Aix*Marseille université
Socialement engagée

NNT : 2023AIXM0001

THÈSE DE DOCTORAT

Soutenue à Aix-Marseille Université
le 5 novembre 2024 par

Nathalie PAVAILLER

Etude de l'évaluation métacognitive des temps de réaction à travers des approches comportementales et électrophysiologiques.

Discipline
Biologie Santé

Spécialité
Neurosciences

École doctorale
Sciences de la Vie et de la Santé – ED62

Laboratoire/Partenaires de recherche
Centre de Recherche en Psychologie et Neurosciences

Composition du jury

Emmanuel PROCYK Université Claude Bernard (Lyon 1)	Rapporteur
Anne BONNEFOND Université de Strasbourg	Rapporteure
Pascal MAMASSIAN Ecole Normale Supérieure	Examinateur
Lucie CHARLES Queen Mary University of London	Examinatrice
Clémence ROGER Université de Lille	Examinatrice
Virginie VAN WASSENHOVE Université Paris-Saclay	Examinatrice
Thomas BROCHIER Aix-Marseille Université	Président du jury
Boris BURLE Aix-Marseille Université	Directeur de thèse

cnrs

CRPN
CENTRE DE RECHERCHE EN PSYCHOLOGIE & NEUROSCIENCES

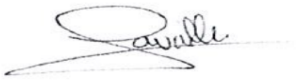
NEUROSCHOOL
Aix*Marseille Université

Affidavit

Je soussignée, Nathalie Pavailler, déclare par la présente que le travail présenté dans ce manuscrit est mon propre travail, réalisé sous la direction scientifique de Boris Burle, dans le respect des principes d'honnêteté, d'intégrité et de responsabilité inhérents à la mission de recherche. Les travaux de recherche et la rédaction de ce manuscrit ont été réalisés dans le respect à la fois de la charte nationale de déontologie des métiers de la recherche et de la charte d'Aix-Marseille Université relative à la lutte contre le plagiat.

Ce travail n'a pas été précédemment soumis en France ou à l'étranger dans une version identique ou similaire à un organisme examinateur.

Fait à Marseille le 04 septembre 2024



Cette œuvre est mise à disposition selon les termes de la Licence Creative Commons Attribution - Pas d'Utilisation Commerciale - Pas de Modification 4.0 International.

Liste de publications et participation aux conférences

Liste des publications réalisées dans le cadre du projet de thèse:

1. **Pavailler, N.**, Gevers, W., Burle, B., (2024). *Temporal metacognition: direct readout or mental construct? The case of introspective reaction time.* In revision. Journal of Experimental Psychology: General.

2. **Pavailler, N.**, Vidal, F., Burle, B., (2024). *Metacognitive context tunes medial prefrontal cortex performance monitoring activity.* Submitted for publication.

Participation aux conférences au cours de la période de thèse:

Présentations orales:

1. **Pavailler, N.**, (2023). *Metacognitive judgments: direct readout or causal inference? The case of introspective reaction times.* Congrès de l'Association for the Scientific Study of Consciousness (ASSC), New York, Etats-Unis.

Posters:

1. **Pavailler, N.**, Vidal, F., Burle, B., (2024). *Metacognitive instructions finely tune medial pre-frontal cortex performance monitoring activity.* Workshop Cognitive Control and Performance Monitoring (CCPM), Oostduinkerke, Belgique.

2. **Pavailler, N.**, Burle, B., (2023). *Metacognitive judgments: direct readout or causal inference? The case of introspective reaction times.* Timing Research Forum (TRF), Lisbonne, Portugal.

3. **Pavailler, N.**, Burle, B., (2022). *What's on a subjective evaluation of time? Dissociating premotor and motor components in RT task reveals the building blocks of time metacognition.* Congrès de l'European Society for Cognitive Psychology (ESCOP), Lille, France.

Résumé

La métacognition, soit la connaissance de nos propres performances et états cognitifs, nous permet d'adapter constamment nos comportements et est ainsi essentielle à l'apprentissage. Dans des situations de choix, les recherches sur la metacognition se sont largement concentrées sur les jugements de confiance ou la détection d'erreur, reflétant tous deux un jugement sur la qualité de la décision. Or, dans de nombreux cas, le temps mis pour réaliser une action est tout aussi important pour atteindre son but. Cette thèse porte spécifiquement sur cette " métacognition temporelle ". Nous avons étudié les capacités de participants adultes à évaluer subjectivement la durée de leurs temps de réaction (TR) lors d'une tâche de détection d'orientation de stimuli visuels.

Les deux premières études montrent que l'estimation du TR repose sur divers indices, incluant des informations temporelles liées à la durée objective du TR et des informations contextuelles non temporelles. Des mesures électromyographiques (EMG) ont également révélé que des processus décisionnels et moteurs sont accessibles pour l'évaluation métacognitive du TR.

Dans deux études suivantes, nous avons montré que l'activité du cortex préfrontal médian (MPFC), mesurée en EEG par l'amplitude de la Négativité d'Erreur (Ne), était sensible au TR, uniquement lorsque l'instruction métacognitive portait sur cette dimension. Ces résultats nous ont permis de mieux caractériser le système d'évaluation de la performance, indiquant qu'il peut être sensible à des dimensions variées de l'action, selon le contexte métacognitif en cours. Nous mettons en évidence que l'activité du MPFC est lié au temps de décision, et qu'il ne constitue qu'une composante du circuit permettant les jugements de TR, au vu de leur multidimensionnalité. Le lien entre cette activité, rapports subjectifs et capacités métacognitives reste à clarifier.

Dans un dernier chapitre empirique, nous montrons que les jugements métacognitifs modifient la performance de premier ordre, à la fois globalement et par des régulations séquentielles. Enfin, nous présentons une nouvelle méthodologie de détection de l'activité EMG en temps réel, ouvrant la voie à des études futures sur le rôle causal du MPFC dans la métacognition temporelle.

L'ensemble de ces travaux offre une compréhension approfondie des mécanismes sous-jacents à l'estimation des TR et inscrit cette capacité dans un cadre plus large d'évaluation de la performance.

Mots-clés: Evaluation de la performance, métacognition, temps de réaction, électrophysiologie.

Abstract

Metacognition, or the awareness of our own cognitive performance and states, allows us to constantly adapt our behavior and is thus essential for learning. In decision-making situations, research on metacognition has largely focused on confidence judgments or error detection, both reflecting an assessment of decision quality. However, in many cases, the time taken to perform an action is equally important for achieving its goal. This thesis specifically addresses this "temporal metacognition." We studied the ability of adult participants to subjectively evaluate the duration of their reaction times (RT) during a visual stimulus orientation detection task.

The first two studies show that RT estimation relies on various cues, including temporal information related to the objective RT duration and non-temporal contextual information. Electromyographic (EMG) measurements also revealed that both decisional and motor processes are accessible for the metacognitive evaluation of RT.

