Journal of Experimental Psychology: General

Temporal Metacognition: Direct Readout or Mental Construct? The Case of Introspective Reaction Time

Nathalie Pavailler, Wim Gevers, and Boris Burle

Online First Publication, January 23, 2025. https://dx.doi.org/10.1037/xge0001708

CITATION
Pavailler, N., Gevers, W., & Burle, B. (2025). Temporal metacognition: Direct readout or mental construct? The case of introspective reaction time. *Journal of Experimental Psychology: General*. Advance online publication. https://dx.doi.org/10.1037/xge0001708

AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Experimental Psychology: General

© 2025 American Psychological Association
ISSN: 0096-3445

https://doi.org/10.1037/xge0001708

Temporal Metacognition: Direct Readout or Mental Construct? The Case of Introspective Reaction Time

Nathalie Pavailler[1], Wim Gevers[2], and Boris Burle[1]

[1] Centre de Recherche en Psychologie et Neurosciences, Centre National de la Recherche Scientifique, Aix-Marseille Université

[2] Center for Research in Cognition and Neurosciences, Neurosciences Institute, Université Libre de Bruxelles

Deciphering whether and which mental processes are accessible for metacognitive judgments is a key question to understand higher cognitive functions. Paralleling the crucial role of reaction times (RT) for unraveling the temporal sequence of mental processes, a comparable chronometric approach can be employed at the second-order level through introspective reaction times (iRT) measures. Although mean iRT correlate with mean RT, suggesting good metacognitive abilities, this would not necessarily imply a direct readout of the duration of the underlying processes as participants may instead rely on inferences based on other salient, nontemporal, cues. In the present study, two experiments investigated information at the basis of iRT. In visual choice reaction time tasks, participants were asked to report their RT on a visual analog scale after each trial. Thanks to linear regression analyses, we could evidence that trial-by-trial RT and iRT were strongly correlated, indicating a good readout of RT duration, but also that subjective evaluation was systematically biased by some experimental conditions. In addition, with electromyographic recordings, each single trial RT could be fractionated into premotor and motor times, allowing to investigate the relative contribution of each subprocess to iRT. This revealed that participants access both decision and motor execution durations. Results show that participants can access the duration of their mental processes but that this readout can be biased by nontemporal cues. The proposed methodology allows to dissociate the two.

Public Significance Statement

Do we have direct access to the content of our mental life? Or do we merely reconstruct it a posteriori? Can we monitor the processes at play during the realization of a task? To advance on these questions, the present study explores the capacity of human participants to estimate the time they took to perform a perceptual discrimination task. A new data processing rationale, coupled with electrophysiological measures, allowed to reveal that participants can reliably access the duration of their processing operations but also that such estimation can be biased by external factors and a priori knowledge. These observations constitute important information to further understand how humans build metacognitive representations of (the duration of) their actions.

Keywords: reaction times, introspection, performance monitoring, electromyography

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Joo-Hyun Song served as action editor.

Boris Burle https://orcid.org/0000-0001-8179-9034

Data and codes used in this study are available on the Open Science Framework at https://osf.io/3rzak/. Parts of the data of this article have been presented at the 22nd conference of the European Society for Cognitive Psychology and at the 26th annual meeting of the Association for the Scientific Study of Consciousness.

Nathalie Pavailler is funded by a PhD grant from the French Ministry for Research. This work has received support from the French government under the Programme "Investissements d'Avenir," Initiative d'Excellence d'Aix-Marseille Université via A*Midex and Agence Nationale de la Recherche fundings (Grants AMX-19-IET-004 and ANR-17-EURE-0029), and by the Agence Nationale de la Recherche (Grant ANR-22-CE28-0027 awarded to Boris Burle). This research was funded in whole, or in part, by the French National Research Agency (ANR; Grant ANR-22-CE28-0027). For the purpose of open access, the author has applied a CC BY public copyright license to any Author Accepted Manuscript version arising from this submission. The authors thank D. Louber and D. Paleressompoulé for developing the response device and F.-X. Alario for fruitful discussions and advice.

Nathalie Pavailler played a lead role in data curation and formal analysis and an equal role in conceptualization, methodology, software, writing–original draft, and writing–review and editing. Wim Gevers played a supporting role in conceptualization, methodology, and validation and an equal role in writing–review and editing. Boris Burle played a lead role in conceptualization, funding acquisition, project administration, and supervision, a supporting role in data curation, formal analysis, and investigation, and an equal role in methodology, resources, software, validation, writing–original draft, and writing–review and editing.

Correspondence concerning this article should be addressed to Boris Burle, Centre de Recherche en Psychologie et Neurosciences, Centre National de la Recherche Scientifique, Aix-Marseille Université, 3 Place Victor Hugo, 13331 Marseille, cedex 3, France. Email: boris.burle@univ-amu.fr

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

In many aspects of life, understanding the cognitive processes that led to our behavior is essential, especially for learning and academic achievement (Stanton et al., 2021). A fundamental question, however, is whether we have access to our own cognitive processes or whether we infer them a posteriori based on our behavior and other environmental cues (Nisbett & Wilson, 1977). Addressing this question is particularly difficult because, as experimenters, we have no direct access to these processes but must infer them from observable behaviors. In the last decades, scholars have focused on somehow simpler forms of metacognition, generally referred to as "performance monitoring." Indeed, keeping our behavior effective requires to constantly evaluate the outcomes of our actions so that adjustments can be made when they differ from planned goals. When external feedback is absent, the only information about action's outcome comes from an internal and subjective evaluation of performance, constituting a large part of metacognition research (Yeung & Summerfield, 2012).

Besides accuracy analysis, our understanding of information processing operations largely owns to reaction time (RT) measures that allowed to reveal the temporal sequencing and relative durations of mental processes (Donders, 1969; Luce, 1986; Meyer et al., 1988; Posner, 1978; Sternberg, 1969). Recent evidence suggests that such a chronometric approach can also be useful to reveal core aspects of metacognition (Corallo et al., 2008), by asking participants to estimate the time it took them to perform a first-order cognitive task.

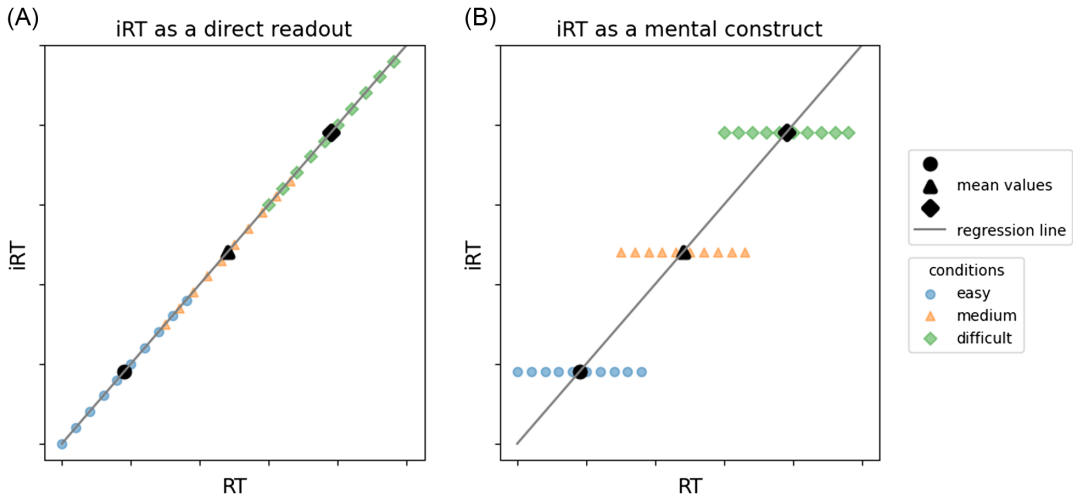
Metacognitive abilities in the temporal domain are supported by recent data indicating that participants can introspectively evaluate the accuracy of their timing performance: In temporal reproduction tasks, participants turned out to be aware of the direction and amplitude of their temporal errors (Akdoğan & Balcı, 2017), leading to the emergence of temporal error monitoring field of study (Akdoğan & Balcı, 2017; Kononowicz et al., 2019; Kononowicz & van Wassenhove, 2019). In such situations, accuracy and timing of the first-order decision are related, as the goal is to explicitly produce a target duration. There is also evidence that participants can evaluate the duration of their cognitive processing in nontemporal decision tasks, potentially using it to set an optimal response threshold (Balci et al., 2011). Indeed, in the early 70s, Sanford (1970) showed that participants were able to meaningfully rate their RTs in four categories from "fast" to "slow." Corallo et al. (2008) relaunched this field of study with a new methodology asking participants to provide on each trial a subjective estimate of their own RT using a visual analog scale (VAS), providing a quantified second-order measure they termed "introspective RT" (iRT). Following studies have also used other methods to assess iRT, including temporal reproduction (Bryce & Bratzke, 2015; Klein & Stolz, 2018), method of constant stimuli (Bratzke & Bryce, 2016), and timeline method (Bryce & Bratzke, 2017), with comparable results. Paralleling the logic used on RT, Corallo et al. (2008) reasoned that if some experimental manipulations impacting RT also impact iRT, this would be an argument that the process affected by the manipulation is introspectively accessible. In contrast, if iRT is blind to this manipulation, this would suggest that the underlying process is not accessible. Many previous studies have indeed observed dissociations between RT and iRT in dual-task paradigms (Bratzke et al., 2014; Bryce & Bratzke, 2015, 2017; Corallo et al., 2008; Marti et al., 2010). The RT/iRT relationships in single tasks remain much less explored. In this latter context, using a number comparison task, Corallo et al. (2008) reported that, across different conditions of task difficulty, mean RT and mean iRT increased in the same way and were consequently positively correlated. While this should necessarily be observed if participants are able to access the duration of their processes, on itself it is not sufficient to affirm that iRT is a direct readout of the temporal content of RT. It is indeed well established that participants can use salient cues to indirectly infer their cognitive processes (Nisbett & Wilson, 1977). For example, iRT could be based on the perception of task difficulty (Bryce & Bratzke, 2014): With some experimental conditions being (objectively and subjectively) more difficult, participants might infer that their RT is likely lengthened in such conditions, leading to an overall longer mean iRT. Although such a strategy could generate strong correlations between mean RT and mean iRT, it would not represent a direct readout of the duration of the processes constituting RT. Such potential confounds are depicted in Figure 1.

In both panels, the black dots represent the (same) observed correlation between mean iRT and mean RT across three experimental conditions. But such correlation on mean may underlie very different relationships at the trial level, illustrated by the colored symbols. On Panel A, the correlation on means is paralleled by an identical correlation at the trial level, within each condition, supporting a direct readout of the true RT duration. However, if participants infer their RT based on, for example, task difficulty, the correlation on means may actually correspond to the situation depicted in Panel B: While the means are strongly correlated, such correlation completely vanishes at the trial level within each condition (a form of Simpson's paradox; Kievit et al., 2013; Simpson, 1951), supporting a reconstruction of RT duration from the different experimental conditions. Besides revealing potential confounds, Figure 1 also points to a solution to distinguish the possible outcomes, thanks to linear models (see Bratzke & Bryce, 2019, 2022; Desender et al., 2017; Questienne et al., 2018, for comparable use in different contexts). If one considers iRT as the dependent variable and both RT and conditions as the predictors, two different statistical patterns are to be expected. In both cases, there would be an effect of conditions on both iRT and RT. However, when the predictor RT is added, and in the case of a direct readout (Panel A), it should entirely explain the differences in mean between experimental conditions; one hence expects conditions to become not significant anymore, all the variance being explained by RT. In contrast, in the case of a mental construct (Panel B), RT should not predict iRT, and only the effect of conditions should be significant. Obviously, intermediate cases are possible, where a true relationship between iRT and RT does exist, but can be biased by conditions. With this methodological approach, the first goal of the present study is hence to dissociate the part of iRT that is temporally readout from the part that is mentally reconstructed from the experimental conditions.

On the other hand, RT is a compound measure and there is a general agreement that it can, at the minimum, be decomposed into decision-related time(s) and nondecision times (Luce, 1986; Ratcliff, 1978; Ratcliff et al., 2016), the latter being composed of stimulus encoding and response execution times. Encoding, decision, and execution are generally thought as being sequential with additive durations (Ratcliff, 1978; Ratcliff et al., 2016). A critical question is whether participants have access to the duration of all the different operations or only to a subset of them. Besides dual-task manipulations, iRT studies mostly used manipulations of decision time. Corallo et al. (2008) and Bratzke and Hansen (2024) manipulated the decision time by changing the number notation as

Figure 1

Scenario That can Underlie a Linear Association Between Mean iRT and Mean RT



Note. (A) iRT is purely based on the duration of RT. (B) iRT is purely inferred from task conditions. Each colored (gray) point represents an individual trial, and black points represent the means. If performed across means, correlations are identical between (A) and (B). iRT = introspective reaction time; RT = reaction time. See the online article for the color version of this figure.

well as the numerical distance in number comparison tasks, and Bryce and Bratzke (2014) used different levels of stimulus degradation. In standard drift diffusion modeling, those variables impact drift rate and are hence considered to affect the difficulty of perceptual decisions. The authors observed that task difficulty affected both RT and iRT. To our knowledge, only one study investigated nondecisional processes in introspective judgment. Miller et al. (2010) used a motor programming manipulation with different response complexity and asked for subjective reports of decision and response times on a rotating clock. However, the obtained results were ambiguous. One likely reason is that, since the authors had no way to directly measure the durations of decision and nondecision processes, they inferred them based on unverifiable and unwarranted psychological and physiological assumptions. Note that even outside the context of introspection, nondecision stages have been much less studied. As a matter of fact, nondecision times are often referred to as "residual times," illustrating the little interest in the field of decision making. However, it is possible to modulate the duration of motor execution, for example, by increasing the force needed to produce a response (Burle et al., 2002; Weindel et al., 2021). Whether this would also affect iRT, and how, remains unknown. While an absence of effect of force manipulation on iRT would evidence that motor execution time cannot be accessed by participants, its presence would not necessarily imply a direct readout, for the reason explained above. One hence needs to have an objective measure of RT's subcomponents to be able to assess if and how they contribute to iRT. The use of electromyographic (EMG) recordings provides a solution by allowing to fractionate RT (Botwinick & Thompson, 1966; Burle et al., 2002): Based on the bursts of muscular activities linked to the behavioral responses, one can define the premotor time (PMT, from stimulus onset to EMG burst onset) and the motor time (MT, from EMG onset to behavioral response) as measures of decision and motor execution times, respectively (Possamaï et al., 2002; Weindel et al., 2021). Note that, per construction, RT = PMT + MT. Extending Corallo et al.'s (2008) approach, one may wonder whether iRT = iPMT + iMT, with iPMT and iMT being introspective PMT and MT, respectively.

To assess this question, three models will be compared:

$$\text{iRT} \sim \text{RT} + \text{exp\_factors}, \quad (1)$$

$$\text{iRT} \sim \text{PMT} + \text{exp\_factors}, \quad (2)$$

$$\text{iRT} \sim \text{PMT} + \text{MT} + \text{exp\_factors}. \quad (3)$$

Depending on which model better predicts the data, different conclusions will be drawn. If Model 1 (Equation 1) best accounts for the data, this would suggest that both premotor and motor components are accessible and taken into account to estimate RT. Note, however, that such an outcome could also be compatible with iRT simply reflecting an estimation of the duration between two external events, corresponding to the global RT: a perceptual one (the stimulus onset) and a motor one (the behavioral response); in such case, it may not reflect any introspective process at all.

In case Model 2 (Equation 2) is preferred, this would indicate that introspection is blind to the motor components of the task. Note that in this case, the externally defined duration estimation process would not hold anymore, as no external event marks the boundary between decision and motor execution. Functionally, this would indicate that the introspective system can track the time it takes to reach a decision but ignores everything following this point.

Model 3 (Equation 3) deserves a bit of a comment. Indeed, by construction, PMT + MT = RT. One may hence first consider that Model 3 is equivalent to Model 1. There is, however, a major difference: Model 3 allows PMT and MT to have different weights (different β) while Model 1, implicitly, forces them to have the same weight. If Model 3 better explains the data, this would indicate that (a) both PMT and MT are accessible, (b) they have different weights, and (c), for the reasons explained for Model 2, this would be incompatible with an external duration estimation process.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

In the present study, we used quantified introspection to assess if participants can directly readout the duration of their cognitive processes in an orientation detection task with different stimulus angles, within two experiments. The first experiment was run online and aimed at (a) refining the relation between individual RTs and iRTs in a simple choice reaction time task and (b) dissociating temporal information from other cues linked to experimental conditions that could be used to infer RT. This was further addressed in the second lab-based experiment, in which we added a response force manipulation. Stimulus angle and response force allowed to manipulate RT in two different ways, by selectively affecting decision and motor execution times. Encoding time was constant, as the physical properties of the stimuli remained the same. Finally, we used RT fractioning with EMG, to explore whether decision and execution processes are both introspectively accessible for the iRT judgment.

