single scalar summary value per trial and could mitigate power issues by keeping more data in the analysis and permitting a multilevel analysis similar to their RT analysis (something like: "cor ~ mo(proximity) + mo(effort) + (mo(proximity) + mo(effort) | id)"). The downside to this approach is that it precludes a timecourse analysis, but as far as I understand, the timecourse of the EMG signal is less of interest than the comparison of experimental conditions in this case. See Devine et al. (2023) for an example of this approach.

Minor comments ----------

1. Interestingly, the blocking procedure used here closely mirrors that used in classical experiments on goal-gradients in rats (Brown, 1948). I am not sure if this was an inspiration or not for the original design, but if not, it is an interesting coincidence.

2. I very much liked the video of the experiment uploaded to OSF! More researchers should do this—it makes evaluating the methodology much easier.

3. Though Yu et al also did not analyze accuracy rates, I wonder if proximity or effort meaningfully influenced accuracy rates in this task. Or was accuracy too high and invariable for a simple task as this?

References ----------

Berger, A., Mitschke, V., Dignath, D., Eder, A., & van Steenbergen, H. (2020). The face of control: Corrugator supercilii tracks aversive conflict signals in the service of adaptive cognitive control. *Psychophysiology*, *57*(4), e13524.

Brown, J. S. (1948). Gradients of approach and avoidance responses and their relation to level of motivation. *Journal of comparative and physiological psychology*, *41*(6), 450.
Devine, S., & Otto, A. R. (2022). Information about task progress modulates cognitive demand avoidance. *Cognition*, *225*, 105107.

Devine, S., Roy, M., Beierholm, U., & Otto, A. R. (2024). Proximity to rewards modulates parameters of effortful control exertion. *Journal of Experimental Psychology: General*.

Devine, S., Vassena, E., & Otto, A. R. (2023). More than a feeling: physiological measures of affect index the integration of effort costs and rewards during anticipatory effort evaluation. *Cognitive, Affective, & Behavioral Neuroscience*, *23*(4), 1129-1140.

Emanuel, A., Katzir, M., & Liberman, N. (2022). Why do people increase effort near a deadline? An opportunity-cost model of goal gradients. *Journal of Experimental Psychology: General*, *151*(11), 2910.

Katzir, M., Emanuel, A., & Liberman, N. (2020). Cognitive performance is enhanced if one knows when the task will end. *Cognition*, *197*, 104189.

Simonsohn, U. (2015). Small telescopes: Detectability and the evaluation of replication results. *Psychological science*, *26*(5), 559-569.

Van Boxtel, A. (2010, August). Facial EMG as a tool for inferring affective states. In *Proceedings of measuring behavior* (Vol. 2010).

References

1. Berger A, Mitschke V, Dignath D, Eder A, et al.: The face of control: Corrugator supercilii tracks aversive conflict signals in the service of adaptive cognitive control.*Psychophysiology*. 2020; **57** (4): e13524 PubMed Abstract | Publisher Full Text
2. BROWN JS: Gradients of approach and avoidance responses and their relation to level of motivation.*J Comp Physiol Psychol*. 1948; **41** (6): 450-65 PubMed Abstract | Publisher Full Text
3. Devine S, Roy M, Beierholm U, Otto AR: Proximity to rewards modulates parameters of effortful control exertion.*J Exp Psychol Gen*. 2024; **153** (5): 1257-1267 PubMed Abstract | Publisher Full Text
4. Devine S, Vassena E, Otto AR: More than a feeling: physiological measures of affect index the integration of effort costs and rewards during anticipatory effort evaluation.*Cogn Affect Behav Neurosci*. 2023; **23** (4): 1129-1140 PubMed Abstract | Publisher Full Text
5. Emanuel A, Katzir M, Liberman N: Why do people increase effort near a deadline? An opportunity-cost model of goal gradients.*J Exp Psychol Gen*. 2022; **151** (11): 2910-2926 PubMed Abstract | Publisher Full Text
6. Katzir M, Emanuel A, Liberman N: Cognitive performance is enhanced if one knows when the task will end.*Cognition*. 2020; **197**: 104189 PubMed Abstract | Publisher Full Text
7. Simonsohn U: Small telescopes: detectability and the evaluation of replication results.*Psychol Sci*. 2015; **26** (5): 559-69 PubMed Abstract | Publisher Full Text
8. Sato W, Kochiyama T: Crosstalk in Facial EMG and Its Reduction Using ICA.*Sensors (Basel)*. 2023; **23** (5). PubMed Abstract | Publisher Full Text

Is the work clearly and accurately presented and does it engage with the current literature?