In two subsequent studies, we demonstrated that the activity of the medial prefrontal cortex (MPFC), measured through EEG by the amplitude of the Error Negativity (Ne), was sensitive to RTs only when the metacognitive instruction focused on this dimension. These findings allowed us to better characterize the performance monitoring system, indicating that it can be sensitive to various action dimensions depending on the ongoing metacognitive context. We highlight that MPFC activity is linked to decision time and that it constitutes only one component of the circuit underlying RT judgments, given their multidimensionality. The relationship between this activity, subjective reports, and metacognitive abilities remains to be clarified.

In a final empirical chapter, we show that metacognitive judgments influence first-order performance, both globally and through sequential adjustments.

Finally, we present a new real-time EMG activity detection methodology, paving the way for future studies on the causal role of the MPFC in temporal metacognition.

Overall, this body of work provides an in-depth understanding of the mechanisms underlying RT estimation and places this ability within a broader performance monitoring framework.

Keywords: performance monitoring, metacognition, reaction times, electrophysiology

Remerciements

"La reconnaissance est non seulement la plus belle, mais encore la mère de toutes les autres vertus."
– Marcus Tullius Cicero

Si je suis en train d'écrire ces lignes, c'est que je suis (très bientôt, et enfin) au bout de mon aventure doctorale. Le chemin n'aura pas été sans embûches, doutes, et parfois désillusions, que je n'aurais pas pu surmonter sans l'entourage d'autant de personnes bienveillantes dont j'ai bénéficié. Ce n'est pas sans émotions que je leur exprime toute ma gratitude aujourd'hui.

En premier lieu, je tiens à remercier les membres de mon jury, pour avoir accepté de lire et d'évaluer ce travail de thèse.

Le plus grand merci revient tout droit à mon directeur. Boris, merci pour ta confiance, depuis l'époque où je n'étais qu'une petite stagiaire, qui me parait si lointaine. Merci de m'avoir inculqué une science rigoureuse, où les bonnes questions sont plus importantes que les bonnes réponses. Merci pour ces "times we enjoyed wasting that were not wasted times", pour avoir canalisé mes débordements de panique, et pour avoir fait de ce manuscrit ton livre de chevet pendant tes vacances d'été.

Merci au reste de l'équipe ContAct/ACDC. Morgane, Aurélie et Gabriel, merci pour vos quelques conseils glissés entre mon arrivée et vos départs. Merci à Laurence, de m'avoir toujours ouvert la porte de ton bureau quand j'en avais besoin, et de m'avoir transmis le goût de l'enseignement. Merci à Franck pour la transmission d'une part de ton immense savoir, du master jusqu'à l'écriture de mon deuxième papier, et de m'avoir acceptée dans le club restreint des passionnés de la Négativité d'erreur. Merci à Jenny, d'avoir été ma manageuse réseau professionnel à Lisbonne. Merci à Thierry, Hélène, et Marie-Hélène.

Merci à Marieke et Thomas, pour les discussions enrichissantes et motivantes de mes comités de suivi de thèse. Merci à Julien, d'avoir été toujours fidèle dans la gestion du bureau des plaintes et du bureau des pleurs. Merci à Didier et Dany, pour les innombrables heures passées à faire fonctionner mon dispositif de réponse. L'électronique n'a plus de secrets pour moi grâce à vous. Merci à Kamila, de m'avoir encadrée pendant mon stage de M2, c'est grâce à toi que je suis devenue experte en pose d'électrodes EEG. Merci à Laure, de m'avoir appris à utiliser Psychopy, et de m'avoir fourni les scripts de marquage EMG qui ont occupé une large partie de mon temps au labo. Merci à Xavier, pour ton aide précieuse dans ma bataille avec les modèles mixtes. Thank you Wim, for your valuable comments and advices for the writing of my first paper. And for letting me take part in an amazing workshop on Cognitive Control and Performance Monitoring. Merci à Vincent et Joëlle, c'est

parce que vous vous battez pour nous avec Nautilus et Etamine, que j'ai pu participer à de super conférences. Merci à l'ensemble des chercheurs, chercheuses et autres membres du laboratoire, et tout spécialement aux résidents du premier étage aile est, pour l'ambiance conviviale qu'ils y créent. Un merci particulier à Pierre-Yves, pour les nouveaux fonds d'écran et autres bêtises, à Aline pour ta douce voix qui résonne tard le soir quand (presque) plus personne n'est là, et à Martine pour les instants où tu viens papoter avec moi au détour d'un passage dans le couloir.

Un grand merci à l'incroyable équipe de doctorant.e.s du LNC/CRPN, dont la plupart ont terminé l'aventure avant moi. Sonia, Soso, Dibisonia, d'une simple collègue de labo tu es devenue une amie exceptionnelle. Je ne saurais te remercier assez pour tout ce que tu m'as apporté pendant ces années. De mes (nombreux) craquages dans ton bureau, à la folie Léon Marchand pendant mon épisode dans ta nouvelle vie parisienne, en passant par nos vacances au Canada, tu as été d'un soutien sans fin, au labo et en dehors. Je n'y serais pas arrivé sans toi. Lisa, partenaire de bureau, d'aventure new-yorkaises, de discussions scientifiques (ou non) autour du café le matin, et un peu de galères il faut le dire. Merci pour ton incroyable bienveillance, pour tes mots toujours rassurants, et pour l'exemple de force et de réussite que tu m'as donné. Merci aux autres partenaires, passés ou présents, du meilleur bureau : Rebecca, merci pour toutes ces petites discussions, pour tes conseils, et pour m'avoir appris à me servir de la relieuse ;). Ali, thank you for supporting our constant gossiping, my stupid jokes, and above all myself in times of doubt. Special thanks for finding out what happened to my neighbor ;). Estelle, merci de ramener toujours plus de potins, avec ton énergie et ton assurance admirables. Hâte de reprendre les sessions mots-croisés pour confirmer notre illettrisme. Marion, notre lab-manager préférée, merci de m'avoir soutenue, et surtout supportée dans ma dernière ligne droite, jusque dans mes traumas cafards et autres pétages de plomb. Gaëlle, on se sert les coudes depuis que je t'ai trainée à remonter cette piste de ski à pied. Merci pour tes "courages fromage" et photos de Loki qui m'auront motivée jusqu'à la fin. Chloé, même si on n'habite plus tout à fait juste à côté, merci d'avoir été une de mes plus fidèles supportrices. You are my sunshine, my only sunshine. Célia, on a commencé ensemble et on finit ensemble, merci pour ta joie de vivre éternelle, et pour tes incroyables blagues sur les dinosaures. Et Paul, merci de m'avoir presque convaincue de regarder l'attaque des titans. Régis, merci pour ces "calibration, validation" qui résonnent dans ma tête depuis les 4572 heures passées à la TMS. On n'oubliera pas mon 50% de réussite en anniversaire surprise. Nada, merci pour les créations de jeux innovants, qu'ils soient de société ou aquatiques, qui ont animé certaines soirées. Benjamin, merci pour les discussions sur la recherche, et surtout pour tes innombrables blagues, plus ou moins réussies. Et parce que t'avoir vu à la télé, ça n'a pas de prix. Antonin Abillamaps, merci pour les souvenirs de yourte et de baleine. Une dédicace à ce commentateur de trampoline Eurosport, qui m'aura donné un peu de légèreté en période de rédaction. Claire, merci de prendre la relève, après avoir été sans aucun doute mon étudiante favorite, et une stagiaire en or. Merci à tous les autres docs et post-docs, d'hier et d'aujourd'hui, avec qui j'ai eu la chance de faire un bout du chemin : Manon, Jean-Yves, Julien, Lauriane, Guillem, Zoé, Raphaëlle, Margaux, Fabrizio, Elie, Alix, Lou, Mariama, Inga, Marieva,

Antoine, Léa, Andréa, Jacob, Sophia, Rosa, Daniela.