Experiment 1

The main goal of Experiment 1 was to assess the robustness of the trial-by-trial relation between RT and iRT. The study was performed online, allowing to recruit a larger sample of participants and thus to estimate the between-subject consistency of this link. The second aim was to investigate to what extent iRT reflects a direct readout of the temporal information contained in RT and/or is inferred from other information linked to the task (see Figure 1).

Method

This study was programmed with PsychoPy (Peirce et al., 2019) and exported to https://pavlovia.org (v2020.2) to be run online.

Participants

To recruit participants, the link to the online experiment was transmitted via various means of communication within the university and lab networks, as well as social media. The experiment was also distributed to the volunteer base of the Cognitive Science Information Network (Relais d'Informations sur les Sciences de la Cognition, a unit of the French Centre National de la Recherche Scientifique). We recruited 148 participants. Fifty-eight were excluded because of high error rates (>50%) or nonrespect of the instructions.[1] Thus, 90 participants (M_{age} = 28.1 years) were included in the analyses. Sixty-one of them reported their gender as female, and 29 as male. Fourteen were left-handed. All participants performed the experiment online, on their own computers. The experiment was self-paced by the participants and took approximately 30 min to complete. Before the beginning of the experiment, a consent form, an information notice, and instructions were presented on the screen and participants had to validate them. The study was approved by the Aix-Marseille University ethical committee (No. 2020-12-03-008).

Stimuli and Task

Stimuli were vertical Gabor patches with a size of 0.7/0.7 height units (relative to the height of the window) and a spatial frequency of five cycles/stimulus. Their contrast was set to 0.8 on a scale from 0 (*uniform gray*) to 1 (*black and white*). The stimulus could be tilted either clockwise or counterclockwise (factor "orientation") by either 1° or 7° (factor "angle"). Participants had to respond as a function of the orientation of the patch. Orientation and angle were randomized across trials in a fully factorial design. Participants performed two tasks: a choice reaction time task and a Go/No-Go task (see below). They performed one block of 100 trials of each task, with 25 trials in each elementary cell (Angle × Orientation). The order of the two tasks was counterbalanced between participants. Before each one of them, a training block of 12 trials was run. All trials had the same time course (see Figure 2): After a 2-s presentation of a fixation cross, a Gabor patch was presented at the center of the screen. The task consisted of responding according to its orientation. Responses were given with the keyboard. In the choice block, participants had to press the "S" key with their left hand when the stimulus was counterclockwise and the "L" key with their right hand when the stimulus was clockwise. In the Go/No-Go block, participants had to press the "S" key with their left hand when the stimulus was counterclockwise and had not to respond when the stimulus was clockwise. Go/No-Go RTs are supposed to be shorter compared to choice RTs, revealing differences in terms of response selection complexity between the two tasks (Donders, 1969; Ulrich et al., 1999; Vidal et al., 2011). Participants were explicitly asked to be as fast and accurate as possible. Responses had to be given in less than 1 s. The stimulus remained on the screen until the response. After each response, a VAS was presented on the screen. The scale ranged from 0 to 1 s and was labeled every 500 ms. Participants had to click on the scale with their mouse or touchpad to evaluate their RT, which was explicitly defined as the time interval between the appearance of the stimulus and the response. Participants had unlimited time to do this evaluation. Once they did so, a red marker was placed at the corresponding location on the scale and remained present for 500 ms before the beginning of the next trial. In trials in which no response was given (>1 s or No-Go trials), the scale appeared in gray for 2 s without the possibility to interact with it. During the training blocks (12 trials), feedback concerning the RT was given to participants on the same scale in the form of a green marker presented for 2 s. No feedback was given during the experimental blocks.

Data Analysis

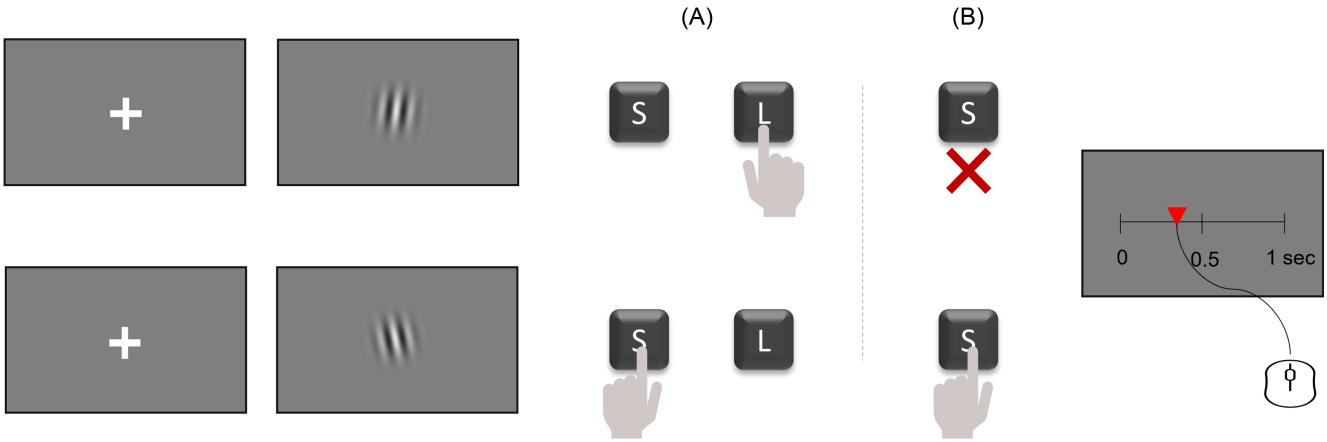
Median absolute deviation (MAD) was computed on the evaluation times to exclude trials for which we can assume that participants were not doing the evaluation properly. Trials with an evaluation time inferior to the median −2 MAD and superior to the median +5 MAD were then excluded (less than 3% of trials). No other censoring of the data was applied. Statistical analyses were performed by means of repeated-measures analyses of variance (ANOVAs), correlations, and mixed models, using Python packages "pingouin" (Vallat, 2018) and "statsmodels" (Seabold & Perktold, 2010) and R package "lme4" (Bates et al., 2015). In the linear mixed model analysis, iRT was used as the dependent variable, with RT, angle, and orientation as fixed effects and participants as random effects, with both random intercepts and random slopes for each predictor. Predictors were chosen based on comparisons of the full model with models including a subset of predictors, with likelihood ratio tests and comparison of the "Akaike information criterion" (AIC; Akaike, 1974). AIC incorporates the likelihood of a model while also penalizing for the number

[1] Nonrespect of the instructions could include no responses in one or several experimental conditions, performing twice the same task, or using the wrong response keys.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Figure 2

Experimental Paradigms of (A) Choice and (B) Go/No-Go Blocks of Experiment 1



Note. The symbol "+" represents the fixation cross. See the online article for the color version of this figure.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

of parameters, with the lowest AIC being preferred. The full model always outperformed the reduced ones (see Appendix A for the AIC of all tested models). The model with interactions was compared to the model without interactions with a likelihood ratio test, and the latter was preferred, $\chi^2(4) = 5.669$, $p = .2253$. Furthermore, given the question of interest (temporal vs. other cues), fixed effects were judged appropriate enough to address it powerfully.

Results

Participants performed two tasks, a Go/No-Go and a choice one, distinguished by the complexity of the decision involved. However, this task manipulation did not succeed: When comparing the left hand (used in both tasks), no difference was observed between the two tasks, neither on RT, $t(89) = -1.14$, $p = .26$, nor on iRTs, $t(89) = 0.19$, $p = .85$. Although speculative, this absence of difference could be due to a lack of control of hand position, as the experiment was conducted online. Indeed, if the participants kept their right finger on the response key during the Go/No-Go task, even if not relevant for this task, this might have been sufficient to cancel the advantage of the Go/No-Go choice. Consequently, for sake of simplicity and coherence with Experiment 2, only results from the choice task will be presented below.[2] Analyses of omissions are reported in Appendix B. In a first set of analyses, only correct trials were included.

Effect of Experimental Conditions on RTs and iRTs

The effect of experimental conditions on mean RT and mean iRT was studied using two-way ANOVAs with angle and orientation as independent within-subjects variables. RTs were affected by both factors: They were slower when angle was smaller, $F(1, 89) = 546.95$, $p < .001$, $\eta_p^2 = 0.86$, and when stimuli were counterclockwise, $F(1, 89) = 11.16$, $p = .001$, $\eta_p^2 = 0.11$.[3] No interaction was found between these two factors, $F(1, 89) = 2.28$, $p = .13$. iRTs were also affected by angle: They were higher when angle was smaller, $F(1, 89) = 132.49$, $p < .001$, $\eta_p^2 = 0.60$. Contrary to RT, no effect of orientation was observed, $F(1, 89) = 1.0$, $p = .32$, but the two factors interacted, $F(1, 89) = 5.91$, $p = .017$, $\eta_p^2 = 0.06$. Descriptive statistics for each experimental condition are presented in Table 1.

Correlation Between RT and iRT

Since both RTs and iRTs were affected by stimulus angle, correlation analyses were performed to quantify the link between these two measures. To assess the common within-individual association between trial-by-trial RTs and iRTs, we computed repeated-measures correlation, allowing to overcome measured variance between participants (Bakdash & Marusich, 2017). Repeated-measures correlation analysis revealed a positive correlation between RTs and iRTs, $r_{rm}(7,371) = 0.56$, $p < .001$. Correlation coefficients were also computed for each participant using Pearson's correlation (see Figure 3). Most subjects (77/90, considering an $\alpha = .05$) presented a positive correlation between RTs and iRTs (range = −0.06–0.89, $M = 0.56$, $SD = 0.23$; see Appendix L for examples of individual data).

Dissociating RT and Experimental Factor Contributions to iRT

We evaluated whether participants based their evaluation solely on time information or whether nontiming information could also impact the subjective evaluation of their timing performance. Paralleling the correlation observed above, mixed model showed a main effect of RT ($\beta = 0.54$, $t = 17.71$, $p < .001$). More importantly, effects of angle ($\beta = -0.002$, $t = -3.5$, $p < .001$; see Figure 4A) and orientation ($\beta = -0.015$, $t = -4.47$, $p < .001$; see Figure 4B) survived the introduction of RT as a regressor, indicating that, for the same objective RT, iRT was higher for smaller angles and for clockwise stimuli. As explained in the introduction, such factor effects indicate a bias in subjective

[2] We nonetheless performed the same analyses on the Go/No-Go task, which showed similar results as the choice one, reported in Appendix C.

[3] Note that, although we describe this effect as an orientation one, this factor is perfectly confounded with the response side. It is hence impossible to know whether it is a perceptual—stimulus orientation—or a motor—response side—effect.

Table 1
Means of the Different Variables for Each Experimental Condition of Experiment 1

	Orientation							
	−7		−1		1		7	
Dependent variable	*M*	95% CI	*M*	95% CI	*M*	95% CI	*M*	95% CI
Omissions (%)	0.8	[0.2, 1.5]	7.6	[5.5, 9.7]	7.2	[4.9, 9.5]	0.9	[0.1, 1.7]
Accuracy (%)	97	[96, 98]	76	[72, 80]	84	[80, 87]	97	[96, 98]
RT (s)	0.463	[0.458, 0.467]	0.566	[0.559, 0.574]	0.547	[0.541, 0.554]	0.452	[0.447, 0.456]
iRT (s)	0.472	[0.467, 0.477]	0.545	[0.537, 0.552]	0.544	[0.537, 0.551]	0.481	[0.476, 0.487]

Note. The 95% CI indicates the confidence interval around the mean. Statistics on omissions and accuracy are reported in Appendix B. RT = reaction time; iRT = introspective reaction time.

temporal evaluation as it reveals that, for the same objective duration, RTs for smaller angles, or for clockwise stimuli, were judged longer. This is further supported by a mediation analysis provided in Appendix D.

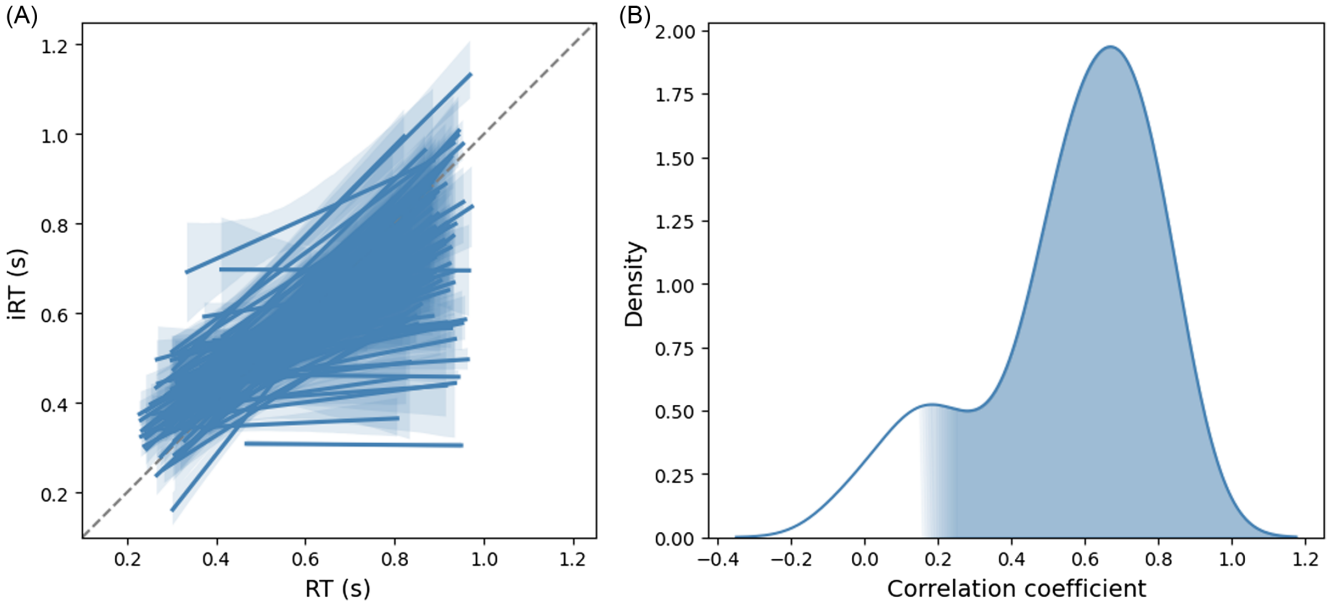
We have so far concentrated on correct trials. Finally, to assess if RTs of errors were estimated in the same way as correct ones, we used a mixed model with iRT as the dependent variable and RT and response accuracy as the predictors. In addition to the main effect of RT ($\beta = 0.57$, $t = 17.67$, $p < .001$), an effect of accuracy ($\beta = 0.014$, $t = 2.38$, $p = .017$) was observed, with smaller iRT for error trials, even though RTs were shorter on correct trials, $t(87) = 6.7$, $p < .001$.

Discussion

Experiment 1 reproduced the results of Corallo et al. (2008) with a similar modulation of mean RT and mean iRT by an experimental modulation of perceptual difficulty in an orientation discrimination task. We confirmed that participants have a good representation of the time needed to respond to a stimulus by looking at trial-by-trial correlations between RT and iRT and that subjective reports of RT given on a visual scale can be reliable. We refined this observation with a larger sample of participants, most (about 85%) of them having a significant metacognitive sensitivity. Interestingly, it was not the case for all of them, and around 15% of participants could not perform the task efficiently and did not present any correlation. Note that, since the experiment was run online, hence without proper control by the experimenter, this could simply reflect a lack of engagement in the task.