Yes

Is the study design appropriate and is the work technically sound?

Yes

Are sufficient details of methods and analysis provided to allow replication by others?

Yes

Are all the source data and materials underlying the results available?

Yes

If applicable, is the statistical analysis and its interpretation appropriate?

Yes

Are the conclusions drawn adequately supported by the results?
Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Cognitive psychology, cognitive effort, goal-gradient effects, EMG analysis, computational modeling, multilevel modeling

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 02 Apr 2025

Charlotte Eben

1. I hate to be the reviewer who requests to cite his own work, but in this case I think it is appropriate. In the introduction, the authors discuss both recent and classical work on the goal-gradient hypothesis but neglect the burgeoning literature on goal-gradients in cognitive effort expenditure (Katzir et al., 2020; Emanuel et al., 2022; Devine et al., 2020; 2024). This work seems especially relevant since the current task leverages an attentional task (the arrow task) not too dissimilar from the one in Devine et al. (2024), where we found similar speeding proximity effects. So, at the least, it might be useful to put Yu et al.'s (2014) results (and the current replication) in conversation with some more recent work exploring these proximity effects in psychology. Moreover, the authors also discuss past work which has used EMG to measure affect, but do not discuss more recent work specifically linking EMG activity to cognitive effort exertion (Devine et al., 2024; Berger et al., 2020). This seems like something that would be directly relevant to the current discussion and justification of EMG in this context. I leave it up to the authors if they choose to include this work and they should not feel compelled to do so just because some of this work is mine, as the reviewer's. I just think some of it is particularly relevant for the topic at hand.

Reply: *Thank you for this suggestion. We have added some of your suggested articles to our introduction of goal gradient on p.4.* "Many goal pursuit activities require individuals to carry out multiple successive actions to obtain a reward. Accordingly, individuals may fail at any stage of the goal pursuit process, which in turn may influence the magnitude of the 'frustration' effect triggered by reward omission or blockage. As individuals make progress in their goal pursuit, their motivation to obtain the goal generally increases (i.e., the goal gradient hypothesis; Hull, 1932). Recent work has shown that people generally exert more cognitive effort when they are provided with information on their progress in a task (Devine & Otto, 2022; Katzir, Emanuel, & Liberman, 2020), and cognitive effort exertion increases near the end of a task (i.e., near goal completion, Devine et al., 2024; Emanuel, Katzir, & Liberman, 2022). Outside the laboratory context, the influence of perceived goal progress on motivation has also been documented. For instance, customers bought coffee more frequently the closer they were to getting a free coffee by collecting enough stamps in a reward program (Kivetz et al., 2006). In addition to influencing people's motivation, goal

progress can also influence the magnitude of the 'frustration' effect when they are blocked from the reward (Amsel, 1958). For instance, Haner and Brown (1955) found that children responded with more force when a failure occurred close to success, compared to when they were still far from achieving their goal." *While facial EMG has indeed been used as an indicator of cognitive effort exertion, in our study we are mainly focussing on the fEMG after receiving the outcome rather than during the task. As such, the fEMG activities here primarily reflect affect triggered by outcomes, rather than cognitive effort exertion. We therefore did not expand on the relationship between fEMG and cognitive effort exertion in our manuscript.*