Ju, Chloé, mes petits soleils, l'aventure aurait été bien plus sombre sans vous. Merci d'être toujours là, les jours de pluie qui mènent à des réservations de week-end à Madrid, comme les jours de beau temps qui mènent au bar ou au frioul. Merci de toujours prendre soin de moi, merci d'être vous. Amandine, loin des yeux mais si près du cœur, merci pour les heures passées au téléphone à se raconter nos malheurs (et nos bonheurs !), et de me dédier tes montées de course à pied pour me pousser dans mes derniers efforts. Salomé, merci pour les cafés du dimanche au cours Ju, et surtout d'être une amie qui, malgré un délai de réponse aux textos d'en moyenne 7 mois, réussit toujours à être là quand on en a besoin. Laurie (et Kevin), merci pour les verres ensoleillés aux berthoms. Eva, merci pour ton sourire qui fait un bien fou quand on arrive à se voir. Léa, merci pour le partage réconfortant de nos expériences du doctorat. Antoine, merci d'avoir été de loin mon participant le plus drôle, et d'être un fidèle représentant de la team mouette. Ridou, merci d'être devenu, de manière improbable à l'époque mais tout à fait évidente aujourd'hui, un si bon ami. Théo, merci pour nos rendez-vous mensuels quasi respectés et autres yellow tuesdays. Patou, merci pour les traquenards de ma première année, et merci au reste de votre équipe IHU de m'avoir laissé vous mettre la misère à la pétanque. Quentin et Théo, merci pour nos quelques week-end retrouvailles passés à ne pas visiter Grenoble.

Et parce que je n'oublierai quand même pas mes racines stéphanoises : Merci à mes girls, Marie, Clara, Natalya. Marie, merci d'être mon pilier depuis la maternelle, même si ça n'a pas été sans interruption, et de toujours me rappeler qu'on est des bêtes de meuf. Et merci pour cette expérience incroyable de pose de toile de verre, que je peux désormais ajouter à mon CV. Natalya, merci pour ces moments de colocation à Marseille, et de vacances à Lausanne. Clara, merci pour nos rituels du jour de l'an.

Enfin un merci tout particulier à ma famille, sans qui rien ne serait possible. Papa et Maman, merci d'avoir toujours cru en moi, de m'avoir supportée pendant 2 confinements consécutifs et d'avoir tant bien que mal essayé de comprendre ma thèse. Julie, merci de m'avoir toujours aidé à surmonter toutes les épreuves, dans la thèse comme dans la vie. Seb, merci de m'avoir donné l'exemple d'une thèse brillante il y a 8 ans, et de m'avoir donné un jour le précieux conseil de ne pas mettre tous les œufs dans le même panier. Marion, merci pour les week-ends passés ensemble, qui ont enrichi mon répertoire de glaciers marseillais. Dom, Lek, Téo et Lenalee, merci pour tous ces bons moments passés en famille. Malo, merci d'être arrivé il y a maintenant plus d'un an, l'échographie accrochée au-dessus de mon ordi aura été mon rappel qu'il y a d'autres choses importantes dans la vie.

"Le premier savoir est le savoir de mon ignorance : c'est le début de l'intelligence."
– Socrate

"Le savant n'est pas l'homme qui fournit les vraies réponses; c'est celui qui pose les vraies questions."
–Claude Levi-Strauss

Table of Contents

Affidavit 2

Liste de publications et participation aux conférences 3

Résumé 4

Abstract 5

Remerciements 6

Table of Contents 10

1. General Introduction 13
- 1.1. Subjective judgments as scientific evidence? . . . 13
  - 1.1.1. Historical aspects about introspection . . . 13
  - 1.1.2. A definition of metacognition . . . 16
  - 1.1.3. Metamemory . . . 17
  - 1.1.4. Metacognition in decision-making . . . 18
  - 1.1.5. The interplay between RT and metacognitive experiences . . . 24
- 1.2. Temporal metacognition . . . 25
  - 1.2.1. Time perception . . . 25
  - 1.2.2. Confidence in time estimation . . . 26
  - 1.2.3. Temporal Error monitoring . . . 28
  - 1.2.4. Reaction times as implicit produced durations . . . 31
  - 1.2.5. Introspective reaction times . . . 34
- 1.3. A generic system of performance monitoring . . . 37
  - 1.3.1. Performance monitoring in the brain . . . 37
  - 1.3.2. Performance monitoring beyond errors . . . 42
  - 1.3.3. Subjective performance monitoring? . . . 44
  - 1.3.4. Performance monitoring beyond action selection . . . 46
- 1.4. Outline of the thesis . . . 48

2. Investigating the building blocks of iRT 49
- 2.1. Note on the present chapter . . . 49
- 2.2. Article: Temporal metacognition: direct readout or mental construct? The case of introspective reaction time . . . 50
  - 2.2.1. Abstract . . . 51
  - 2.2.2. Introduction . . . 51
  - 2.2.3. Experiment 1 . . . 54

2.2.4. Experiment 2 . . . 58
2.2.5. General discussion . . . 67
2.2.6. Conclusion . . . 69
2.2.7. References . . . 70

3. MPFC contribution to subjective evaluation 77
3.1. Note on the present chapter . . . 78
3.2. Article: Metacognitive context tunes MPFC performance monitoring activity . . . 79
3.2.1. Introduction . . . 80
3.2.2. Results . . . 82
3.2.3. Discussion . . . 84
3.2.4. Methods . . . 87
3.2.5. Extended Data Figures . . . 94
3.2.6. References . . . 98
3.2.7. Supplementary Information . . . 103
3.3. From objective to subjective performance monitoring: MPFC activity and beyond . . . 110
3.3.1. Introduction . . . 110
3.3.2. Methods . . . 111
3.3.3. Results . . . 112
3.3.4. Discussion . . . 117

4. Second order judgments influence first order behavior 122
4.1. Note on the present chapter . . . 122
4.2. Introduction . . . 122
4.3. Methods . . . 126
4.3.1. Comparison of first order measures . . . 126
4.3.2. Mixed models for sequential effects . . . 126
4.4. Results . . . 127
4.4.1. Effect of global metacognitive context on first-order behavior . . . 127
4.4.2. Effect of local metacognitive judgments on subsequent behavior 129
4.5. Discussion . . . 131