Although RT and iRT covariate, systematic biases occurred, meaning that iRT depends, to some extent, on other information than the actual RT duration. Indeed, in the mixed model analysis, effects of orientation and angle survived the global effect of RT duration. First, iRT were overestimated for clockwise stimuli despite faster RT in this condition. This could reflect the fact that participants could be overestimating response times of their fastest side, the same

Figure 3
Experiment 1: Individual Correlations Between iRT and RT

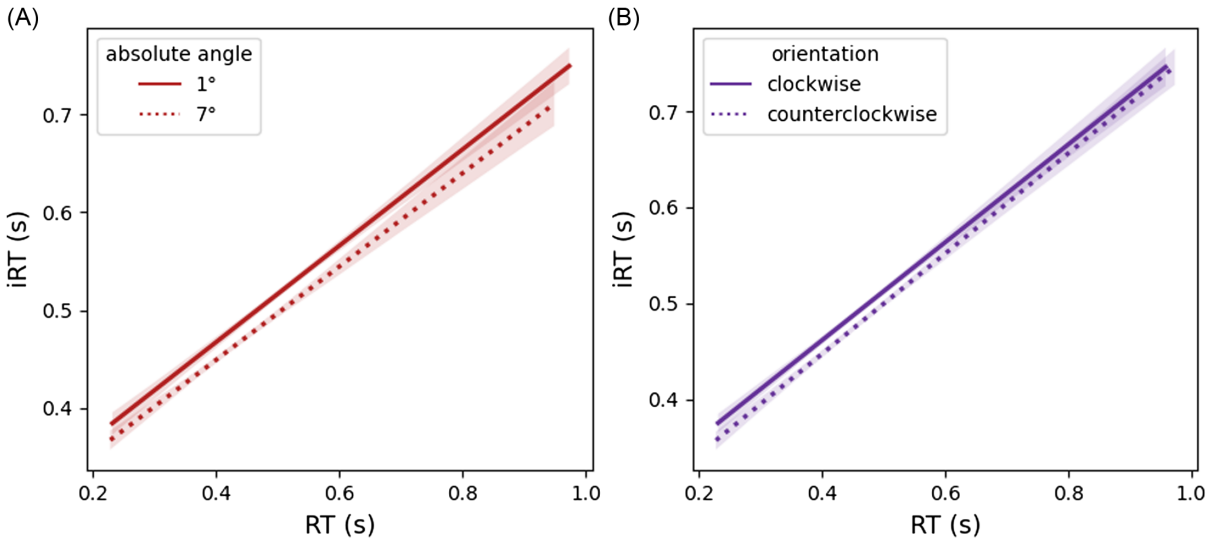


Note. (A) Individual trial-by-trial correlations between RT and iRT. Each regression line represents a participant. The gray dashed line is the $x = y$ line. (B) Distribution of individual correlation coefficients (r). Blue (gray) zone corresponds to positive correlations and white zone to null correlations. The zone of uncertainty concerning correlation significance is represented by a linear gradient around critical r values for an α of .05 and minimal and maximal numbers of observations ($n = 70$ and $n = 100$). RT = reaction time; iRT = introspective reaction time. See the online article for the color version of this figure.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Figure 4

Experiment 1: Linear Regression Between Trial-by-Trial RTs and iRTs for (A) the Two Stimulus Angles and (B) the Two Orientations



Note. RT = reaction time; iRT = introspective reaction time. See the online article for the color version of this figure.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

way higher performers tend to underestimate their skills (Kruger & Dunning, 1999). Alternatively, this could reflect the presence of some compatibility effect between the orientation of the stimulus or the response side, both being confounded in correct trials, and the visual scale used to represent the mental timeline. Participants could have a tendency to use the right part of the scale, corresponding to higher durations, after giving a right response or seeing a stimulus pointing right (Droit-Volet & Coull, 2015).

Second, for the same effective RT duration, participants also estimated themselves slower for small angles, that is, in the condition for which perceptual difficulty is the highest. Would iRT depend only on actual RT, then the angle effect should be captured by the increase of RT duration. Instead, the persistence of an effect of angle on iRT indicates that participants partially inferred the time they took to respond from their perception of task difficulty or something related. This last result is in line with the observation that iRTs are primarily based on experienced difficulty during a psychological refractory period paradigm (Bryce & Bratzke, 2014). Overall, these main effects of angle and orientation, surviving the effect of RT duration, suggest that iRT is more than the estimation of the time interval separating a stimulus and a response. In addition, the RT of error trials was underestimated compared to the same RT of correct trials, reflecting another type of bias.

The present results hence suggest that temporal metacognition is a mix between a direct readout and a mental construct (see Figure 1) and that participants do not only use temporal information to build a representation of self-produced RTs but also rely on other cues. Experiment 2 was designed and run in the laboratory to deepen our understanding of these effects of factors and objective RT.

Experiment 2

Experiment 2 was based on the same protocol as Experiment 1 but performed in the laboratory, allowing better experimental control. Several adjustments were introduced. First, to evaluate whether iRTs really reflect a (internal) time estimation, we added an externally defined duration estimation task to explore the correlation between the metacognitive evaluation of RT duration and general timing abilities. Second, to better assess the effect of task difficulty, we increased the number of stimulus angles to five (for each orientation) to get a more continuous variable. Third, we also manipulated the force level necessary for the response to be recorded. This manipulation is known to selectively affect motor execution time (Burle et al., 2002; Weindel et al., 2021) contrary to stimulus angle that (rather) specifically affects decision time.

Consequently, these two experimental manipulations permitted to test if iRT was sensitive to both manipulated factors. Besides decision stage, if motor execution is taken into account in iRT, we expect an effect of force on both RT and iRT. However, since experimental factors have been shown to potentially bias iRT, independently of actual RT duration, observing a force effect on iRT would be necessary but not sufficient to demonstrate a contribution of motor execution time. Consequently, we also recorded the EMG activity of the muscles involved in response execution and used it to fractionate RT into PMT (from stimulus onset to EMG onset) and MT (from EMG onset to recorded mechanical response; see Figure 5). We could thus dissociate the impact of decision and execution durations on iRT and evaluate whether taking into account the two subprocesses was better in predicting the estimation than considering only global RT.

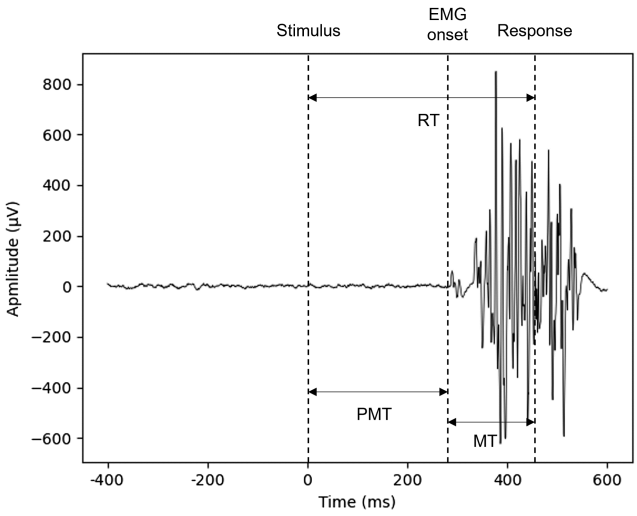
Method

Participants

We recruited 32 participants from the Aix-Marseille University community. All participants had normal or corrected-to-normal vision and no neurological disorders. The study was approved by the "Comité de Protection des Personnes Sud-Est VI" (No. 2021-A01548-33).

Figure 5

Experiment 2: RT Fractioning Into PMT and MT



Note. RT = reaction time; PMT = premotor time; MT = motor time.

Participants gave their written consent and were compensated at a rate of 15€ per hour. Two participants were excluded due to a high error rate (>50%). Finally, data of 30 participants (M_{age} = 25.07) were included in the analyses. Fourteen of them reported their gender as male and 16 as female. Eight were left-handed. Sample size was determined using G*Power (Faul et al., 2009) with effect sizes available from correlation analyses in the existing literature and Experiment 1 (r = 0.5). Based on this effect, an α of .05, and an 80% power, the required number of participants was 21. However, we assumed smaller effect sizes given RT fractioning, so the sample size was increased to 30 participants.[4]

Apparatus

Participants performed the experiment in a Faraday cage. They were comfortably seated in front of a computer screen placed at a 1-m distance, with a frame rate of 120 Hz. Responses were given by pressing either a left or right button with the corresponding thumb. Buttons were placed on two cylinders mounted on force sensors allowing to continuously measure the force produced with a sampling frequency of 2048 Hz. Force threshold needed for response recording was set by the experimenter thanks to this device. At button press, participants heard a 3-ms sound feedback at 400 Hz (resembling a small click). The left button was fixed in a vertical position by a magnet located at the bottom end of the cylinder (see Figure 6A). When the magnet was off, the button could be freely tilted to the right (toward the center of the device) up to 45°, which moved a cursor on the screen. This apparatus allowed participants to provide their subjective evaluation without having to remove the hand from the response button, which would have decreased the quality of EMG recordings (see below).

We measured the EMG activity of the *flexor pollicis brevis* of both hands with two electrodes placed on each thenar eminence. We used a Biosemi Active II system (BioSemi Instrumentation, Amsterdam, the Netherlands) for the recordings, with a sampling rate of 2048 Hz. During the whole experiment, the forearms and hypothenar muscles of the participants rested comfortably on the table to minimize tonic muscular activity, which would compromise the detection of voluntary EMG bursts. Furthermore, the EMG signal was monitored by the experimenter who asked the participant to relax his/her muscles if noise was detected. All participants performed one single 2-hr session including a short-duration estimation task of approximately 15 min and an RT estimation task with one training block and 12 experimental blocks, separated by self-paced breaks. The second task was similar to the online choice task, consisting in discriminating the orientation of a Gabor patch as fast and accurately as possible.

Stimuli

Stimuli presentation was controlled by the PsychoPy software, Version 2021.1.4 (Peirce et al., 2019). Each stimulus was a Gabor patch presented in the middle of the screen. It had a spatial frequency of 0.5 cycle/visual angle degree and a size of 15 visual angle degree. Its contrast was set to 0.5 on a scale from 0 (*uniform gray*) to 1 (*black and white*). Five stimulus angles were chosen: 1, 3, 5, 10, and 20° from vertical. These values were determined from a pilot study where they revealed a large range of performances: from 100% to 70% accuracy and optimal RT variations. Reports of durations and RTs were collected via a horizontally presented VAS. The scale was marked with vertical lines at 0 (left end), 0.33, 0.66, 1, and 1.33 s (right end). Only the 1-s tick was labeled as a cue for participants. We set the right end of the scale at a value superior to 1 s to avoid biasing estimates of longer durations.

Maximum Voluntary Force

Before the experimental tasks, maximum voluntary force was measured to adapt response thresholds to each participant. They were asked to alternate three left and three right presses. They were told that their presses should be brief and as strong as possible. The maximum force was determined for each press, and the highest maximum for each hand separately was retained. The maximum force of each participant was chosen as being the minimum of these two maximum values, to avoid tiring the weakest hand, and was used to determine the two response forces used (see below).

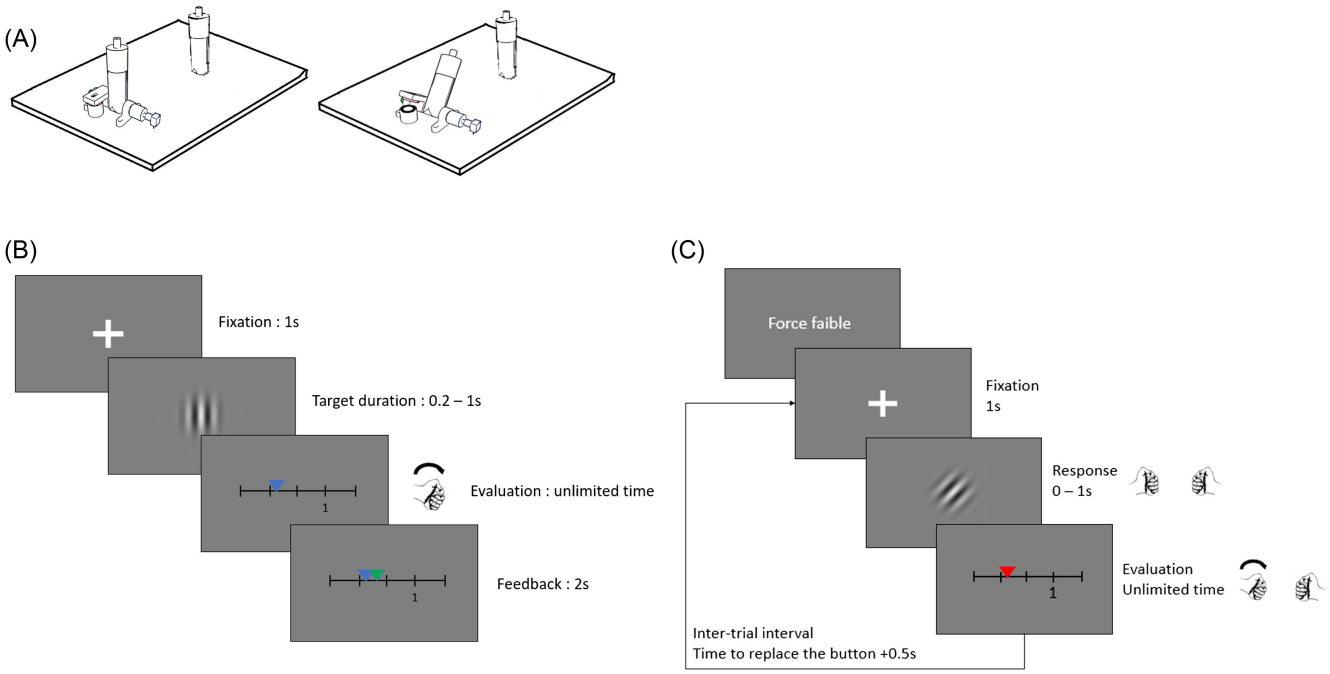
Externally Defined Duration Estimation Task

To assess participants' general timing abilities, they first performed an externally defined duration estimation task (see Figure 6B), as opposed to the RT estimation task where RT is a self-produced duration. Each trial began with the presentation of a fixation cross in the center of the screen. After 1 s, the stimulus appeared in place of the cross. During this task, all stimuli were vertical Gabor patches. Stimulus duration varied from trial to trial between 0 and 1 s. Possible durations were defined beforehand from a lognormal distribution (μ = 0.45 and SD = 0.3) to be similar to a RT distribution (right skewed) with a comparable range of durations to estimate between the tasks. Durations were randomized across trials. After each stimulus presentation, the VAS was presented on the screen with a blue triangle marker placed on its left end. Participants had to precisely evaluate the duration of the stimulus previously presented. When the

[4] Upon a reviewer's suggestion, we performed a power analysis using the R package simr, revealing a power of 99% for the RT coefficient of our mixed model.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Figure 6
Experiment 2: Experimental Display



Note. (A) Response device. Response buttons are mounted on force sensors. The left button is fixed by a magnet and can be released and tilted to the right. (B) Duration estimation protocol. (C) Reaction time evaluation protocol. The symbol "+" represents the fixation cross. See the online article for the color version of this figure.

scale appeared, the left button magnet was turned off so that the participant could tilt it to the right. This tilt allowed the visual marker to move on the scale. Participants were asked to place the marker at the position corresponding to their estimated duration of the stimulus and to press the left button to validate their estimation. The force threshold for this validation was set constant at 6 N (600 g). The marker stayed on the scale for 0.5 s followed by a feedback given as a green triangle marker placed at the position corresponding to the effective duration of the stimulus. The scale with both markers was kept on the screen for 2 s before the next trial. The transition to the next trial did not take place until the left button was replaced to its vertical position to be locked again. Every participant performed 80 trials.

Choice Reaction Time Task

Participants then performed the orientation discrimination task and had to estimate their RTs after each trial (see Figure 6C). Each trial began with the presentation of a fixation cross in the center of the screen. After 1 s, the stimulus replaced the cross. In this task, stimuli were clockwise or counterclockwise rotated Gabor patches displayed on the screen. Gabor patches were presented for a constant duration of 150 ms, to prevent the subjective evaluation to be based on the actual duration of the stimulus. Stimulus angle was fully randomized across trials. Participants had to respond to the stimulus orientation by pressing either the left (counterclockwise stimuli) or the right (clockwise stimuli) button. They were asked to respond as fast and as accurately as possible. The force level needed to provide a response could take two values, calculated from the maximum force of each participant. They constituted a weak force condition and a strong force condition, with force levels set at 5% and 20% of the maximum force, respectively (see Appendix E). Participants alternated three blocks of each force (to avoid the confounding effect of time on task, while reducing switches), and the starting force level was counterbalanced. At the beginning of each block, force level was indicated on the screen with the French instruction "Force faible" (weak force) or "Force forte" (strong force). Within each trial, when a response was given, the scale appeared with a red triangle marker placed on its left end. At the same moment, the left button was released, allowing the participant to tilt it to move the marker on the scale. In this task, participants had to precisely evaluate their RT, that is, the time between the apparition of the stimulus and the button press corresponding to the response. For that, participants were asked to place the marker at the position corresponding to their estimated RT and to press the left button to validate their estimation. The force threshold for this validation was set constant at 6 N (600 g). The scale remained on the screen for 0.5 s before the next trial. If no response to the Gabor patch was given within 1 s, the same visual scale was presented but the left button remained locked so that no interaction was possible. The scale was displayed for 1.25 s before the next trial. Intertrial interval was set to 0.5 s and did not begin until the left button was replaced in its vertical position to be locked again. Participants performed 12 blocks of 60 trials. Before the experimental blocks, participants performed 20 training trials to make sure that they understood the task properly. During this task, no feedback was ever provided.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

EMG Processing

EMG signal was processed with MYOnset (Spieser & Burle, 2025) a custom-made Python program to detect EMG onsets (released under an open-source license, available at https://github.com/lspieser/myonset). Briefly, the signal was first high pass filtered at 10 Hz, and then for each trial, the baseline mean (μ) and standard deviation (σ) of the signal were computed from 0.5 s preceding stimulus onset, for each hand separately. On the poststimulus signal, it was then evaluated whether, and when, EMG activity was significantly above a predetermined threshold ($\mu + n \times \sigma$) in either hand's channels (n was being optimized for each participant and force). When the signal was above this threshold, the precise burst onset was identified with an algorithm based on the "Integrated Profile" of the EMG burst (Liu & Liu, 2016). If the algorithm failed to locate or detect the EMG burst onset, the experimenter corrected or added them manually through the visual interface. Every muscular event (above-threshold change in the signal followed by a return to the baseline) in the trial was marked, even when the activation was not immediately followed by an overt response. If the trial was too noisy or if EMG activity started before the stimulus, the trial was not marked. In trials where a single EMG burst was detected, MT was defined as the time between the onset of EMG burst and the force threshold crossing recorded. PMT was defined as the time between stimulus onset and the EMG burst onset (see Figure 5).