2. One key element which I found unclear throughout the paper was *which* effects Eben and colleagues were aiming to replicate. In their Discussion they state: "Our RT data showed participants sped up in the arrow task the closer they got to the reward. Once they were blocked, participants seemed to be surprised about the outcome and confirmed the outcome with a slowed response [...O]ur current results are not in line with the original study" (p. 14). Yet, in reading Yu et al. (2014), they specifically de-emphasize the confirm-RTs in their analysis: "[Participants] were told that "Confirming the outcome is not a speeded task, and that you have plenty of time (4 sec) to make this response" (p. 167). Further, only RTs to arrows (not confirm-RTs) are reported in Yu et al.'s main text figures and tables. In fact, apart from listing average confirm-RTs, I am not sure where Yu et al. report confirm-RT results whatsoever. Thus, I am not entirely sure what to make of the discussed discrepancy in confirm-RTs. In any case, I am not sure it is correct to position the current replication as strongly inconsistent with the original findings—especially when the majority of behavioral and self-report measures do seem consistent. Perhaps the authors could further clarify their view on the importance of confirm-RTs to the original findings.
Reply: *As indicated on p.4 (highlighted in the revised version, see also below), we are mainly interested in outcome-related effects of frustration, which included the confirm RT, self-reported surprise, frustration and the confirm force (which due to technical issues was invalid, and hence not included in the manuscript). For consistency with the original work, we also examined the other variables, namely arrow RTs and self-reported motivation in the task.* "The ubiquitous influence of goal progress on motivated behavior requires a better understanding of how such influence arises. Here we specifically focus on how goal progress may influence the magnitude of the 'frustration' effect when individuals are blocked from a reward." Yu and colleagues indeed found no effects on confirm RTs. We think the confirm RT is still an important variable to consider though, for several reasons. First, much previous work on feedback-related behavioral adaptation uses response times as the main indicator, such as post-loss speeding and post-reinforcement pause in the gambling context (and in reinforcement learning more generally), and post-error slowing in the cognitive control field. Using confirm RTs allows us to potentially link the current finding to this body of work on response times. Second, specifically within the gambling context, how quickly people react to previous outcomes (which are captured by confirm RTs here) are related to event frequency, which is how many gambling events one may experience within a certain amount of time. Event frequency is an important structure feature of gambling products and has been proposed to play a key role in gambling-related harms. Examining the 'frustration' effect via confirm RTs therefore also has potential practical implications. *We agree with the reviewer that most of our results are consistent with the previous findings. For instance, when using the original analysis approach by Yu and colleagues, we also observed non-significant effects of proximity and expended effort on confirm RTs. For a better comparison*

between the current and the original results, we included the original plots in Figure 2. This figure shows that most of the behavioral results are highly consistent between the original and our replication study. It also highlights the discrepancy, namely the effects on surprise. The effect of proximity on surprise together with our exploratory analyses on confirm RTs suggest that there might be an orienting response to blocked rewards when the goal proximity was high. This is a finding that is not in line with the original findings. We adjusted our conclusion to weaken the claim that our findings are not in line with the original study.

*3.*Regarding statistical power, I think there is a bit more work to be done here. Firstly, I was not clear how the authors computed power for their analyses. They report a benchmark Cohen's d (0.5) which they are able to detect with 80% power. But both their replication analyses (ANOVAs) and exploratory analyses (Bayesian MLMs) do not have effect sizes on this metrics. So perhaps a conversion was done?
Reply: We have adjusted our sample size justification to be completely transparent (see p. 6) *:* "In the original analysis, Yu and colleagues examined the effects of proximity and expended effort by combining the p values from multiple repeated-measures ANOVAs using the Stouffer's method. As far as we know, no analytical method exists to compute the statistical power for the Stouffer's combined p values. We therefore decided to go beyond the sample sizes used in Yu et al. (between 20 and 27), but still within our resource constraints. We were additionally interested in potential differences in the response force after wins and losses, for which Yu et al. did not report the results. We therefore assumed a medium effect size (Cohen's d = 0.5), for which 32 participants would provide approximately 80% power to detect in a within-subject design. Our response force data, however, was compromised and we never followed up on that specific effect. The eventual sample size of 32 was therefore chosen given all these considerations, to (1) exceed the original sample sizes, (2) be within our resource constraints, and (3) provide sufficient power to detect potential response force differences between wins and losses."