5. A new method for the online detection of EMG activity 134
5.1. Note on the present chapter . . . 134
5.2. Towards an online detection of EMG onsets . . . 135
5.3. Introduction . . . 136
5.3.1. EMG as a marker of interest for cognitive architecture and cognitive control . . . 136
5.4. Existing methods of EMG onset detection . . . 138
5.5. General constraints of online processing . . . 139
5.6. Core algorithm and specific constraints . . . 140
5.6.1. Specific constraints of our acquisition system . . . 140

5.6.2. Core detection algorithm 140
5.7. Technical solutions tested for online EMG onset detection 142
5.7.1. OpenVibe 142
5.7.2. Full custom python program through TCP/IP protocol 144
5.7.3. Other potential solutions to solve the remaining issues 147
5.8. Conclusion 149
5.9. Appendix 151
5.9.1. OpenVibe python box 151
5.9.2. Python code for online processing 153

6. General Discussion **157**
6.1. A general overview of the main findings 157
6.1.1. iRT as an accurate representation of RT 157
6.1.2. iRT as a composite introspective measure 158
6.1.3. RT estimation is embedded within a broader performance monitoring process 161
6.2. An integrative mechanism for RT estimation 162
6.2.1. Performance monitoring as a prediction error computation . . . 162
6.2.2. Temporal metacognition as a two-stage accumulation process . 164
6.2.3. An integrative model for decision time estimation 164
6.3. What about motor execution time? 167
6.4. Knowing how long it took: what for? 169
6.5. The measurement problem 170
6.6. Conclusion 173

Bibliography **174**

APPENDICES **206**
A. Behavioral results of EEG experiment 207
A.1. Dissociating RT and experimental factors contributions to iRT . 207
A.2. Dissociating force and experimental factors contributions to iForce 207
A.3. Impact of PMT and MT on iRT 208
B. Supplementary EEG analyses 208
B.1. Ne and iForce 208
B.2. Ne sensitivity to RT as a function of RT estimation ability 209
B.3. Ne trial-by-trial analyses 210
C. Supplementary behavioral analyses 211
C.1. Dissociation between effects of block on RT and iRT 211
C.2. Individual models with factor effects and timing ability 212

List of Figures **214**

List of acronyms **218**

1. General Introduction

1.1. Subjective judgments as scientific evidence?

"Introspective observation is what we have to rely on first and foremost and always. The word introspection need hardly be defined - it means, of course, the looking into our own minds and reporting what we there discover."
– William James, *The Principles of Psychology (1890)*

1.1.1. Historical aspects about introspection

Our every-day life perceptions and behaviors are constantly accompanied by subjective experiences. For example, when standing in front of a beautiful sunset, it is not only our visual system that interprets the different shapes and colors, we also have a sense of our own existence and individuality, experiencing the scene in a personal way, with evoked emotions and sensations. Through introspection, we can gain insight into these internal experiences and mental states, and eventually report them.

Introspection has for long been a central theme of philosophical discussion, dating back to ancient times with the work of greek philosophers who pondered questions about human consciousness and self-awareness. The maxim "Know thyself", inscribed upon the temple of Apollo in Delphi, underscores the importance of self-knowledge and the critical examination of our assumptions and beliefs for a deeper understanding of ourselves. Socrate, who famously said "The unexamined life is not worth living", promoted introspection and self-awareness as essential to personal growth, and aknowledged the necessity of recognizing our own ignorance in the pursuit of true knoweldge (Plato, 390). Plato further delved into the concept of introspection in his dialogues, positing that true wisdom arises from knowing oneself.

Centuries later, thinking about self-reflection can be seen in the work of Descartes: his famous statement *"Cogito, ergo sum"*, "I think therefore I am", (Descartes, 1637) is often considered a foundational concept in introspective inquiry. Through this radical skepticism, he arrived at the realization that while he could doubt the existence of the external world and even the reliability of his senses, he could not doubt the existence of his own thoughts. This insight placed introspection at the heart of epistemology, as it demonstrated that the mind's awareness of itself is the first indubitable truth.

The term 'introspection', derived from the latin words for 'to look within', did not appear until the end of the seventeenth century, as the philosophical exploration of self-reflection continued to evolve. At this period, John Locke argued that all knowledge originates from experience. He distinguished two types of experience: the

external experience of the physical world, which we acquire through our traditional senses, and the internal experience of our mental world. To describe the latter, Locke often referred to the concept of "inner sense", which accounts for our ability to reflect on and understand our own thoughts and feelings.

These early philosophical questionings laid the groundwork for the development of psychology research. Psychology being the study of human mind and internal states, introspection seems to be a relevant and easy way to access mental contents.

However, introspection as a methodology for psychology has always been controversial. Already in the 18th century, there were critics who questioned the validity of introspective methods. According to David Hume, introspection does not yield objective knowledge because it is subject to personal biases and inconsistencies (Hume, 1739). Immanuel Kant's posited that introspective data are fragile, because private and evanescent, and that it is hence impossible to build a science on it (Kant, 1786). From this point of view, the use of introspection cannot go beyond a descriptive psychology. Another line of criticism comes from Auguste Comte, who challenged the very principle of introspection, considering that self-observation is impossible because a subject cannot act and observe himself acting at the same time (Comte, 1830). Introspection has always faced these contradictory aspects: on one hand it is necessary because how else could we access the human mind? But on the other hand, it does not appear as completely reliable. This paradoxical view has significantly influenced the evolution of its use in psychology.

Introspection was first introduced as a method for experimental psychology by Wilhelm Wundt in the late 19th century (Wundt, 1873). He proposed that psychology should be approached as a scientific discipline, and he believed that introspection could be made more reliable by conducting experiments in highly controlled laboratory settings. He designed experiments in which participants were exposed to specific stimuli under controlled conditions and compared the introspective reports of different individuals or groups under various experimental conditions. He aimed at breaking down complex conscious experiences into their elemental components and at identifying their general principles and regularities. These ideas led to the formalization of structuralism as a school of thought by Wundt's student Edward B. Titchener.

However, introspection as a method faced criticism due to its subjective nature and lack of reliability. Critics argued that different individuals might interpret and report their experiences differently. Consequently, it makes it difficult to draw generalizable conclusions from introspective data, unlike physics or chemistry which are based on stable elements, making experiments repeatable (for an historical overview, see Sackur, 2009). In the early 20th century, John Watson, considered as the father of behaviorism, argued that psychology should focus on observable behavior rather than unobservable mental processes like those studied through introspection (Watson, 1913). Behaviorists believed that mental states and introspective reports are inherently subjective and unscientific, emphasizing the importance of objective, measurable behaviors in the study of psychology. This school of psychology led to a shift away from introspection as a reliable source of scientific evidence. Behaviorists argued that all

behavior is learned through interaction with the environment, and they emphasized the importance of external stimuli in shaping behavior. Key concepts in behaviorism include conditioning, where behaviors are acquired through associations with positive or negative reinforcements (Skinner, 1938).