Data Analysis

The first trial of each block was removed. As in the first experiment, MAD was computed on the evaluation times to exclude trials for which we can assume that participants were not doing the evaluation properly. Trials with an evaluation time inferior to the median −2 MAD and superior to the median +5 MAD were then excluded (<4% of trials). Trials with no EMG marking were excluded (<3%). Statistical analyses were performed by means of repeated-measures ANOVAs, correlations, and mixed models, using Python packages "pingouin" (Vallat, 2018) and "statsmodels" (Seabold & Perktold, 2010) and R package "lme4" (Bates et al., 2015). For ANOVAs, when the condition of sphericity was violated, the Greenhouse–Geisser correction was applied. Linear mixed models were first used for the same purpose as in Experiment 1. iRT was used as the dependent variable, with RT, angle, force, and orientation as fixed effects and participants as random effects, with both random intercepts and random slopes for each predictor. Predictors were chosen based on comparisons of the full model with models including a subset of predictors, with likelihood ratio tests. The full model always outperformed the reduced ones (see Appendix F for AIC of all tested models). The addition of interactions led to convergence issues and made the models overcomplex. In accordance with Experiment 1, model without interaction was judged appropriate enough to address our question of interest. Note that force and orientation were categorical factors with two modalities and that the β hence reflect the difference between both modalities with strong force and counterclockwise stimuli as references, respectively. In the present experiment, mixed models were also performed to study the contribution of PMT and MT to iRT. A global model including RT as a fixed effect was compared to a reduced model with only PMT as a fixed effect and to a full model with both PMT and MT as fixed effects (Models 1, 2, and 3 presented in the introduction). Models were compared based on their AIC, and a likelihood ratio test determined if they were different.

Results

Analyses of omission rate and accuracy are reported in Appendix G. Overall, overt errors were observed in 10% of trials and trials with multiple EMG activities were observed in 12% of trials. This type of trial has already been reported before (Servant et al., 2016; Weindel et al., 2021), but a clear theoretical interpretation of these multiple bursts is still lacking. Anyhow, they imply that the decision–execution sequence is not always respected. In a first step, only pure correct trials, that is, correct trials with a unique EMG burst, were analyzed. The number of the remaining trials per experimental condition is provided in Appendix H. The impact of those multiple bursts was analyzed in a second step.

Pure Correct Trials

Effect of Experimental Conditions on RT and iRT. The effect of the experimental conditions was studied using three-way ANOVAs with angle, orientation, and force as within-subjects variables. RTs were affected by the three factors. RTs increased as stimulus angle decreased, $F(4, 116) = 107.46$, $p < .001$, $\eta_p^2 = 0.79$, $\varepsilon = 0.35$, and were slower in the strong force condition, $F(1, 29) = 32.45$, $p < .001$, $\eta_p^2 = 0.53$, and for counterclockwise stimuli, $F(1, 29) = 22.92$, $p < .001$, $\eta_p^2 = 0.44$ (see Figure 7A). No first- nor second-order interactions were observed (all $p > .1$).

These effects on RT were underlied by effects on its subcomponents: PMT and MT. As expected, PMT was increased when stimulus angle decreased, $F(4, 116) = 94.68$, $p < .001$, $\eta_p^2 = 0.77$, $\varepsilon = 0.36$ (see Figure 7C). It was also smaller for clockwise stimuli, $F(1, 29) = 7.80$, $p = .009$, $\eta_p^2 = 0.21$. No effect of force, $F(1, 29) = 0.16$, $p = .69$, nor interaction was found (all $p > .1$). MT was longer in the strong force condition, $F(1, 29) = 114.29$, $p < .001$, $\eta_p^2 = 0.8$, and for counterclockwise stimuli, $F(1, 29) = 26.76$, $p < .001$, $\eta_p^2 = 0.5$ (see Figure 7D). MT was also increased when stimulus angle decreased, $F(4, 116) = 15.06$, $p < .001$, $\eta_p^2 = 0.34$, $\varepsilon = 0.41$. No interaction was found (all $p > .1$).

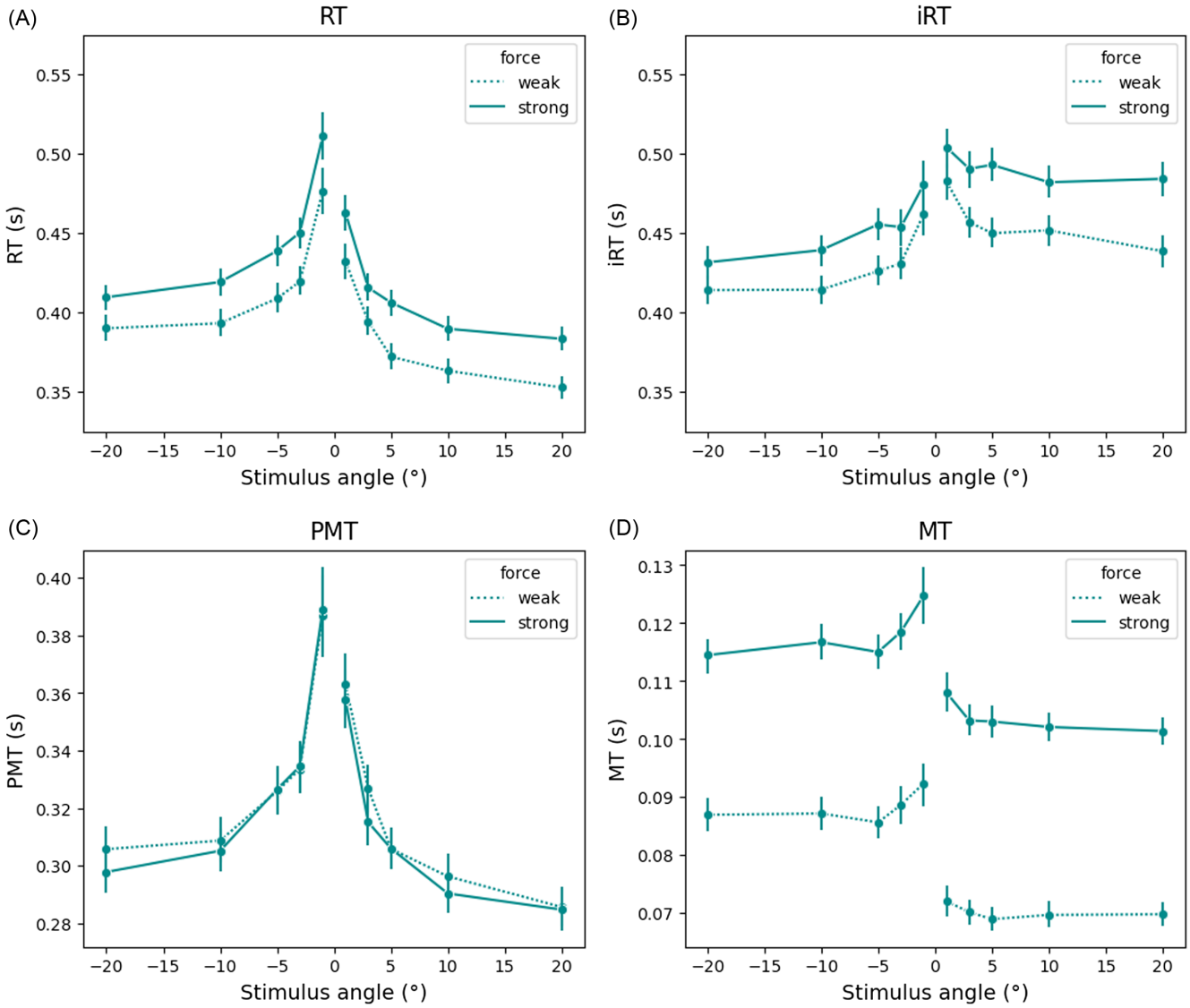
Same effects of angle, $F(4, 116) = 17.04$, $p < .001$, $\eta_p^2 = 0.37$, $\varepsilon = 0.38$, and force, $F(1, 29) = 9.96$, $p = .004$, $\eta_p^2 = 0.26$, as those observed on RTs were observed on iRTs. In contrast, unlike RTs, iRTs were higher for clockwise stimuli, $F(1, 29) = 11.78$, $p = .002$, $\eta_p^2 = 0.29$ (see Figure 7B). Angle and orientation interacted, $F(4, 116) = 3.70$, $p = .013$, $\eta_p^2 = 0.11$, $\varepsilon = 0.79$. No other first-order interaction was observed (all $p > .1$). A second-order interaction was also present, $F(4, 116) = 3.91$, $p = .014$, $\eta_p^2 = 0.12$, $\varepsilon = 0.70$.

Correlation Between RT and iRT. Mean RTs and iRTs showed a similar dependence on both force and stimulus angle. To assess RT and iRT relationship at the individual trial level, we performed correlation analyses. Repeated-measures correlation analysis revealed a positive correlation between trial-by-trial RTs and iRTs, $r_{rm}(15{,}806) = 0.39$, $p < .001$. Correlation coefficients were also computed for each participant using Pearson's correlation (see Figure 8). All participants but one presented a positive correlation between RTs and iRTs (range = −0.01–0.7, $M = 0.42$, $SD = 0.16$; see Appendix L for examples of individual data).

Link With External Timing Abilities. If participants really rely on time estimation, their capacity to correctly estimate their RT

Figure 7

Experiment 2: Mean (A) RT, (B) iRT, (C) PMT, and (D) MT as a Function of Stimulus Angle and Response Force



Note. Negative and positive angles represent counter- and clockwise orientations, respectively. RT = reaction time; iRT = introspective reaction time; PMT = premotor time; MT = motor time. See the online article for the color version of this figure.

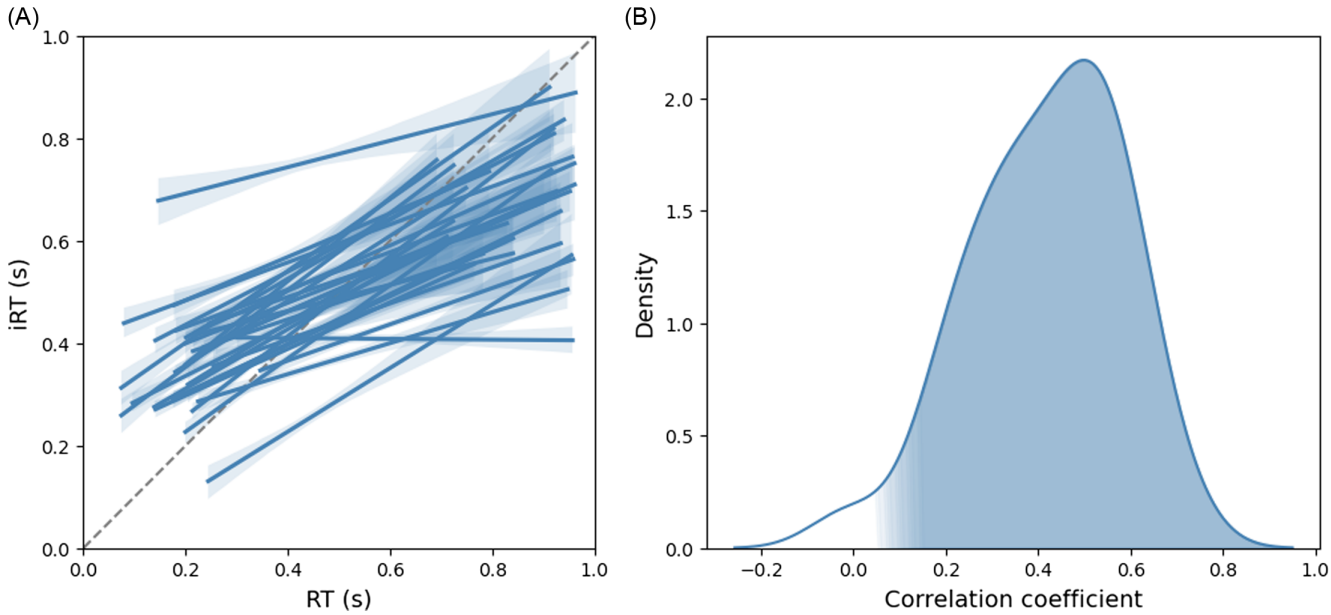
This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

should depend on their general time estimation ability. To investigate this link, we analyzed participants' performance in the time estimation task performed at the beginning of the experiment, and we then explored the relation between performance in RT estimation task and external duration estimation task. Correlation analyses were also performed for the duration estimation task, between the target duration, and the estimated one (as assessed by the position of the cursor on the scale). Repeated-measures correlation analysis revealed a positive correlation between target and estimated durations, $r_{rm}(2,369) = 0.72$, $p < 001$. Individual Pearson's correlations showed that all subjects presented a positive correlation between target and estimated durations (range = 0.27–0.9, $M = 0.73$, $SD = 0.11$; see Figure 9). Not surprisingly, since feedback was given in this task, correlation coefficients were larger than for the RT evaluation task, $t(29) = -10.97$, $p < .001$.

To investigate the link between RT evaluation and duration estimation abilities, we computed a simple linear regression to predict RT evaluation task correlation coefficients based on duration estimation task correlation coefficients (see Figure 10). Outlier detection using the Local Outlier Factor (Breunig et al., 2000) brought out one outlier value within each task (easily detectable on Figure 10). Regression was thus performed on $n = 28$ observations. Result of this analysis indicated that there was a significant association between the two variables ($R^2 = 0.2$) with r of the duration estimation task being a significant predictor of r of the RT evaluation task ($\beta = 0.92$, $t = 2.61$, $p = .015$).

Figure 8

Experiment 2: Individual Correlations Between RT and iRT



Note. (A) Individual trial-by-trial correlations between RT and iRT. Each regression line represents a participant. The gray dashed line is the $x = y$ line. (B) Distribution of individual correlation coefficients (r). Blue (gray) zone represents positive correlations, and white zone represents null correlations. The zone of uncertainty concerning correlation significance is represented by a linear gradient around critical r values for an α of .05 and minimal and maximal numbers of observations (n = 408 and n = 648). RT = reaction time; iRT = introspective reaction time. See the online article for the color version of this figure.

Dissociating RT and Experimental Factor Contributions to iRT. As in Experiment 1, we used mixed models with iRT as the dependent variable and RT and the experimental factors as the regressors to assess if nontemporal information could also participate to RT evaluation. This model showed the main effects of RT (β = 0.48, t = 14.49, $p < .001$), force ($\beta = -0.018$, $t = -1.99$, $p = .047$), and orientation (β = 0.043, t = 6.78, $p < .001$), but no effect of stimulus angle ($\beta < 0.001$, $t = 0.021$, $p = .98$). It means that for the same RT, iRTs were higher in the strong force condition and for clockwise stimuli. Multiple regression performed on means showed similar results with the main effects of RT (β = 0.44, t = 6.47, $p <$ .001), force ($\beta = -0.02$, $t = -4.99$, $p < .001$; see Figure 11A), and orientation (β = 0.04, t = 13.47, $p < .001$; see Figure 11B), but no effect of stimulus angle ($\beta < -0.001$, $t = -0.37$, $p = .72$). All results are confirmed by the mediation analyses provided in Appendix I.