4.Leaving this question aside, the authors' rightfully highlight that using Yu et al.'s ANOVA-based analysis, they too fail to find a significant confirm-RT effect (replicating the original null finding). However, when using a multilevel approach, this proximity effect on confirm-RTs emerges. I wholly agree that the multilevel approach is the superior one to the ANOVA-based analysis. However, calculating power in an MLM context is quite different than doing so for ANOVA and usually requires simulation. If indeed the MLM approach afforded increased power over the inferior ANOVA approach, then it could be that the original Yu et al. study was simply underpowered to detect the proximity effect on confirm-RTs. Again, it is hard to judge since, from what I can see, Yu et al. do not report or visualize any confirm-RT data in their original publication. But if this is indeed just a power issue (e.g., Yu et al also found numerical increases in confirm-RTs as a function of proximity, but which failed to achieve significance under an ANOVA approach), this would be a helpful piece of information for explaining the discrepancy in results.
Reply: *We discuss the possibility of a statistical power issue on p. 14. "*Note that when we used the same data analysis approach as in the original study, we also did not observe statistically significant effects of proximity or effort. This discrepancy in results might arise because the exploratory analysis reported here included proximity and effort simultaneously in one model and analyzed data on each trial (rather than aggregated means in each condition), and thus might have more statistical power to detect an effect.

Yu et al. did not report or visualize the results on confirm RTs, so it is unclear whether a similar pattern was also present in their data."

5. Going beyond replication-by-significance, it would be interesting to see a brief discussion of relative effect sizes. It is difficult to compare right now because arrow-RT effects are reported by ANOVA in Yu et al. but by MLM in Eben et al. Nevertheless, eyeballing Figure 2 in Eben et al. against Figure 3 in Yu et al., the RT and self-report effects seem remarkably similar (e.g., a roughly ~40ms difference from 1/4 to 4/4 schedules in both samples). At the most, something like a small-telescopes approach could be used (Simonsohn, 2015), but even just a qualitative comparison of the relative effect sizes—particularly for the confirm-RTs (see power issue from comment #3 above)—could be very informative.
Reply: *We added the original study's plots to our plots for a better comparison (p. 24). This plot now shows that for arrow RT, self-reported motivation, and self-reported frustration, the effect sizes were highly similar between the current study and Yu et al. It also highlights the difference in self-reported surprise. For confirm RT, unfortunately, since Yu et al. did not report detailed results, such a comparison is not possible. We also added the analyses on the RTs to the arrows (as done in the original study) in our results table and in the main text (p.11).*

6. Some more details about the fEMG analysis would be helpful. What preprocessing steps were taken? Were signals bandpassed filtered, and if so at what frequency range? Were they rectified and low-pass filtered? (see van Boxtel, 2010). Some of these cleaning steps may help more subtle effects emerge, especially under conditions of potentially low power as the authors highlight on p. 13.
Reply: *We report these specifics in the paragraph "Electrophysiological recording and data reduction" on p. 9.* "According to common guidelines and practices (Fridlund & Cacioppo, 1986; Korb et al., 2017), we attached shielded EMG electrodes over the right zygomaticus major and corrugator supercilii areas and placed a ground electrode at the top of the central forehead. Using the Biopac EMG100C amplifier with the MP150 module and the Acqknowledge software (www.biopac.com), we acquired continuous EMG at 1000 Hz sampling rate, with a 100 Hz high-pass and a 500 Hz low-pass. The data reduction and preprocessing pipeline was based on Korb et al. (2017), van Boxtel (2010) and Whelan (2003), and adjusted for our design. We applied a 20-200 Hz band pass filter and full-wave rectified the data by flipping all negative values of the signal to positive values. To exclude high-frequency noise, we then applied a 40 Hz low-pass filter. We epoched the data time-locked to the outcome presentation (win or block) from -750 ms to 2000 ms. Before cleaning the data, we compared the behavioral and EMG data to identify extra "ghost" triggers (triggers that Biopac picked up but were never sent) and deleted these. Then, we excluded error trials and trials where the baseline was more than 2 standard deviations from the participant's baseline mean. The signal was normalized by expressing it as a percentage of the baseline mean per trial. This was done to ensure comparability between individuals and muscles, as well as to control for changes in muscle tone over time. Finally, we excluded the epochs with responses two standard deviations higher than the participant's condition mean."