The decline of behaviorism as the dominant paradigm in psychology began in the mid-20th century, largely due to its limitations in addressing the complexity of human cognition. While behaviorism excelled in explaining observable behaviors through conditioning and reinforcement, it struggled to account for higher-order cognitive processes. Noam Chomsky's 1959 critique of B.F. Skinner's work on language acquisition (Chomsky, 1959) highlighted these shortcomings, arguing that behaviorism could not explain the generative aspects of language. Simultaneously, the cognitive revolution emerged, emphasizing the importance of studying internal mental processes and likening the mind to an information-processing system (G. A. Miller, 2003).

This shift in focus led to a renewed interest in introspection as a method for studying cognition. However, unlike the early introspectionists who relied solely on subjective self-reports, cognitive psychologists adopted a more rigorous and systematic approach to introspection, by combining them with other methods and data. Rather than asking participants to freely report their experiences, cognitive psychologists developed structured introspective techniques that guided participants to observe and report on specific aspects of their cognitive processes. Furthermore, cognitive psychologists supplemented introspective reports with other objective measures, such as reaction times, neural imaging, and behavioral experiments, to validate and complement introspective findings. This multi-method approach helped to overcome some of the limitations of traditional introspection and enhance the reliability and validity of introspective data.

The common controversy surrounding the use of introspection in experimental psychology stems from the methodological challenge it presents. The scientific method states that objective data are the primary source of evidence, leading to disregard and avoid subjectivity. However, this approach seems impossible concerning mental events as they, by definition, cannot be observed "from the outside" but are only accessible from the first-person point of view. The only solution to access them is through subjective reports, with the challenge of using them within the framework of experimental psychology (Overgaard, 2006). The classic work of Nisbett and Wilson (1977) challenged once again the reliability of these introspective reports in understanding cognitive processes. They argue that people often lack access to the underlying causes of their behavior and mental states, leading to inaccuracies and confabulations in verbal reports about their thoughts and motivations. They conducted experiments demonstrating that individuals' explanations for their behaviors are often based on plausible but incorrect attributions, rather than true causal factors. However, their conclusion may not generalize to all situations or cognitive processes, and subjective judgments might still be informative in simpler tasks. In addition, dissociations between subjective reports and objective data can be interpreted to show that the subject's experience can differ from what can be analyzed from his behavior and hence inform us about the introspective process itself. It does not automatically follow

that the introspective report is invalid. In fact, while introspection has for long been considered a method to access mental states, it has now become an object of study by itself, notably with the vast development of consciousness research. The coupling of these subjective data with other objective evidence, behavioral or neurophysiological, allows to establish correlations and infer what can or cannot be accessible to humans' introspection.

1.1.2. A definition of metacognition

Cognitive science research on subjective experience has been referring to the term "metacognition". The prefix "meta" derives from Greek and means "beyond", to indicate a concept that is the abstraction of another, often with a self-reference. Thus, metacognition concerns processes "beyond" cognition. Flavell (1979) was the one to introduce this term in the psychological literature and defined it as "cognition about cognition". Metacognition can therefore refer to the subjective evaluation of both cognitive processes and their outcomes. Addressing the question of the evaluation of the processes themselves is inherently difficult, as experimenters cannot directly observe these internal operations. However, assessing the evaluation of performance offers a more practical approach to studying metacognition. Performance can be objectively measured, making it possible to directly compare these measurements with individuals' subjective evaluations. This comparison provides a clearer and more manageable way to explore metacognitive judgments. In the present case, we will hence restrict our definition of metacognition to the ability to evaluate our own performance in a cognitive task without explicit feedback (Yeung & Summerfield, 2012).

Two levels are hence part of metacognitive processes: an object-level, which is the first-order cognitive task, and a meta-level, which is the second-order representation of the object-level (Nelson, 1990). Both levels interact by means of a monitoring process, by which individuals evaluate ongoing cognitive activities and their effectiveness. In return, the information gathered through monitoring allows individuals to adjust their cognitive processes and strategies to improve performance and/or achieve their goals (see Fig. 1.1).

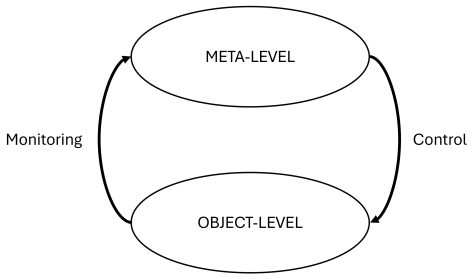


Figure 1.1. *Schematic representation of the Nelson and Narens model, based on Nelson (1990).*

Hence, metacognition covers both introspective and self-regulatory processes (Nor-

man et al., 2019). Flavell (1979) distinguished three facets of metacognition. *Metacognitive knowledge* refers to a person's understanding of cognitive abilities and strategies. This includes knowledge about personal strengths and weaknesses, the nature of different cognitive tasks, and the effectiveness of various learning strategies. *Metacognitive strategies* involve voluntary actions taken to control cognitive activities, such as spending much time studying if level of knowledge appears insufficient. *Metacognitive experiences* refer to the feelings and judgments that arise during a cognitive task. The tip-of-the-tongue phenomenon is a famous example of metacognitive experience, when one have a strong sense that it knows a specific word, but cannot retrieve it (R. Brown & McNeill, 1966).

1.1.3. Metamemory

Metacognitive judgments in the field of memory (metamemory) and learning occupies a central position in the history of metacognition research (A. L. Brown, 1977; Flavell, 1979; Hart, 1965; Nelson, 1990, see Dunlosky and Tauber, 2016). In this context, metacognition can be defined as "knowing about knowing". Individuals are actually able to judge their ability to recall a previously learned item (judgment of learning, Nelson and Dunlosky, 1991) and show higher confidence with greater memory performance (Dunlosky & Metcalfe, 2008; Koriat & Goldsmith, 1996).

A first account of these metacognitive processes stands that it directly relies on the strength of the memory trace (Cohen et al., 1991). Such account would predict a strong correlation between metacognitive judgments and performance in the first-order memory task. However, dissociations can be observed between first and second-order performance. This is the case for the well-known feeling-of-knowing (Hart, 1965) and tip-of-the-tong (R. Brown & McNeill, 1966) phenomenons, where people report knowing an item, but are unable to access it at the moment. Metamemory research has also revealed instances of metacognitive illusions or biases that can influence individuals' memory judgments. For example, the "illusion of knowing" refers to individuals' tendency to overestimate their memory abilities, particularly for familiar or well-learned information (Eakin, 2005). Another theory was developed to account for these dissociations and stipulates that metacognitive judgments are rather inferred from multiple internal or external sources of information, and that the relevance of these sources determines the correlation between first and second order performance (Koriat, 1997). It introduces two accounts of metacognitive monitoring: a direct access, describing a privileged and direct access of task performance (Koriat, 1993), and a cue-utilization process, in which task performance is predicted by the integration of various cues (Koriat, 1997; Schwartz, 1994).