Impact of PMT and MT on iRT. To assess to what extent premotor and motor components of RT contribute to iRT, we compared three models (Models 1, 2, and 3 exposed in the introduction). In a first model (Model 1), we used global RT as a predictor of iRT. In a second model (Model 2), we entered only PMT as a predictor, to assess whether the motor component was used to report iRT or not. In a third model (see Model 3), we entered both PMT and MT as predictors (see introduction for the difference between Models 1 and 3). In all three models, force and orientation were also included as predictors, as they had significant effect on iRT in the previous analysis. Participants were included as random effect, and we used both random intercept and slopes for each factor. Model 3 provided a better fit than Model 2, $\chi^2(6) = 482.68$, $p < .001$ (see Table 2). Both PMT ($\beta = 0.47$, $t = 13.88$, $p < .001$) and MT (β = 0.57, t = 7.68, $p < .001$) predicted iRT, indicating that MT is likely taken into account in iRT. In a last step, we compared Model 3 and Model 1 to assess whether PMT and MT have the same weight in iRT (see the introduction for the rationale). Model 3 (PMT + MT, different weights) better accounts for the data, $\chi^2(6) =$ 124.93, $p < .001$, than Model 1 (including only RT), indicating that the weights of PMT and MT are different, independently of their respective durations. Even if AIC takes into account the complexity of the model (Akaike, 1974), it could be that the increased flexibility allowed by splitting RT in two subintervals always leads to a better fit. To evaluate this possibility, we performed pseudorandom splits of RT (N = 1,000) and compared them to Model 1 and Model 3. Details and results are reported in Appendix J: Briefly, Model 3 largely outperformed pseudorandom split.

Beyond Pure Correct Trials

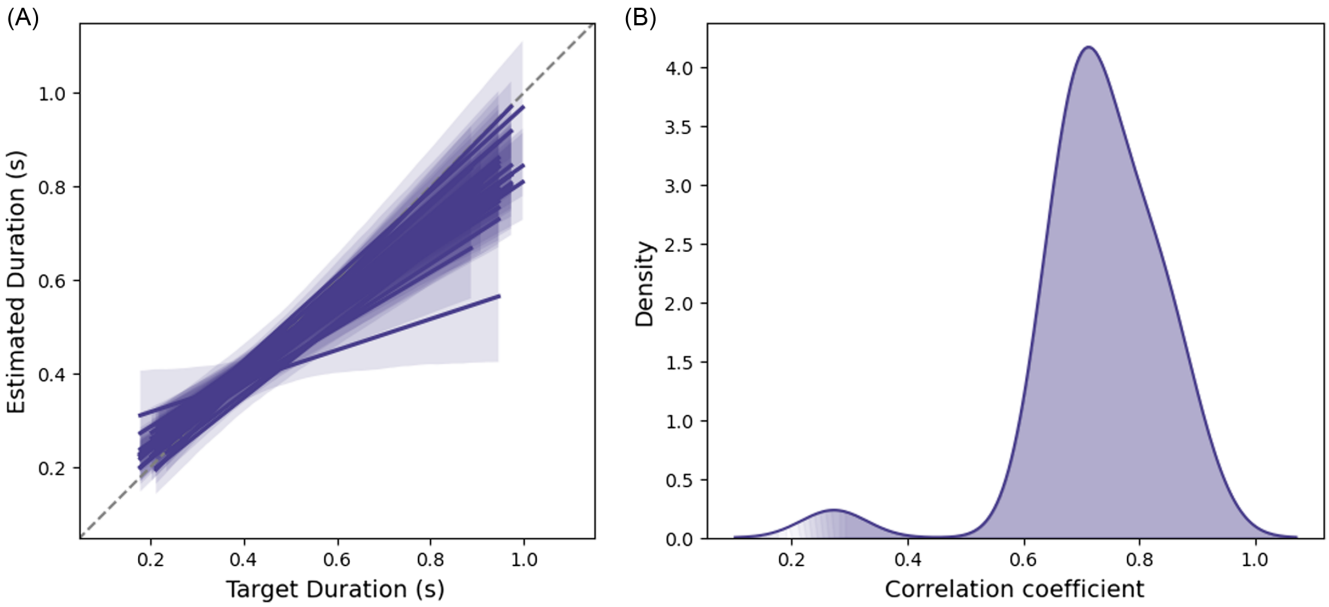
All previous analyses were made on pure correct trials, namely correct trials with only one EMG burst. Hence, two types of trials were left aside: erroneous responses and nonpure correct trials. Indeed, in some trials, some EMG activations are observed before the one leading to the overt response, revealing some forms of "hesitation." The effect of experimental conditions on the number of EMG activations follows the effect observed on accuracy and is depicted in Appendix K. As both overt errors and motor hesitation related to multiple EMG activations could impact the way participants estimate their RTs, we ran some further analyses including all trials.

Impact of Accuracy on iRT. To see if the RTs of trials containing multiple bursts were evaluated differently, we ran a mixed model on all correct trials with iRT as the dependent variable

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Figure 9

Experiment 2: Individual Correlations Between Target and Estimated Durations

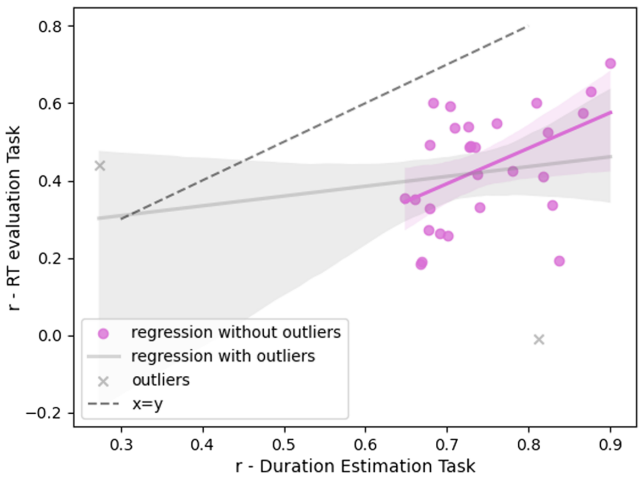


Note. (A) Individual trial-by-trial correlations between target and estimated durations. Each regression line represents a participant. The gray dashed line is the $x = y$ line. (B) Distribution of individual correlation coefficients (r). Purple (gray) zone represents positive correlations, and white zone represents null correlations (considering an $\alpha = .05$). The zone of uncertainty concerning correlation significance is represented by a linear gradient around critical r value for an α of .05 and number of observations ($n = 80$). See the online article for the color version of this figure.

and RT and the number of EMG bursts as the predictors, in addition to response force, as it has an effect on the number of bursts and was a significant predictor of iRT. Participants were included as random effect, and we used both random intercept and slopes for each predictor. In addition to the main effects of RT ($\beta = 0.48$, $t = 13.64$, $p < .001$) and force ($\beta = -0.018$, $t = -1.96$, $p = .05$), an effect of the number of bursts ($\beta = 0.019$, $t = 2.79$, $p = .005$) was observed, with iRT being higher for trials with multiple EMG activations, revealing that for the same objective RT, trials with multiple bursts are judged longer. Finally, we studied the effect of the overt errors on iRT, by using another mixed model with iRT as the dependent variable and RT and response accuracy (correct vs. errors) as the predictors. In addition to the main effect of RT ($\beta = 0.52$, $t = 17.03$, $p < .001$), an effect of accuracy ($\beta = 0.034$, $t = 5.46$, $p < .001$) was observed, with iRT being smaller for error trials, even though no difference in RT was observed between correct and error trials, $t(29) = -0.7$, $p = .49$, replicating results obtained in Experiment 1.

Figure 10

Experiment 2: Linear Regression Between Individual Duration Estimation Task and RT Evaluation Task Correlation Coefficients



Note. Each dot is a subject, and the crosses are outliers. Regressions are presented with (in light gray) and without (in dark gray) the two outliers (respectively in gray and pink in the colored version). The dashed line is the $x = y$ line. RT = reaction time. See the online article for the color version of this figure.

Discussion

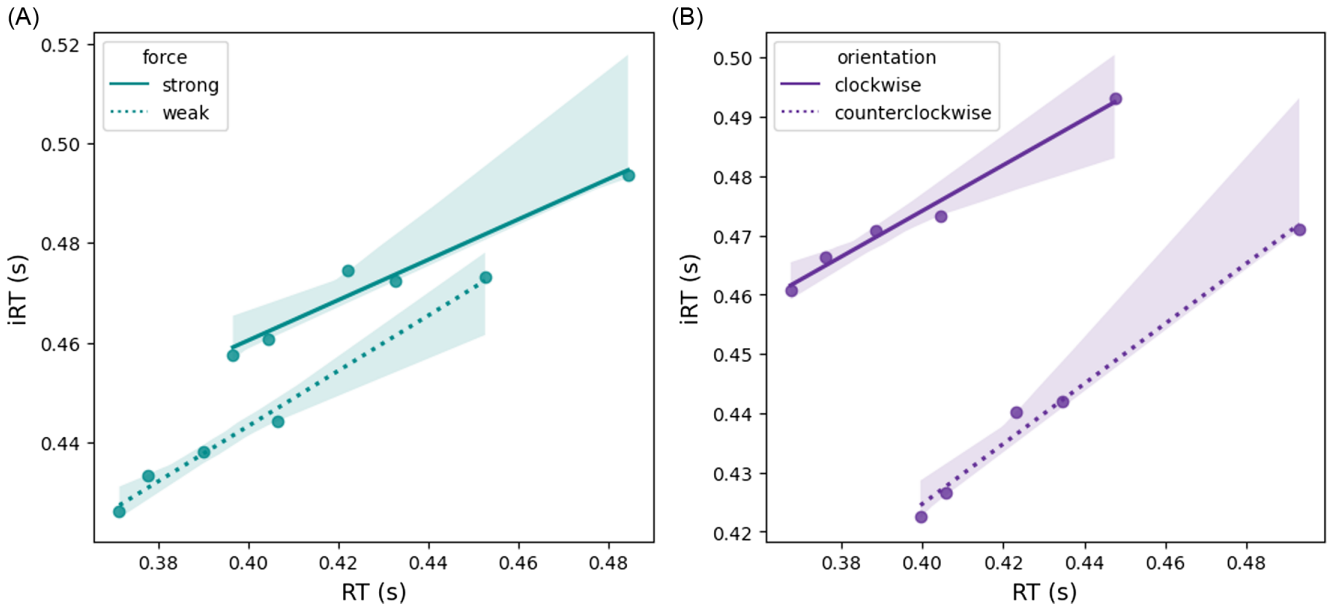
The correlations between RT and iRT confirm and extend the observations of Experiment 1 and of the literature (Bratzke & Bryce, 2019; Corallo et al., 2008). In the more controlled environment of the laboratory, most participants presented a positive correlation between iRT and RT at the trial level, indicating that they can accurately evaluate the duration of their RT.

As in Experiment 1, RT and iRT were affected similarly by stimulus angle but showed opposite modulations by orientation: RT was lower while iRT was higher for clockwise stimuli. The fact that this effect is replicated despite different devices used to provide the evaluation tends to exclude artifacts linked to the way this evaluation was provided. In Experiment 1, we speculated that this asymmetry might be due to a bias toward the right part of the evaluation scale induced by the clockwise stimulus or the right response. The present

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Figure 11

Experiment 2: Linear Regression Between Mean iRTs and RTs for (A) the Two Force Levels and (B) the Two Orientations



Note. RT = reaction time; iRT = introspective reaction time. See the online article for the color version of this figure.

data do not provide any new argument for or against such an explanation, which certainly requires more investigation. Anyways, whatever the reason for this asymmetry, it reveals a systematic deviation between subjective evaluation and actual RT.

Experiment 2 introduced a new manipulation, namely response force. As expected, force manipulation specifically affected the motor execution stage, as it increased MT but left PMT unchanged. Increasing response force lengthened iRTs, which can at first sight suggest that motor execution time is taken into account in iRT. However, analyzing iRT predictors in the mixed models, an effect of force was observed that goes beyond its effect on RT. This indicates that, for the same effective RT duration, participants estimated themselves as being slower in the strong force condition. As it will be discussed below, this means that an effect of force on iRT is necessary but not sufficient to conclude on the consideration of MT.

On the other hand, and contrary to Experiment 1, the effect of angle on iRT did not survive the introduction of RT in the mixed model, in addition of RT and force effects, indicating that all the effect of angle on iRT was accounted for by its effect on RT. Although this may sound surprising and appear as a replication failure of Experiment 1, we will argue in the general discussion that this discrepancy might reveal interesting properties of metacognitive evaluation.

In a further step, action monitoring and error detection being key features of metacognitive evaluation, we explored how (partial) errors are subjectively evaluated. We looked at whether overt incorrect responses were evaluated differently than correct ones. Replicating Experiment 1, we observed that RT was underestimated for error trials compared to correct ones. Besides accuracy classification in a binary way (correct or incorrect) based on overt behavior, analyses of the EMG activity showed that overt correct responses can be preceded by one or several early subthreshold bursts. In conflict task, subliminal activations recorded on the hand associated with the incorrect response (Coles et al., 1985), called "partial errors" (Burle et al., 2002), index automatic incorrect response activation. In nonconflicting settings, subliminal EMG activation can also occur on the correct hand, likely revealing processing hesitations (Questienne et al., 2018; Weindel et al., 2021; see Dendauw et al., 2024, for a possible theoretical account). The presence of multiple EMG bursts increased iRT, besides their lengthening effect on RT, as these trials were judged as longer than pure correct trials with the same actual RT. This lengthening is surprising, given that most of these partial EMG

Table 2

Experiment 2: Estimates for the Fixed Effects of the Three Mixed Models Predicting iRT, and Corresponding AIC

Model	Intercept	RT	PMT	MT	Force	Orientation	AIC
RT	0.255***	0.48***			−0.018*	0.043***	−28,279
PMT + MT	0.242***		0.475***	0.57***	−0.016*	0.046***	−28,392
PMT	0.314***		0.462***		−0.033***	0.037***	−27,921

Note. iRT = introspective reaction time; AIC = Akaike information criterion; RT = reaction time; PMT = premotor time; MT = motor time.
* $p < .1$. *** $p < .001$.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

activations likely remain unconscious (Ficarella et al., 2019; Rochet et al., 2014). The possible nature of the subjective (largely unconscious) feelings biasing iRT will be discussed in the general discussion.

Overall, confirming Experiment 1, Experiment 2 shows that although systematic biases do occur, the best predictor of iRT is actual RT. Furthermore, the accuracy of iRT estimation seems to depend on general timing abilities (see, however, Bratzke & Bryce, 2022; Klein & Stolz, 2018). Besides linking iRT to actual RT, the second experiment also went a step further in assessing whether, and which, subprocesses included in RT are also accessible. Fractioning RT based on EMG activity allowed to dissociate the duration of decision-related and motor execution-related subcomponents. Mixed model analysis showed that both decision and execution subprocesses are considered to estimate RT, with different weights though. The implications of all those results will be discussed in the General Discussion section.

Transparency and Openness

In accordance with the Transparency and Openness Promotion Guidelines, all data, software code, and other methods developed by others are appropriately acknowledged. Data and codes used in this article are available on the Open Science Framework at https://osf.io/3rzak/. The research was not preregistered. We used open-source software for all data analyses, ensuring that others can reproduce our findings without restrictions.

General Discussion

Acting in time is essential for adapted behavior. While the processes underlying time perception have been studied for long, it is only very recently that scholars got interested in our capacity to evaluate the timing of our actions. Following previous research (Corallo et al., 2008; Sanford, 1970), the goal of the present study was to deepen our understanding of temporal metacognition, by investigating whether participants can evaluate the duration of their cognitive processes in a choice task. We sought to disentangle temporal and nontemporal information contributing to this subjective evaluation. Two experiments, one online and one in the laboratory, were run using quantified introspection to measure both actual RT and its introspective equivalent, iRT. Replicating previous observations (e.g., Corallo et al., 2008), mean RT and mean iRT were sensitive in a comparable way to experimental manipulations, confirming that participants have a rather good representation of the time needed to respond to a stimulus. Importantly, and as exposed in the introduction, observing a positive correlation on the means is not sufficient to conclude that iRT is a direct readout of RT temporal content as it may simply reflect a mere average estimation of the sensitivity of RT to experimental conditions. The use of linear mixed models on individual trials allowed us to decipher the information at the basis of iRT. First, in all mixed models of both experiments, individual RTs were the main predictor of individual iRTs, establishing a strong link between the two. Despite some interindividual variability, this effect was consistent, with a large majority of participants showing a substantial link and only a few being unable to estimate their RT. This indicates that participants, at least partly, estimated the effective temporal interval constituting RT. In addition, Experiment 2 showed that participants who were the most accurate in estimating external time intervals were also better at estimating their own RT, suggesting a shared mechanism of time estimation in both tasks. Overall, these results provide strong arguments for the idea that iRT really reflects a time estimation process (see, however, Bratzke & Bryce, 2022; Klein & Stolz, 2018, in a dual-task context).