7. Additionally, here the authors adopted a grand-mean ERP approach for computing fEMG signals (p. 8). Instead, the authors could consider a single-trial approach, in which a summary value for each muscle is computed per trial per subject (e.g., the percent change

in maximal constriction from baseline). This approach yields a single scalar summary value per trial and could mitigate power issues by keeping more data in the analysis and permitting a multilevel analysis similar to their RT analysis (something like: "cor ~ mo(proximity) + mo(effort) + (mo(proximity) + mo(effort) | id)"). The downside to this approach is that it precludes a timecourse analysis, but as far as I understand, the timecourse of the EMG signal is less of interest than the comparison of experimental conditions in this case. See Devine et al. (2023) for an example of this approach.
Reply: *Following the reviewer's suggestion, we took a single-trial approach. For each participant, muscle and trial, we found the maximum percent change from baseline in the 250 – 2000 ms window after outcome presentation. The trial-level corrugator activity data after blocked outcomes were then standardized within participants and analyzed with a mixed-effects model in brms. The pseudo-code for the brms model was: corrugator ~ mo(proximity) + mo(effort) + (mo(proximity) + mo(effort)|participant). We used the student's t distribution as the likelihood function, to reduce the potential influence of extreme values. The coefficients for proximity and effort were not statistically credible, estimate = 0.002, 95% CI = [-0.038, 0.043], and estimate = -0.016, 95% CI = [-0.059, 0.022], respectively. We similarly analyzed the trial-level zygomaticus data after blocked outcomes. Here, the effect of effort was again not statistically credible, estimate = 0.004, 95% CI = [-0.023, 0.039]. The effect of proximity was statistically credible, estimate = 0.022, 95% CI = [0.002, 0.041], indicating that as participants were blocked proximal to a reward, the activity of zygomaticus tended to increase. As noted in the main text, the number of schedules within each condition was small, which was the reason why we decided to compare the win versus blocked outcomes, rather than examining the effects of different blocked schedules on fEMG data, as suggested here. Figures 5 and 7 also show large variations across participants in the fEMG data. It is thus unclear how valid and reliable the results obtained above are. For these reasons, we prefer to keep our original grand-mean ERP approach in the manuscript.*

Minor comments ----------

1. Interestingly, the blocking procedure used here closely mirrors that used in classical experiments on goal-gradients in rats (Brown, 1948). I am not sure if this was an inspiration or not for the original design, but if not, it is an interesting coincidence.

Reply: *That is indeed very interesting. Since Yu and colleagues did not mention Brown (1948) at all, we assume that this is an interesting coincidence.*

1. I very much liked the video of the experiment uploaded to OSF! More researchers should do this—it makes evaluating the methodology much easier.

Reply: *Thanks so much for this feedback.*

1. Though Yu et al also did not analyze accuracy rates, I wonder if proximity or effort meaningfully influenced accuracy rates in this task. Or was accuracy too high and invariable for a simple task as this?

Reply: *As suggested, we analyzed accuracy as a function of proximity and effort, in a multilevel logistic regression. The pseudo-code for the* brms *model was: accuracy ~ mo(proximity) + mo(effort) + (mo(proximity) + mo(effort)|participant). The mo() function indicates that the variables were modelled as monotonic predictors. We observed that as proximity increased, accuracy decreased, estimate = -0.298, 95% CI = [-0.441, -0.160]. The effect of effort was not*

statistically credible, estimate = -0.068, 95% CI = [-0.201, 0.068]. These results, combined with the observation that participants tended to respond more quickly as proximity increased, suggest that proximity modulates the speed-accuracy trade-off. As they were getting close to a reward, they emphasized speed over accuracy, making responses more quickly at the cost of being less accurate. This result seems to differ from that of Devine et al. (2024), where participants responded more quickly, but not less accurately, when they were close to a reward. Our task put a strong emphasis on speed (i.e., beating the computer by responding quicker than an unknown deadline), and the arrow task was rather simple, potentially explaining the difference in the results. Devine et al. showed that their results were best explained by reduced response caution and increased information processing when participants were proximal to a goal. We speculate that in our task, proximity might similarly reduce response caution (thereby explaining the speed-accuracy trade-off), but did not strongly increase information processing due to the simple nature of the task. While this is not the focus of Yu et al. nor our current replication attempt, the task with the schedules used here might be a useful tool to disentangle the effects of proximity and expended effort in cognitive effort exertion.

Competing Interests: No competing interests were disclosed.