Recent advancements in neuroimaging have provided empirical support for the biological basis of metamemory. Comparing metamory to non-metamemory tasks, fMRI study revealed that metamemory was associated with greater activity in medial prefrontal, medial parietal, and lateral parietal regions (Chua et al., 2009). A study conducted on macaque monkeys also highlighted the involvement of regions of the prefrontal cortex (PFC), demonstrating that their inactivation led to selective impair-

ments in metacognitive ability without affecting the underlying memory processes (Miyamoto et al., 2017). This points to an important role of the PFC in metacognitive processing.

In addition, this ability to monitor memory processes plays a crucial role in enabling learners to regulate and control their own learning processes. By being aware of how well they understand or remember information, learners can make informed decisions about when, and even how, to continue studying (Bjork et al., 2013). This has significant implications for educational research, seeking ways of optimizing learning in students (Stanton et al., 2021).

Research on metamemory has uncovered various mechanisms that apply to other cognitive domains. In particular, a parallel can be drawn with metacognition in the field of decision making, which represents the second dominant body of literature on the topic.

1.1.4. Metacognition in decision-making

1.1.4.1. Confidence

At the perceptual level, scholars have largely focused on the confidence we have in "simple" perceptual decision-making, such as deciding whether a stimulus has been presented or not or deciding which of two possible stimuli has been presented, in two-alternative forced choice (2AFC) tasks (see Mamassian, 2016 for an overview of visual confidence). In fact, when we make a decision, we also concurrently evaluate the likelihood that it will result in a favorable outcome.

In studies of human perceptual decision-making, confidence is often measured with retrospective judgment, where subjects are asked to report their confidence about a previous choice (for a review, see Fleming, 2024). The most common method is confidence ratings, with scales ranging from complete uncertainty to complete certainty. Other methods were developed, some of them being tailored to be applicable to non-humans animals (see below). For example, post-decision wagering was proposed as an objective measure of awareness (Persaud et al., 2007), with the amount wagered serving as an index of decision confidence. Alternatively, other authors introduced a confidence forced-choice paradigm, in which participants indicate which of two decisions produces the highest feeling of confidence (Barthelmé & Mamassian, 2009; Gardelle & Mamassian, 2015), hence avoiding the need to constantly map internal confidence on a rating scale.

Most of the time, humans report confidence judgments that correlate with their objective performance, in that higher confidence is associated with higher accuracy. Evidence for confidence in non-human animals has also been observed, where they opt out of a decision more frequently for difficult trials (Hampton, 2001; Kiani & Shadlen, 2009), decide to abort and restart them (Kepecs et al., 2008), wait longer for a reward after a more certain decision (Kepecs & Mainen, 2012), or seek additional information when unsure (Hampton et al., 2004; Stoll et al., 2016).

But how to precisely measure metacognitive ability? (for a review see Fleming and

Lau, 2014) To assess metacognitive abilities and reliability of confidence judgments, one common approach is to measure the correlation between accuracy and confidence. However, this method alone is limited in its ability to separate true metacognitive sensitivity –the ability to distinguish between correct and incorrect responses– from other factors such as performance and metacognitive bias –the tendency to consistently over- or under-estimate confidence. One mostly used method is the meta d' model, which uses a signal detection theory framework to assess metacognitive sensitivity by estimating the first order performance (d') that would be expected to produce the observed confidence (Galvin et al., 2003; Maniscalco & Lau, 2012).

One fundamental question raised by metacognitive research concerns the sources of metacognitive judgments, including decision confidence, and whether they reflect a direct access to the first order process. Perceptual choices can be described with great details by models of evidence accumulation such as the Drift Diffusion Model (DDM) (Ratcliff & McKoon, 2008), in which sensory evidence is accumulated until it reaches a decision threshold (see Fig. 1.2). It was first proposed that confidence is based on the same information as the first-order decision, that is on the amount of accumulated evidence (Kiani & Shadlen, 2009; Moreno-Bote, 2010). However, as it is the case for memory, some studies have observed dissociations between accuracy and confidence, with cases where estimation of choice confidence is impaired without affecting performance (Del Cul et al., 2009; Fleming, Ryu, et al., 2014). Second, and crucially, subjects occasionally change their mind after their first order decision (Resulaj et al., 2009). This led to the idea that evidence accumulation could continue after the decision in the first order task (Pereira et al., 2021; Pleskac & Busemeyer, 2010). Other models suggest that confidence judgments might be based on different information than the one(s) used to take the decision (Fleming & Daw, 2017), where sensory evidence underpinning decisions and confidence judgments are distinct but coupled. In this framework, self-evaluation of decision performance is achieved by using the confidence sample and one's own actions to infer the performance of the coupled decision system. Actually, it has been shown that confidence judgments were influenced by response times (Kiani et al., 2014; Moreno-Bote, 2010), and also by the motor response itself (Filevich et al., 2020; Fleming et al., 2015; Gajdos et al., 2019; Wokke et al., 2020). Thus, confidence appears to be shaped by both the information that underpins the initial first-order decision and additional information processed after the decision has been made.

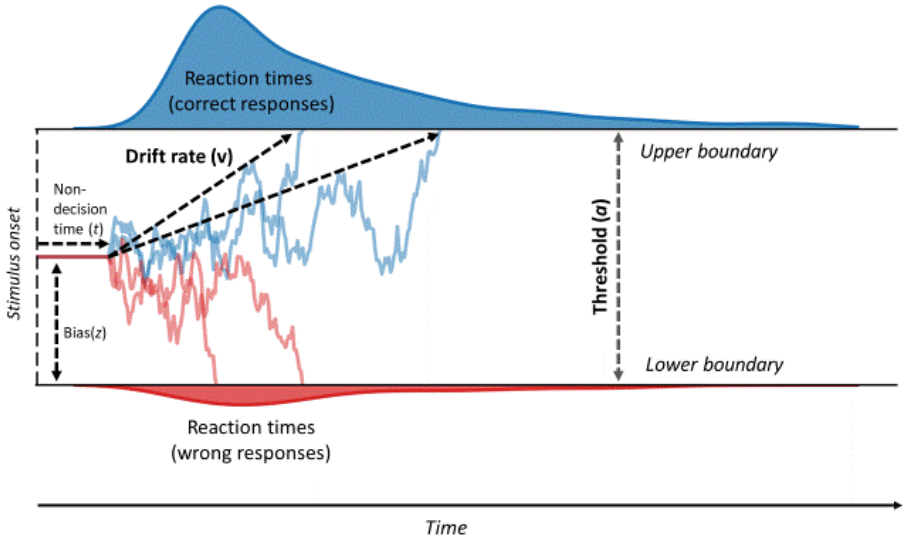


Figure 1.2. *Schematic representation of the two-choice Drift Diffusion Model for several trials. The decision process starts from an initial point and evolves over time, with evidence accumulated towards one of two decision boundaries (upper or lower, representing correct and incorrect responses). The time taken to reach a boundary corresponds to the reaction time, including decision and non-decision times (t). The threshold (a) represents the amount of evidence required to make a decision. The drift rate (v) is the average slope of the evidence accumulation process across a large number of trials.*

Neuroscience research on decision confidence has identified key brain regions which seem to support these metacognitive judgments. First evidence for a cerebral substrate of confidence comes from neuropsychology patients with damages to the PFC who showed impairment in confidence (Fitzgerald et al., 2022; Fleming, Ryu, et al., 2014). Furthermore, interindividual variability in the relationship between performance and confidence is associated with grey matter volume in anterior PFC and white matter density connecting this region (Fleming et al., 2010). Numerous other studies have highlighted a role of different regions of the PFC in metacognitive sensitivity, in humans, monkeys and rodents (for reviews see Fleming and Dolan, 2012; Grimaldi et al., 2015). This apparent domain-general role of the PFC has been nuanced by differences observed between perceptual and memory metacognition (Baird et al., 2013; Fleming, Ryu, et al., 2014; McCurdy et al., 2013), and with the importance and differences in connections with other brain regions (Baird et al., 2013, 2015; Zheng et al., 2021). Overall, the PFC has a pivotal role in confidence judgments, and interacts with a number of other cerebral regions, notably the insula and the cingulate cortex, to promote accurate judgments of performance.