Of Which Processes Are the Durations Evaluated: Contribution of Decision and Motor Execution

Given that participants were able to estimate the effective durations of their RT, we sought to determine if, and to what extent, the different processes at stake during RT were considered in iRT. We here focused on decision- and motor execution-related processes.

In both experiments, we manipulated the angle of the Gabor patches, affecting the decision process. iRT was affected in a way similar to RT by this experimental factor, suggesting that the duration of the decision process is taken into account for the subjective evaluation. In Experiment 2, we additionally manipulated response force, known to (rather selectively) affect motor execution processes (Burle et al., 2002; Weindel et al., 2021). iRT was also sensitive to this factor, supporting the idea that the duration of the motor processes is also incorporated within iRT. However, if the sensitivity of iRT to these manipulations is *necessary* to conclude that the targeted processes are evaluated, it is not *sufficient*. Indeed, as will be discussed below, the mixed model analysis revealed that this sensitivity can reflect, at least partly, a bias linked to the experimental conditions.

To better establish the respective roles of decision and motor processes, EMG fractioning was used to get objective measures of the two processes' durations. Mixed models comparison revealed that PMT and MT better predicted iRT than PMT only, suggesting that participants did take into account both subdurations in their RT evaluation.

Previous research studied the implication of motor processes in metacognition (Fleming et al., 2015; Gajdos et al., 2019; Jovanovic et al., 2021; Mamassian, 2008), but mainly on confidence. Jovanovic et al. (2021) also investigated the accessibility of the duration of motor execution in a synchronization/pointing task. Their data suggest that participants can have access to the duration of their movement time, in agreement with the present data. However, as reported by the authors, their task implied large movements, and how such large movements are timed is a matter of debate (see Leib et al., 2017). Conversely, RT tasks often involve isometric presses (which is the case in the present studies) without any actual movement (or very limited ones), eliminating the dynamical aspect of movement time.

Furthermore, and interestingly, the mixed model with both PMT and MT (whose sum is exactly equal to RT) as predictors outperformed the model with only RT. This is so because the PMT + MT model allows to have different weights for PMT and MT, which improved the prediction. The fact that these two components have different weights rules out the possibility that RT estimation simply relies on the perception of the duration separating two external events (from stimulus onset to response). Indeed, would participants base their estimation on this external duration, fractioning RT based on EMG should represent a random split of RT and hence should not improve iRT prediction. The most likely conclusion is hence that participants really made a metacognitive judgment on the time they took to decide, and on the time required to produce the response, and did not estimate external durations covarying with it. Another important consequence of this result is that iRT is very likely based

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

on the sum of the estimations of the two intervals, an iPMT and an iMT.

Such access to PMT duration is at odds with previous reports suggesting that participants have poor evaluation of their decision time (Miller et al., 2010) in a Libet's paradigm where participants have to estimate the time of their intention to act on a rotating clock (Libet et al., 1983). Such a conclusion, however, relies on very strong assumptions concerning the nondecision time, and the lack of an objective decision time measure prevents a precise determination of the judgments' accuracy. On the other hand, although some studies were interested in the estimation of movement time (De Kock et al., 2021, 2023), no studies have directly asked participants to estimate the duration of their motor execution time in nontiming tasks, which must be inferred from indirect measures without the use of EMG. Hence, to confirm and further explore the idea that participants have distinct representations of the times they took to decide and act, both objective and subjective measures of these processes' durations are needed.

iRT Inference From Multiple Cues

Although iRT is largely based on the duration of the underlying processes, other, nontemporal, information is also impacting it. This is in accordance with previous work demonstrating the influence of nontemporal cues on iRT, such as perceived difficulty in dual task (Bryce & Bratzke, 2014) or cue-stimulus and response-stimulus intervals in task switching (Bratzke & Bryce, 2019). It was not guaranteed that a similar observation would be made in a single task context, where the attentional demand is lower. However, in a similar way, in both of the present experiments, mixed models revealed the effects of experimental factors on iRT, which go beyond their mere effect on RT. Those factors broke the linearity between RT and iRT, leading to systematic under- or overestimations. More precisely, such shifts in the RT/iRT relationship reveal that, for the same objective RT, participants judged themselves as slower under some experimental conditions. In Experiment 1, angle (i.e., perceptual difficulty) induced such a bias, with RTs in the most difficult condition being perceived as longer than their equivalent in the easiest one (the fact that this effect disappeared in Experiment 2 will be discussed below). A similar effect of force (response difficulty) was observed in Experiment 2. Two possible theoretical explanations, not necessarily incompatible, can account for this effect.

A first possibility is that the manipulated factors might have given rise to some sort of subjective feelings that could bias the RT duration evaluation. For example, the feeling of difficulty (Bryce & Bratzke, 2014) or the effort put in realizing the current trial, either at the perceptual or the motor level (see De Kock et al., 2021, for an impact of motor difficulty on time estimation), could be used as an index of RT, as the two might be strongly related. Note that such an explanation could easily account for the effect of EMG bursts: The presence of multiple bursts could lead to a feeling of difficulty/effort/hesitation (Morsella et al., 2009; Questienne et al., 2018), leading to infer that RT was longer than it actually was. Whether it is part of an explicit strategy, that is, "It was difficult/effortful, I must have been slow," remains an open question.

Such an account opens the question as to whether both force and stimulus angle manipulations lead to a common subjective feeling (e.g., a common sense of "difficulty" would be shared between perceptual and motor processes) or whether they give rise to two different types of feelings that combine to give rise to the subjective duration of RT. A reverse observation has been made in studies where RT duration influences other subjective perceptions, such as difficulty (Desender et al., 2017), or the experience of urge-to-err (Questienne et al., 2018). As an example, Desender et al. (2017) concluded that the experience of difficulty relies on multiple cues as they observed that RT, among other factors, influenced this subjective perception. As it stands, whether the feeling of "difficulty/effort" determines iRT or the other way around remains a chicken-and-egg problem, which will certainly necessitate extremely well-designed protocols to distinguish which is the cause of the other. Anyways, in such an account, participants could have a global metacognitive experience emerging from information about multiple cues. In that sense, iRT would rather be a mental construct affected by multiple internal representations (including timing) rather than a simple interval timing of the duration of the underlying processes.

Such an account, however, is challenged by some of the current results. First, error iRTs are underestimated. This would indicate that errors are perceived as less difficult/effortful than correct trials. Although possible, this sounds like a post hoc explanation. The disappearance, or at least the large reduction, of the angle effect in Experiment 2 is also a challenge, as there is no reason why the feeling of difficulty would have diminished while the effect on RT was very clear.

An alternative interpretation could be that the biases in RT evaluation derive from a priori assumptions about the effect of the manipulated factors. For example, participants might systematically associate the stronger force level, or the lower stimulus angle, with longer RT, irrespective of their actual performance, leading to the categorical effects previously described. For such a bias to occur, two conditions are necessary. First, one needs a prior knowledge of the expected effect of each manipulation on RT. Manipulating such knowledge could allow to control the presence or disappearance of such categorical effect on iRT. Second, participants need to identify which experimental condition they are currently running. Indeed, and despite an obvious association between perceptual difficulty and RT, the categorical effect of stimulus angle was only present in Experiment 1 and disappeared in Experiment 2. The major difference between the two experiments, as far as perceptual difficulty is concerned, is the number of possible angles: While only two were present in Experiment 1, and very easily categorizable (1° = hard vs. 7° = easy), they were more numerous (5) in Experiment 2 and some of them were much closer (1°, 3°, and 5°). Hence, while it was very easy on every trial to grasp the experimental condition the participant was currently running in Experiment 1, it was arguably much more difficult in Experiment 2. When facing such a complexity, it might become less advantageous (e.g., consumes too much resources; see below) to rely on the experimental task categories to (partly) infer the RT.

Note that the force manipulation of Experiment 2, being part of the instruction given to the participants before each block, was very easy to identify, making it a reliable category to infer the RT. Delimiting the conditions leading participants to use or not such categorical information would provide essential information on how metacognitive experience is built.

Interestingly, the categorical effects observed on iRT are not restricted to objective experimental factors. The fact that participants also underestimated their RT for error trials could come from a prior belief as well. They may in fact have a more or less explicit

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

knowledge of the speed–accuracy trade-off, that is, knowing that responding hastily increases the risk of an error. Reversing the inference implies that error trials are likely faster than correct ones. The difference lies in the fact that accuracy is not an information given to the participant and hence has to be consciously accessed (Rabbitt, 1966). Having detected they made an error, participants might have inferred that it was because they responded too fast, leading to an underestimation of their RT. Conversely, the effect of multiple EMG bursts seems difficult to reconcile with such a priori assumption hypothesis, as participants are probably unaware of these partial activities (Rochet et al., 2014).

Why Use Nontemporal Cues?

If participants do have access to the time they took to respond, why do they base their RT estimation on other cues? Resource saving might be the key. Indeed, time estimation is a resource-consuming process (Brown, 1997; Burle & Casini, 2001; Macar et al., 1994), and when judging their own RTs, participants are, obviously, allocating resources to respond to the stimulus. Consequently, if any easier and less consuming way of estimating RT exists, it is very efficient to use it (Klein & Stolz, 2018). Accordingly, a salient subjective feeling or easily categorizable experimental condition could constitute an information more effortlessly accessible than objective RT, usable as a good proxy, as they are related to its duration. Besides, each source of information is likely noisy, and not very reliable, but the origins of variations are probably rather independent. As a consequence, combining them might decrease the actual level of noise, hence improving the estimation.

It remains challenging to dissociate the different information considered because we cannot verify what participants are introspecting at the time of the subjective judgment, but it is likely that the relative contribution and weight of each source are not fixed and vary according to different factors. As argued above, the disappearance of the effect of stimulus angle in Experiment 2 is probably due to the increase in the number of angles creating a higher demand to discriminate them and making evidence of difficulty less beneficial. This is in agreement with the fact that the context and demand of the task could determine what information is more prone to be used (Reyes & Sackur, 2014). Moreover, individual abilities and training (Desender et al., 2017) may also be a modulating factor, as, although speculative, different participants could base their judgments on different clues depending on their sensitivity to each of them. More research will be needed to support this possibility.

Conclusion

Altogether, the results showed that the subjective evaluation of RT emerges from a combination of temporal and nontemporal sources. iRT appeared to be partly inferred from prior knowledge and other subjective feelings linked to the task (which could be respectively assimilated to "metacognition knowledge" and "metacognition experience," according to Flavell's, 1979, terminology). This observation supports the idea that internally defined durations can be approximated from nontemporal metacognitive cues (Klein & Stolz, 2018) and is, to some extent, in agreement with the suggestion that mental contents are inferred from a priori causal theories (Nisbett & Wilson, 1977). Besides these inferences, participants could still directly readout RT temporal content and were found to have access to the duration of the processes involved, namely decision and motor execution.

Constraints on Generality

Applying a new data processing rationale on self-produced time allowed to reveal that participants can reliably access the duration of their underlying processing operations but that this estimation can be biased by external factors and a priori knowledge. While we do believe that these conclusions also hold for dimensions other than time, the proposed methodology might not be applicable for all other dimensions, as it requires that both the first- and second-order judgments are continuous and can be regressed one onto the other. One hence cannot be sure that the present conclusions generalize to noncontinuous dimensions. Another potential generality constraint relates to the population used in the second experiment, which is mainly psychology and neuroscience students. However, the first experiment was run online and hence very likely allowed to recruit a much more heterogeneous sample. Since the results of the two experiments are largely coherent, we have good reasons to assume that the results of the second experiment can also be generalized to a larger part of the adult population, as the ones of Experiment 1.

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*(6), 716–723. https://doi.org/10.1007/978-1-4612-1694-0_16

Akdoğan, B., & Balcı, F. (2017). Are you early or late?: Temporal error monitoring. *Journal of Experimental Psychology: General*, *146*(3), 347–361. https://doi.org/10.1037/xge0000265

Bakdash, J. Z., & Marusich, L. R. (2017). Repeated measures correlation. *Frontiers in Psychology*, *8*, Article 456. https://doi.org/10.3389/fpsyg.2017.00456

Balci, F., Simen, P., Niyogi, R., Saxe, A., Hughes, J. A., Holmes, P., & Cohen, J. D. (2011). Acquisition of decision making criteria: Reward rate ultimately beats accuracy. *Attention, Perception & Psychophysics*, *73*(2), 640–657. https://doi.org/10.3758/s13414-010-0049-7

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1–48. https://doi.org/10.18637/jss.v067.i01

Botwinick, J., & Thompson, L. W. (1966). Premotor and motor components of reaction time. *Journal of Experimental Psychology*, *71*(1), 9–15. https://doi.org/10.1037/h0022634

Bratzke, D., & Bryce, D. (2016). Temporal discrimination of one's own reaction times in dual-task performance: Context effects and methodological constraints. *Attention, Perception, & Psychophysics*, *78*(6), 1806–1816. https://doi.org/10.3758/s13414-016-1161-0

Bratzke, D., & Bryce, D. (2019). Introspection is not always blind to the costs of multitasking: The case of task switching. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *45*(6), 980–992. https://doi.org/10.1037/xlm0000635

Bratzke, D., & Bryce, D. (2022). Timing of internal processes: Investigating introspection about the costs of task switching and memory search. *Attention, Perception & Psychophysics*, *84*(5), 1501–1508. https://doi.org/10.3758/s13414-022-02510-6

Bratzke, D., Bryce, D., & Seifried-Dübon, T. (2014). Distorted subjective reports of stimulus onsets under dual-task conditions: Delayed conscious perception or estimation bias? *Consciousness and Cognition*, *30*, 36–47. https://doi.org/10.1016/j.concog.2014.07.016

Bratzke, D., & Hansen, A. (2024). How does it feel? Passage of time judgments in speeded RT performance. *Psychological Research*, *88*(1), 141–147. https://doi.org/10.1007/s00426-023-01854-4

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Breunig, M. M., Kriegel, H.-P., Ng, R. T., & Sander, J. (2000). LOF: Identifying density-based local outliers. *Proceedings of the 2000 ACM SIGMOD international conference on Management of data* (Vol. 29, pp. 93–104). Association for Computing Machinery. https://doi.org/10.1145/335191.335388

Brown, S. W. (1997). Attentional resources in timing: Interference effects in concurrent temporal and nontemporal working memory tasks. *Perception & Psychophysics*, *59*(7), 1118–1140. https://doi.org/10.3758/BF03205526

Bryce, D., & Bratzke, D. (2014). Introspective reports of reaction times in dual-tasks reflect experienced difficulty rather than timing of cognitive processes. *Consciousness and Cognition*, *27*, 254–267. https://doi.org/10.1016/j.concog.2014.05.011

Bryce, D., & Bratzke, D. (2015). Are introspective reaction times affected by the method of time estimation? A comparison of visual analogue scales and reproduction. *Attention, Perception, & Psychophysics*, *77*(3), 978–984. https://doi.org/10.3758/s13414-014-0804-2

Bryce, D., & Bratzke, D. (2017). Are participants' reports of their own reaction times reliable? Re-examining introspective limitations in active and passive dual-task paradigms. *Acta Psychologica*, *172*, 1–9. https://doi.org/10.1016/j.actpsy.2016.10.007

Burle, B., & Casini, L. (2001). Dissociation between activation and attention effects in time estimation: Implications for internal clock models. *Journal of Experimental Psychology: Human Perception and Performance*, *27*(1), 195–205. https://doi.org/10.1037//0096-1523.27.1.195

Burle, B., Possamaï, C.-A., Vidal, F., Bonnet, M., & Hasbroucq, T. (2002). Executive control in the Simon effect: An electromyographic and distributional analysis. *Psychological Research*, *66*(4), 324–336. https://doi.org/10.1007/s00426-002-0105-6

Coles, M. G. H., Gratton, G., Bashore, T. R., Eriksen, C. W., & Donchin, E. (1985). A psychophysiological investigation of the continuous flow model of human information processing. *Journal of Experimental Psychology: Human Perception and Performance*, *11*(5), 529–553. https://doi.org/10.1037/0096-1523.11.5.529

Corallo, G., Sackur, J., Dehaene, S., & Sigman, M. (2008). Limits on introspection: Distorted subjective time during the dual-task bottleneck. *Psychological Science*, *19*(11), 1110–1117. https://doi.org/10.1111/j.1467-9280.2008.02211.x

De Kock, R., Zhou, W., Datta, P., Joiner, M. W., & Wiener, M. (2023). The role of consciously timed movements in shaping and improving auditory timing. *Proceedings of the Royal Society B: Biological Sciences*, *290*(1992), Article 20222060. https://doi.org/10.1098/rspb.2022.2060

De Kock, R., Zhou, W., Joiner, W. M., & Wiener, M. (2021). Slowing the body slows down time perception. *eLife*, *10*, Article e63607. https://doi.org/10.7554/eLife.63607