Metacognitive judgments have been hypothesized to have a function of behavioral control. The sense of confidence can indeed play a crucial role in regulating decision making and is intuitively important to plan subsequent actions (Schulz et al., 2023). Indeed, the decision to seek additional information before making a choice depends on the initial confidence in this choice. In animals, low confidence prompts explo-

ration before taking a decision (Call & Carpenter, 2001). One challenge, however, is that as decision accuracy and confidence are often highly correlated, it is difficult to determine if confidence is causally related to adaptive behavior or if they are both driven by lower-level processes (Le Pelley, 2012). Matching for levels of accuracy, Desender et al. (2018) showed that participants were indeed more likely to sample additional information before giving a response in a low confidence condition. This relationship between confidence and information seeking can be flexibly modulated by the perceived value of additional information (Friedemann et al., 2023). For example, for matched confidence, participants were more prone to seek additional information when higher reward value was at stake.

1.1.4.2. Error monitoring

While confidence judgments reflect the probability of having made the correct choice, the opposite act of reporting the likelihood of having made an error has been largely used in studies examining error monitoring. Error monitoring is the metacognitive process that involves detecting and responding to errors in real-time. It is an essential function that allows individuals to recognize when something has gone wrong and to take corrective action.

Research on error monitoring has mostly been using tasks with temporal pressure, in which most action errors are detected by the participants (P. M. Rabbitt, 1966). Rabbitt pioneering work also established that people can rapidly correct their own errors without explicit feedback, hence based on a fast internal signal. The observation, among others, that this ability is reduced when stimulus duration decreases (P. M. A. Rabbitt & Vyas, 1981), led to the idea that a further accumulation of evidence towards the correct choice after an initial accumulation towards the incorrect one induces error detection/correction. This idea shows great similarities with recent theories of decision confidence (Pleskac & Busemeyer, 2010) and changes of mind (Resulaj et al., 2009; van den Berg et al., 2016). Interestingly, error awareness is much slower, and is not necessary for error correction, with individuals sometimes being unaware of errors that they still correct (Nieuwenhuis et al., 2001). In addition, in these type of tasks, electromyographic recordings of the muscles involved in responding allow to reveal small activities in the incorrect side, too weak to reach the response threshold (M. G. H. Coles et al., 1985). These 'partial errors' have been interpreted as errors that have been detected, inhibited and corrected 'online', appearing as a peripheral index of the executive control taking place within a trial (Burle, Possamaï, et al., 2002; Hasbroucq et al., 1999; Scheffers et al., 1996). Most of the time, they remain consciously undetected (Rochet et al., 2014), suggesting once again that this online correction mechanism does not depend on conscious detection.

Research on error monitoring has gained substantial interest after the discovery of EEG activities occurring time-locked to decision errors. The so-called Error Negativity (Ne–Falkenstein et al., 1991, or Error-Related-Negativity–ERN, Gehring et al., 1993) is a negative deflection occurring at fronto-central sites only tens of millisecond after the muscle activation of an incorrect response (see Fig.1.3). Error Negativity

(Ne) has hence been originally interpreted as an error detection mechanism with an amplitude related to the degree of mismatch between the representation of the correct response and the representation of the actual response (M. G. Coles et al., 2001). In this conception, errors are linked to an incomplete processing of the stimulus, and its further processing leads to the representation of the correct response. Some studies suggest that this activity is generated in the anterior cingulate cortex (ACC, Dehaene et al., 1994; Miltner et al., 2003), or in the supplementary motor area (SMA, Bonini et al., 2014; Herrmann et al., 2004; Roger et al., 2010). Cognitive neuroimaging research has consistently shown that the PFC and the anterior cingulate cortex (ACC) play key roles in error processing (Garavan et al., 2002; for a review see Ridderinkhof et al., 2004). Additionally, reduced activity in this neural network has been linked to impairments in error-related processing and subsequent behavioral adjustments in individuals with schizophrenia (Morris et al., 2006), Attention Deficit Hyperactivity Disorder (ADHD, Burgio-Murphy et al., 2007), Alzheimer's disease (Mathalon et al., 2003), and substance use disorders (Franken et al., 2007).

The Ne is followed by a more parietally distributed Error Positivity (Pe) occurring between 200 and 400 ms after the erroneous response (Overbeek et al., 2005, Fig.1.3). The Pe appears to reflect distributed activities across different cerebral sources, with several studies suggesting that the ACC and the insula are among these sources (Ullsperger et al., 2010; Vocat et al., 2008). The functional role of both components is still debated. Models of the Ne include error detection (M. G. Coles et al., 2001), reinforcement learning (Holroyd & Coles, 2002) and conflict monitoring (M. M. Botvinick et al., 2001) among others and it has been interpreted as being related to the rapid and automatic process of detecting error leading to correction. Pe is modulated by different factors and has been associated with conscious error detection, emotional response and assessment of error significance. However, both have been found to vary gradually with subjective confidence (Boldt & Yeung, 2015; Scheffers & Coles, 2000), and their role in error awareness remains to be elucidated (see section 1.3.3).

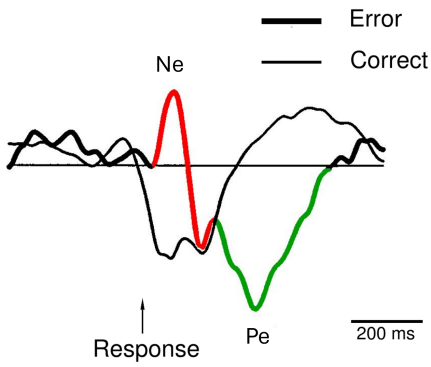


Figure 1.3. *Monopolar grand-averages time-locked to the response, of correct and error trials. The Ne (in red) and the Pe (in green) are highlighted as large negative and positive deflecttions. Adapted from Falkenstein et al. (2000).*

In addition, error awareness is accompanied by vegetative changes such as heart rate, pupil diameter and skin conductance (R. G. O'Connell et al., 2007; J. Wessel et al., 2011), even though it remains difficult to know whether they are a cause or a consequence of error detection (J. Wessel et al., 2011; J. R. Wessel, 2012). This led the authors to propose that several sources of information, including the Ne and sensory inputs can be used as cues for the system to evaluate its performance concerning accuracy, which is likely reflected in the Pe.