Dendauw, E., Evans, N. J., Logan, G. D., Haffen, E., Bennabi, D., Gajdos, T., & Servant, M. (2024). The gated cascade diffusion model: An integrated theory of decision making, motor preparation, and motor execution. *Psychological Review*, *131*(4), 825–857. https://doi.org/10.1037/rev0000464

Desender, K., Van Opstal, F., & Van den Bussche, E. (2017). Subjective experience of difficulty depends on multiple cues. *Scientific Reports*, *7*, Article 44222. https://doi.org/10.1038/srep44222

Donders, F. C. (1969). Over de snelheid van psychische processen [On the speed of mental processes]. *Acta Psychologica*, *30*, 412–431. https://doi.org/10.1016/0001-6918(69)90065-1 (Reprinted from Attention and performance II, by W. G. Koster, Trans., 1868, North Holland Publishing Company)

Droit-Volet, S., & Coull, J. (2015). The developmental emergence of the mental time-line: Spatial and numerical distortion of time judgement. *PLOS ONE*, *10*(7), Article e0130465. https://doi.org/10.1371/journal.pone.0130465

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*(4), 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

Ficarella, S. C., Rochet, N., & Burle, B. (2019). Becoming aware of subliminal responses: An EEG/EMG study on partial error detection and correction in humans. *Cortex*, *120*, 443–456. https://doi.org/10.1016/j.cortex.2019.07.007

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *American Psychologist*, *34*(10), 906–911. https://doi.org/10.1037/0003-066X.34.10.906

Fleming, S. M., Maniscalco, B., Ko, Y., Amendi, N., Ro, T., & Lau, H. (2015). Action-specific disruption of perceptual confidence. *Psychological Science*, *26*(1), 89–98. https://doi.org/10.1177/0956797614557697

Gajdos, T., Fleming, S. M., Saez Garcia, M., Weindel, G., & Davranche, K. (2019). Revealing subthreshold motor contributions to perceptual confidence. *Neuroscience of Consciousness*, *2019*(1), Article niz001. https://doi.org/10.1093/nc/niz001

Jovanovic, L., López-Moliner, J., & Mamassian, P. (2021). Contrasting contributions of movement onset and duration to self-evaluation of sensorimotor timing performance. *European Journal of Neuroscience*, *54*(3), 5092–5111. https://doi.org/10.1111/ejn.15378

Kievit, R., Frankenhuis, W., Waldorp, L., & Borsboom, D. (2013). Simpson's paradox in psychological science: A practical guide. *Frontiers in Psychology*, *4*, Article 513. https://doi.org/10.3389/fpsyg.2013.00513

Klein, M. D., & Stolz, J. A. (2018). Making time: Estimation of internally versus externally defined durations. *Attention, Perception, & Psychophysics*, *80*(1), 292–306. https://doi.org/10.3758/s13414-017-1414-6

Kononowicz, T. W., Roger, C., & van Wassenhove, V. (2019). Temporal metacognition as the decoding of self-generated brain dynamics. *Cerebral Cortex*, *29*(10), 4366–4380. https://doi.org/10.1093/cercor/bhy318

Kononowicz, T. W., & van Wassenhove, V. (2019). Evaluation of self-generated behavior: Untangling metacognitive readout and error detection. *Journal of Cognitive Neuroscience*, *31*(11), 1641–1657. https://doi.org/10.1162/jocn_a_01442

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology*, *77*(6), 1121–1134. https://doi.org/10.1037/0022-3514.77.6.1121

Leib, R., Karniel, A., & Mussa-Ivaldi, F. A. (2017). The mechanical representation of temporal delays. *Scientific Reports*, *7*(1), Article 7669. https://doi.org/10.1038/s41598-017-07289-3

Libet, B., Gleason, C. A., Wright, E. W., & Pearl, D. K. (1983). Time of conscious intention to act in relation to onset of cerebral activity (readiness-potential): The unconscious initiation of a freely voluntary act. *Brain: A Journal of Neurology*, *106*(3), 623–642. https://doi.org/10.1093/brain/106.3.623

Liu, J., & Liu, Q. (2016). Use of the integrated profile for voluntary muscle activity detection using EMG signals with spurious background spikes: A study with incomplete spinal cord injury. *Biomedical Signal Processing and Control*, *24*, 19–24. https://doi.org/10.1016/j.bspc.2015.09.004

Luce, R. (1986). *Response times: Their role in inferring elementary mental organization*. Oxford University Press. https://www.semanticscholar.org/paper/Response-Times%3A-Their-Role-in-Inferring-Elementary-Luce/4f714b26b7ceb23ccad0e47d19a6fd94ec10bf4f

Macar, F., Grondin, S., & Casini, L. (1994). Controlled attention sharing influences time estimation. *Memory & Cognition*, *22*(6), 673–686. https://doi.org/10.3758/BF03209252

Mamassian, P. (2008). Overconfidence in an objective anticipatory motor task. *Psychological Science*, *19*(6), 601–606. https://doi.org/10.1111/j.1467-9280.2008.02129.x

Marti, S., Sackur, J., Sigman, M., & Dehaene, S. (2010). Mapping introspection's blind spot: Reconstruction of dual-task phenomenology using quantified introspection. *Cognition*, *115*(2), 303–313. https://doi.org/10.1016/j.cognition.2010.01.003

Meyer, D. E., Osman, A. M., Irwin, D. E., & Yantis, S. (1988). Modern mental chronometry. *Biological Psychology*, *26*(1–3), 3–67. https://doi.org/10.1016/0301-0511(88)90013-0

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Miller, J., Vieweg, P., Kruize, N., & McLea, B. (2010). Subjective reports of stimulus, response, and decision times in speeded tasks: How accurate are decision time reports? *Consciousness and Cognition*, *19*(4), 1013–1036. https://doi.org/10.1016/j.concog.2010.06.001

Morsella, E., Wilson, L. E., Berger, C. C., Honhongva, M., Gazzaley, A., & Bargh, J. A. (2009). Subjective aspects of cognitive control at different stages of processing. *Attention, Perception & Psychophysics*, *71*(8), 1807–1824. https://doi.org/10.3758/APP.71.8.1807

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, *84*(3), 231–259. https://doi.org/10.1037/0033-295X.84.3.231

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195–203. https://doi.org/10.3758/s13428-018-01193-y

Posner, M. I. (1978). *Chronometric explorations of mind*. Lawrence Erlbaum.

Possamaï, C.-A., Burle, B., Osman, A., & Hasbroucq, T. (2002). Partial advance information, number of alternatives, and motor processes: An electromyographic study. *Acta Psychologica*, *111*(1), 125–139. https://doi.org/10.1016/S0001-6918(02)00019-7

Questienne, L., Atas, A., Burle, B., & Gevers, W. (2018). Objectifying the subjective: Building blocks of metacognitive experiences in conflict tasks. *Journal of Experimental Psychology: General*, *147*(1), 125–131. https://doi.org/10.1037/xge0000370

Rabbitt, P. M. (1966). Errors and error correction in choice-response tasks. *Journal of Experimental Psychology*, *71*(2), 264–272. https://doi.org/10.1037/h0022853

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, *85*, 59–108. https://doi.org/10.1037/0033-295X.85.2.59

Ratcliff, R., Smith, P. L., Brown, S. D., & McKoon, G. (2016). Diffusion decision model: Current issues and history. *Trends in Cognitive Sciences*, *20*(4), 260–281. https://doi.org/10.1016/j.tics.2016.01.007

Reyes, G., & Sackur, J. (2014). Introspection during visual search. *Consciousness and Cognition*, *29*, 212–229. https://doi.org/10.1016/j.concog.2014.08.009

Rochet, N., Spieser, L., Casini, L., Hasbroucq, T., & Burle, B. (2014). Detecting and correcting partial errors: Evidence for efficient control without conscious access. *Cognitive, Affective, & Behavioral Neuroscience*, *14*(3), 970–982. https://doi.org/10.3758/s13415-013-0232-0

Sanford, A. J. (1970). Rating the speed of a simple reaction. *Psychonomic Science*, *21*(6), 333–334. https://doi.org/10.3758/BF03335809

Seabold, S., & Perktold, J. (2010). Statsmodels: Econometric and statistical modeling with python. *Proceedings of the 9th python in science conference* (pp. 92–96). https://doi.org/10.25080/Majora-92bf1922-011

Servant, M., White, C., Montagnini, A., & Burle, B. (2016). Linking theoretical decision-making mechanisms in the Simon task with electrophysiological data: A model-based neuroscience study in humans. *Journal of Cognitive Neuroscience*, *28*(10), 1501–1521. https://doi.org/10.1162/jocn_a_00989

Simpson, E. H. (1951). The interpretation of interaction in contingency tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, *13*(2), 238–241. https://doi.org/10.1111/j.2517-6161.1951.tb00088.x

Spieser, L., & Burle, B. (2025). *MYOnset: A python package to detect EMG onset for electrophysiological studies*. Technical report.

Stanton, J. D., Sebesta, A. J., & Dunlosky, J. (2021). Fostering metacognition to support student learning and performance. *CBE Life Sciences Education*, *20*(2), Article fe3. https://doi.org/10.1187/cbe.20-12-0289

Sternberg, S. (1969). The discovery of processing stages: Extensions of Donders' method. *Acta Psychologica*, *30*, 276–315. https://doi.org/10.1016/0001-6918(69)90055-9

Ulrich, R., Mattes, S., & Miller, J. (1999). Donders's assumption of pure insertion: An evaluation on the basis of response dynamics. *Acta Psychologica*, *102*(1), 43–76. https://doi.org/10.1016/S0001-6918(99)00019-0

Vallat, R. (2018). Pingouin: Statistics in Python. *Journal of Open Source Software*, *3*(31), Article 1026. https://doi.org/10.21105/joss.01026

Vidal, F., Burle, B., Grapperon, J., & Hasbroucq, T. (2011). An ERP study of cognitive architecture and the insertion of mental processes: Donders revisited: ERPs and the insertion assumption. *Psychophysiology*, *48*(9), 1242–1251. https://doi.org/10.1111/j.1469-8986.2011.01186.x

Weindel, G., Anders, R., Alario, F.-X., & Burle, B. (2021). Assessing model-based inferences in decision making with single-trial response time decomposition. *Journal of Experimental Psychology: General*, *150*(8), 1528–1555. https://doi.org/10.1037/xge0001010

Yeung, N., & Summerfield, C. (2012). Metacognition in human decision-making: Confidence and error monitoring. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1594), 1310–1321. https://doi.org/10.1098/rstb.2011.0416

(*Appendices follow*)

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Appendix A

Choice of Mixed Model, Experiment 1

To determine which predictors to include in the mixed model analysis, we compared several models including either reaction time (RT) only, RT and a subset of experimental factors, or RT and all experimental factors as fixed effects. The different models tested and their corresponding Akaike information criterion (AIC) are presented in Table A1. Likelihood ratio tests revealed that the full model with all experimental factors outperformed the others (all $p < .001$).

Table A1
Experiment 1: Different Models Tested and Their Corresponding AIC

Model	AIC
RT	−14,327
RT + angle	−14,395
RT + orientation	−14,433
RT + angle + orientation	−14,492

Note. RT = reaction time; AIC = Akaike information criterion.

Appendix B

Mean Omission Rate and Accuracy, Experiment 1

Trials without any recorded response, for example, when no response was given in less than 1 s, were considered as omissions (3.9% of trials). Omission and error rates were arcsine transformed, and two-way analyses of variance (ANOVAs) were performed to assess the effect of experimental conditions. The omission rate was higher with smaller angles, $F(1, 89) = 91.9$, $p < .001$, $\eta_p^2 = 0.51$, but no effect of orientation, $F(1, 89) = 0.62$, $p = .43$, nor any interaction between these two factors, $F(1, 89) = 0.68$, $p = .41$, was observed. Angle and orientation also affected accuracy, with higher error rate with smaller angles, $F(1, 89) = 444.38$, $p < .001$, $\eta_p^2 = 0.83$, and for counterclockwise stimuli, $F(1, 89) = 4.6$, $p = .035$, $\eta_p^2 = 0.05$.

Appendix C

Results of the Go/No-Go Task, Experiment 1

Results of the choice task were replicated in the Go/No-Go task. In the Go/No-Go task, RTs were slower when stimulus angle was smaller, $t(89) = -18.56$, $p < .001$. This was also the case for introspective reaction time (iRT), $t(89) = -9.32$, $p < .001$. Repeated-measures correlation analysis revealed a positive correlation between the two variables, $r_{rm}(3,807) = 0.61$, $p < .001$. Correlation coefficients were also computed for each participant using Pearson's correlation (range = −0.16–0.92, $M = 0.59$, $SD = 0.23$). Mixed model with RT and angle as predictors of iRT showed the main effects of RT ($\beta = 0.59$, $t = 42.22$, $p < .001$) and angle ($\beta = -0.001$, $t = -2.5$, $p = .013$). This means that, as in the choice task, for the same value of RT, iRT was higher for smaller angles.

Appendix D

Mediation Analysis, Experiment 1

To see if the whole angle effect on iRT was mediated by RT in Experiment 1, we performed a mediation analysis with angle as the predictor, RT as the mediator, and iRT as the dependent variable. Results, presented in Table D1, showed that even after accounting for the mediating role of RT (Average Causal Mediation Effect [ACME] = −0.009, $p < .001$), angle still had a direct effect on iRT (Average Direct Effect [ADE] = −0.002, $p < .001$).

Table D1
Mediation Analysis Results With Data From Experiment 1

Mediation effect	Estimate	95% CI	*p*
ACME	−0.009	[−0.0103, 0.0076]	<.001
ADE	−0.002	[−0.0035, −0.0009]	<.001
Total Effect	−0.011	[−0.013, −0.009]	<.001
Prop. Mediated	0.79	[0.72, 0.90]	<.001

Note. Stimulus angle was used as the predictor, RT as the mediator, and iRT as the dependent variable. The 95% CI indicates the confidence intervals. ACME = Average Causal Mediation Effect; ADE = Average Direct Effect; Total Effect = combined direct and indirect effects; Prop. Mediated = the ratio of these estimates; RT = reaction time; iRT = introspective reaction time.

(*Appendices continue*)

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Appendix E

Range of Used Forces, Experiment 2

In Table E1, we report the range of response forces used among participants.

Table E1
Range of Response Forces Used

Parameter	MVF	Weak force	Strong force
Minimum	39.07	1.95	7.81
Maximum	98.80	4.94	19.76
M	66.94	3.35	13.39
SD	18.41	0.92	3.68

Note. Forces are presented in Newtons. MVF = maximum voluntary force.

Appendix F

Choice of Mixed Model, Experiment 2

To determine which predictors to include in the mixed model analysis, we compared several models including either RT only, RT and a subset of experimental factors, or RT and all experimental factors as fixed effects. The different models tested and their corresponding AIC are presented in Table F1. Likelihood ratio tests revealed that the full model with all experimental factors outperformed the others (all $p < .001$).

Table F1
Experiment 2: Different Models Tested and Their Corresponding AIC

Model	AIC
RT	−26,351
RT + angle	−26,406
RT + force	−27,143
RT + force + angle	−27,273
RT + angle + force + orientation	−28,338

Note. RT = reaction time; AIC = Akaike information criterion.

Appendix G

Mean Omission Rate and Accuracy, Experiment 2

Trials without any recorded response, for example, when no response was given in less than 1 s, were considered as omissions (1.5% of trials). Omission and error rates were arcsine transformed, and two-way ANOVAs were performed to look at the effect of experimental conditions. Omission rate increased as stimulus angle decreased, $F(4, 116) = 24.51$, $p < .001$, $\eta_p^2 = 0.46$, $\varepsilon = 0.84$, and was higher in the strong force condition, $F(1, 29) = 19.32$, $p < .001$, $\eta_p^2 = 0.40$, and for counterclockwise stimuli, $F(1, 29) = 5.52$, $p = .026$, $\eta_p^2 = 0.16$ (see Figure L2A). No first- nor second-order interactions were observed (all $p > .1$). Accuracy increased with stimulus angle, $F(4, 116) = 126.55$, $p < .001$, $\eta_p^2 = 0.81$, $\varepsilon = 0.53$, and was higher in the strong force condition, $F(1, 29) = 22.89$, $p < .001$, $\eta_p^2 = 0.44$, and for clockwise stimuli, $F(1, 29) = 12.19$, $p = .002$, $\eta_p^2 = 0.30$ (see Figure L2B). An interaction between angle and orientation was observed, $F(4, 116) = 7.11$, $p < .001$, $\eta_p^2 = 0.20$, $\varepsilon = 0.49$. No other interaction was present ($p > .1$).

(*Appendices continue*)

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

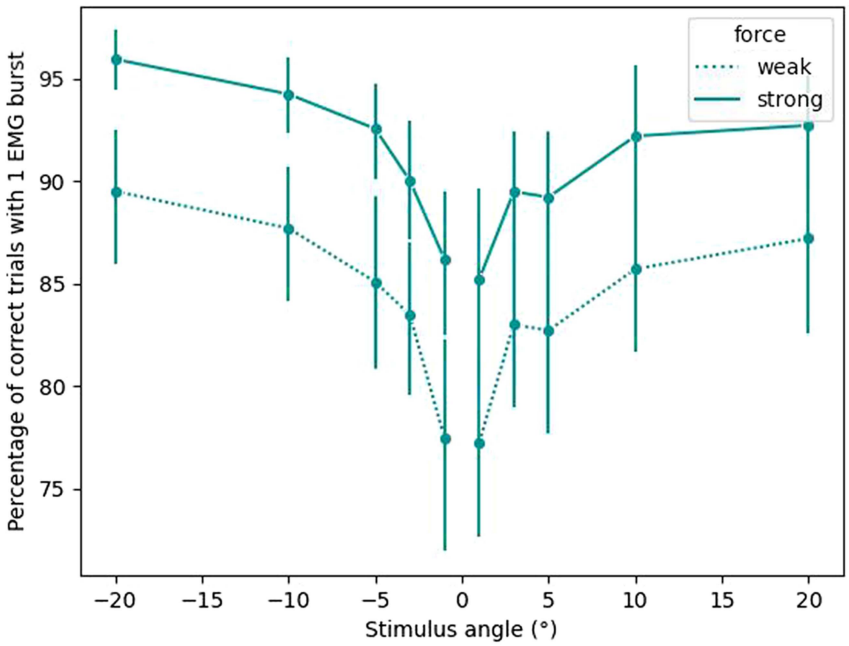
Appendix H

Number of Trials With One EMG Burst, Experiment 2

In Experiment 2, trials with multiple EMG bursts were not included in the first step of analyses. The percentage of remaining correct trials in each condition is reported on Figure H1. A three-way repeated-measures ANOVA revealed effects of angle, $F(4, 116) = 34.29$, $p < .001$, $\eta_p^2 = 0.54$, $\varepsilon = 0.77$, and force, $F(1, 29) = 47.84$, $p < .001$, $\eta_p^2 = 0.62$, on the percentage of trials.

Figure H1

Percentage of Correct Trials Included, That Is, With One EMG Burst, as a Function of Stimulus Angle and Response Force



Note. Negative and positive angles represent counter- and clockwise orientations, respectively. EMG = electromyographic. See the online article for the color version of this figure.

(*Appendices continue*)

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Appendix I

Mediation Analyses, Experiment 2

To see if angle and force had a direct effect on iRT besides their effect, we ran two mediation analyses with angle or force as the predictor, RT as the mediator, and iRT as the dependent variable. Results, presented in Table I1, showed that even after accounting for the mediating role of RT (ACME = −0.0127, $p < .001$), force still had a direct effect on iRT (ADE = −0.0175, $p < .001$). On the contrary, all the angle effect was mediated by RT (ACME = −0.001325, $p < .001$), with no direct effect remaining (ADE = −0.000233, $p = .054$).

Table I1
Mediation Analysis Results With Data From Experiment 2

Mediation effect	Estimate	95% CI	*p*
Response force			
ACME	−0.013	[−0.014, 0.011]	<.001
ADE	−0.018	[−0.021, −0.014]	<.001
Total Effect	−0.030	[−0.034, −0.027]	<.001
Prop. Med.	0.42	[0.37, 0.48]	<.001
Stimulus angle			
ACME	−0.0013	[−0.0014, −0.0012]	<.001
ADE	−0.00023	[−4.9e−04, −7.3e−07]	.054
Total Effect	−0.0016	[−0.0018, −0.0013]	<.001
Prop. Med.	0.849	[0.732, 1.001]	<.001

Note. Stimulus angle or response force was used as the predictor, RT as the mediator, and iRT as the dependent variable. The 95% CI indicates the confidence intervals. ACME = Average Causal Mediation Effect; ADE = Average Direct Effect; Total Effect = combined direct and indirect effects; Prop. Med. = the ratio of these estimates. RT = reaction time; iRT = introspective reaction time.

(*Appendices continue*)

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

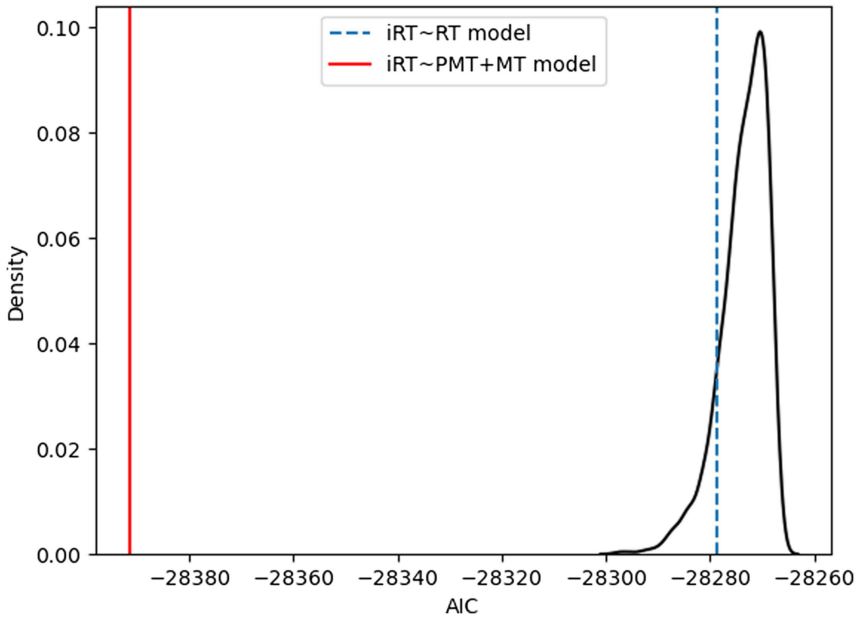
Appendix J

(Pseudo) Random Splits of RTs, Experiment 2

Prediction of iRT was better with both PMT and MT rather than RT (which is exactly the sum of PMT and MT). To ensure that such an outcome is not a mechanical effect of splitting RT into two subintervals, we compared the results to shuffled data. To do so, for each trial, we measured the ratios PMT/RT and MT/RT (how much PMT and MT contribute to RT; by construction, the two sum to 1). We obviously kept constant the couple iRT–RT, but shuffled the ratios across trials and computed new PMT and MT for each trial based on these shuffled ratios (note that the mean ratios PMT/RT and MT/RT are hence kept constant in all simulations). With these new PMT and MT, we ran the model iRT ~ shuffled PMT + shuffled MT and computed the corresponding AIC. We ran 1,000 simulations (1,000 shufflings) and built the distribution of the obtained AIC. The AIC resulting from these 1,000 shufflings had a mean of −28274 and a standard deviation of 4.58. While the AIC of the iRT ~ RT model was included in this distribution (z score = −1.1), the AIC of the iRT ~ PMT + MT model was clearly lower (z score = −25.76; see Figure J1).

Figure J1

Distribution of AICs of Models With Randomly Generated PMTs and MTs as Predictors



Note. The red (solid) line represents the AIC of the iRT ~ PMT + MT model, and the blue (dashed) line represents the AIC of the iRT ~ RT model. iRT = introspective reaction time; PMT = premotor time; MT = motor time; RT = reaction time; AIC = Akaike information criterion. See the online article for the color version of this figure.

(*Appendices continue*)

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

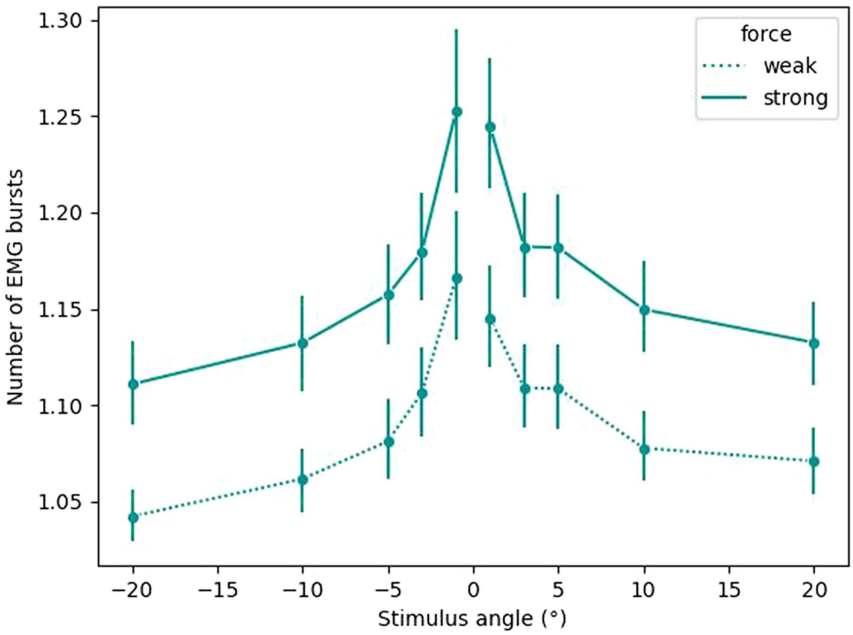
Appendix K

Number of EMG Bursts, Experiment 2

We looked at the effect of experimental conditions on the number of EMG bursts using three-way ANOVAs with stimulus angle, orientation, and force as within-subjects variables (see Figure K1). The number of bursts was increased when stimulus angle decreased, $F(4, 116) = 36.0$, $p < .001$, $\eta_p^2 = 0.55$, $\varepsilon = 0.61$. It was also smaller in the weak force condition, $F(1, 29) = 49.29$, $p < .001$, $\eta_p^2 = 0.63$.

Figure K1

Mean Number of EMG Bursts as a Function of Stimulus Angle and Response Force



Note. Negative and positive angles represent counter- and clockwise orientations, respectively. EMG = electromyographic. See the online article for the color version of this figure.

(*Appendices continue*)

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

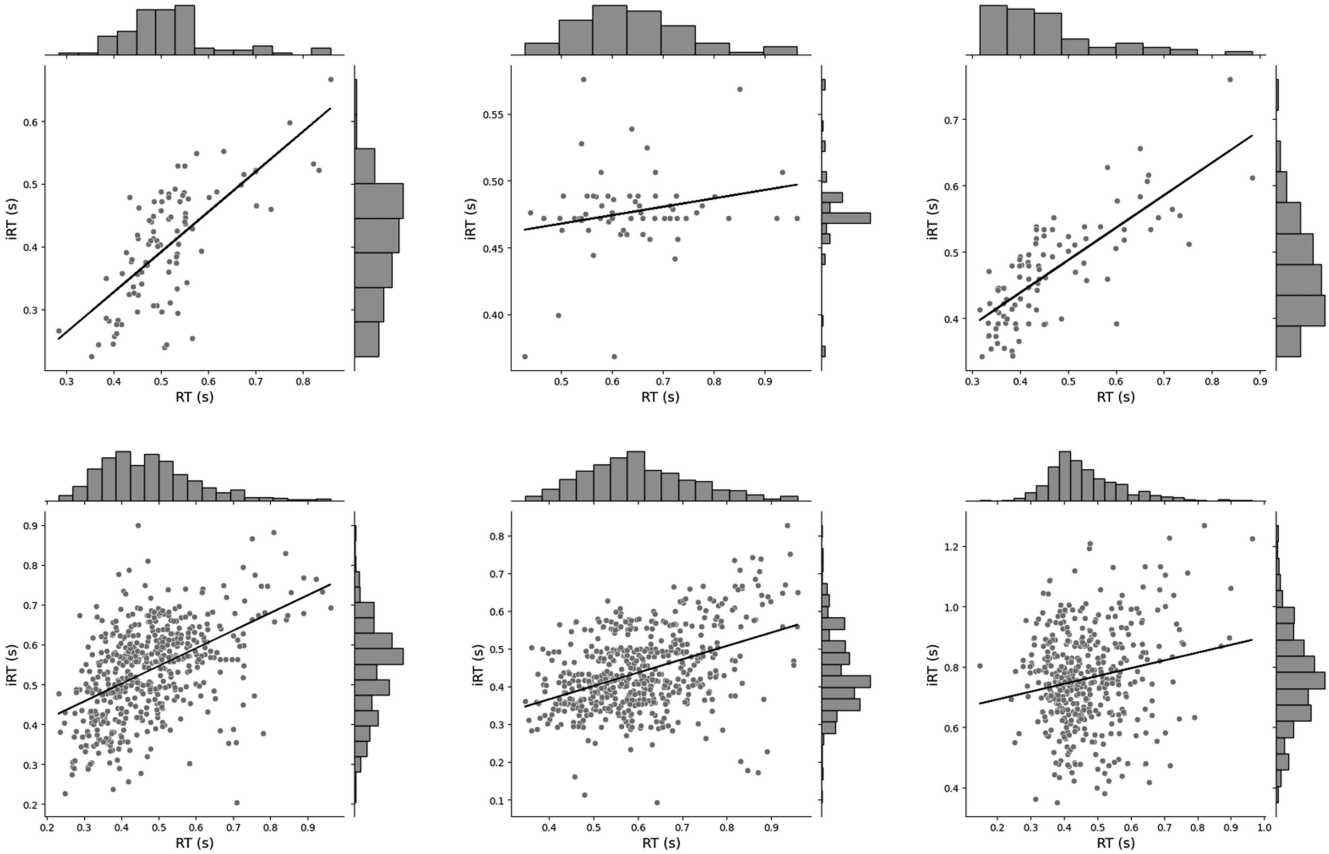
Appendix L

Examples of Individual RT–iRT Correlations

To provide a sense of intraindividual variability in the relationship between RT and iRT, we present the scatterplots of the three first participants of Experiments 1 and 2 in Figure L1.

Figure L1

Scatterplots of Individual iRT as a Function of RT for the Three First Participants of Experiment 1 (Top Row) and Experiment 2 (Bottom Row)



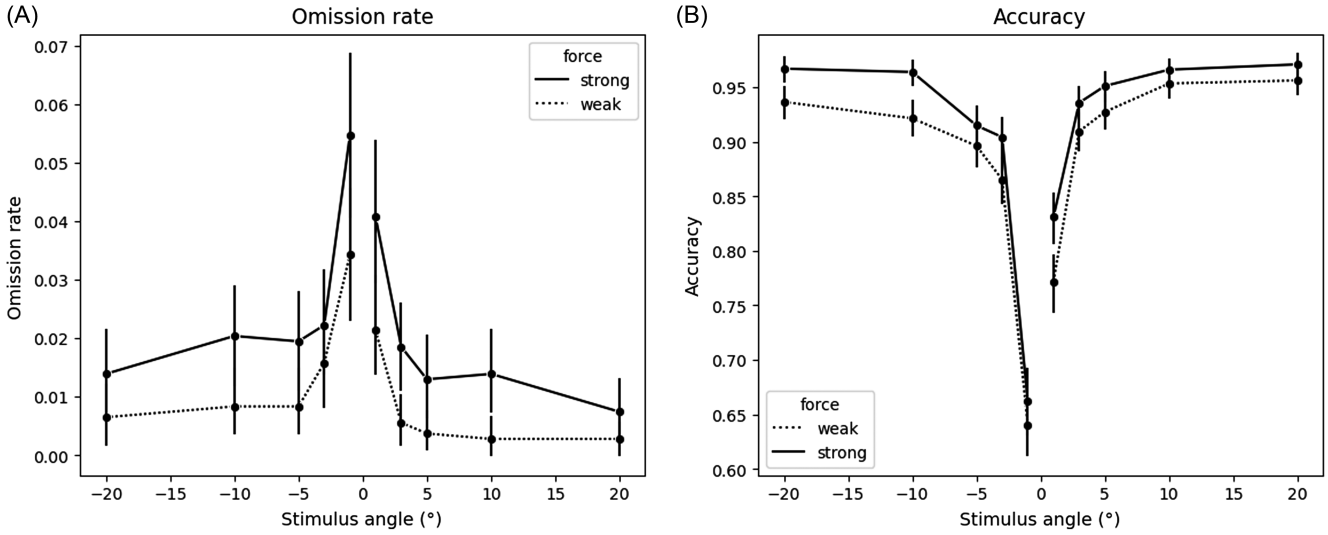
Note. Each scatterplot is associated with histograms of both variables and the linear regression line. RT = reaction time; iRT = introspective reaction time.

(*Appendices continue*)

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Figure L2

Experiment 2: (A) Omission Rate and (B) Mean Accuracy as a Function of Stimulus Angle and Response Force



Note. Negative and positive angles represent counter- and clockwise orientations, respectively.

Received March 7, 2024
Revision received October 19, 2024
Accepted October 23, 2024 ■

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.