Besides their detection and correction, the presence of errors can also modify subsequent behavior. Reaction times (RTs) are longer on trials following erroneous actions (Allain et al., 2009; Laming, 1979b; P. M. Rabbitt, 1966). This post-error slowing (PES) has been interpreted as an adaptive behavior consisting in processing information more carefully to avoid making the mistake again (Dutilh et al., 2012). While this can be supported by correlations between PES and post error accuracy, some authors have challenged this functional interpretation of PES (Burns, 1971; Notebaert et al., 2009). The relationship between this adaptive behavior and conscious error detection has received diverging evidence. While some authors noticed a PES only after errors consciously detected by the participant (Endrass et al., 2007; Nieuwenhuis et al., 2001; J. Wessel et al., 2011), suggesting a metacognitive control mechanism, others have failed to observe such relationship (Harsay et al., 2018; Hester et al., 2005; van Gaal et al., 2009). Some studies on the alternative measure of post error adjustments, post error accuracy, similarly did not reveal effects of error awareness (Di Gregorio et al., 2016; Endrass et al., 2012). Conversely, in a modified Go/No-go task, aware errors were more frequently followed by correct inhibitory performance (Dali et al., 2022). A similar sequential effect has been observed following partial errors (Allain et al., 2009), initially suggesting that control mechanisms can occur without conscious access, as they remain poorly detected (Rochet et al., 2014). However, a later study has observed post partial error slowing only after detected partial errors (Ficarella et al., 2019).

Overall, confidence judgments and error signaling have been both largely used to explore individuals' ability to reflect on the accuracy of their choices, and appear essential for the development of adaptive behaviors. Their use in different task contexts revealed some discrepancies even though they might reflect two sides of the same coin and share cerebral substrates (Boldt & Yeung, 2015; Yeung & Summerfield, 2012).

The accuracy of decision-making has traditionally been the central focus of metacognition research, with studies often asking participants to evaluate the correctness of their decisions. However, significantly less research has explored the evaluation of the temporal dynamics of decision-making processes, although accuracy and RTs are commonly analyzed together, and closely related. In addition, RT is not independant of various metacognitive experiences (see below, section 1.1.5). Just as researchers use measures of accuracy and RTs to gain insights into cognitive processes, it is plausible that individuals use these performance metrics to monitor and reflect on their own cognitive processes. By examining both indices from a subjective point of view, we may gain a better understanding of how humans can access their decision processes.

1.1.5. The interplay between RT and metacognitive experiences

RTs are related to accuracy and confidence: speed-accuracy trade-off implies that short RTs are most of the time associated with higher percentage of errors (Bogacz et al., 2010), and RTs are negatively correlated with subjective confidence in certain settings (Pleskac & Busemeyer, 2010; Rosenbaum et al., 2022; Vickers & Packer, 1982). Models of confidence indeed predict a negative relationship with an increase drift rate implying an increase in mean confidence. It also predicts a positive relationship with a confidence that should increase with choice threshold or quantity of evidence accumulated. In fact, in condition where accuracy is emphasized over speed, higher threshold leads to longer response time but also higher confidence because more evidence are accumulated. Some authors suggested that RT can serve as a cue to produce response confidence (Audley, 1960; Pleskac & Busemeyer, 2010; Zylberberg et al., 2012). Kiani et al. (2014) observed that the negative link between RT and confidence was also present for error trials, with a higher confidence for faster errors. They suggested that elapsed decision time informs certainty because it serves as a proxy for task difficulty.

This idea is in accordance with the observation that RT contributes to the feeling of difficulty in a conflict task (Desender et al., 2017), in the same way as response conflict and response repetition. Other authors have been interested in subjective experiences arising while performing conflict tasks such as feelings of difficulty, control, competition or urge-to err (Morsella et al., 2009). Interestingly, all these judgments were stronger in experimental conditions associated with high response incongruency which is also related to longer RT. Authors argue that it is difficult to eradicate the effects of processing speed, fluency or effort on the reported subjective judgments. Using objective measures of RT, visual congruency and motor hesitation with recordings of multiple electromyographic activities in some trials, Questienne, Atas, et al. (2018) were able to assess the contribution of these features to the experience of urge-to-err in a subliminal priming conflict task. They showed that reaction times and motor hesitation with response competition were major determinants of this metacognitive experience. Just as confidence judgment and error monitoring can lead to behavioral adjustments, these different subjective experiences are thought to be important for the deployment of cognitive control (Desender et al., 2014).

Altogether, these studies shows that several subjective experiences are in fact based on multiple cues related to performance, and that RT is often one of these cues. External temporal regularity has also been shown to increase confidence (Stevenson & Carlson, 2020). Anyhow, this would mean that this temporal dimension of performance can be accessed by the participants from a metacognitive point of view. Evidence of metacognitive abilities in estimating temporal intervals is the object of the next chapter.

1.2. Temporal metacognition

"The only reason for time is so that everything doesn't happen at once."
– Albert Einstein

1.2.1. Time perception

Timing is ubiquitous and a crucial component of our behavior. In fact, correctly estimating durations is of great importance in many daily activities such as walking, speaking, driving, playing sport or music. Research on cognitive and cerebral processes underlying duration estimation is largely spread and concerns timescales from subsecond intervals to circadian rhythms (for reviews, see Buhusi and Meck, 2005; Grondin, 2010). Intervals below a second in the human timing literature have received a great deal of emphasis, probably because they are related to fundamental adaptive behavior, such as speech processing, motor coordination, and music perception. Interval timing has been studied through four main experimental paradigms: verbal estimation, comparison tasks, production and reproduction. They allowed to show that humans, as well as non-human animals (Malapani & Fairhurst, 2002; Meck, 2005), are able to estimate durations in various conditions, and that time perception can vary as a function of age (Carrasco et al., 2001), attentional (S. W. Brown & Boltz, 2002; Coull et al., 2004) or emotional (Droit-Volet & Meck, 2007) contexts.

Variations in time estimation can be observed through two types of deviations from the target. First, change in accuracy reflects the difference between the objective and mean perceived duration. When they differ, a bias can be observed towards over or underestimation. Second, timing behavior can vary in precision, which is related to the variance of perceived durations across repeated trials. This trial-to-trial variability in temporal judgments has well-defined properties, namely time-scale invariance, which is a form of Weber's law, where variability increases linearly with the duration of the target interval to be estimated (Gibbon, 1977; Gibbon et al., 1984).

Several types of models have been proposed for time estimation. The first, and still mostly used, derives from the original model proposed by Treisman (1963), which stipulates that a pacemaker is emitting pulses collected by an accumulator. The number of accumulated pulses determines the perceived duration. Within this framework, variations in time perception can originate from either modulation of the clock speed, or the rate of the accumulation. For example, increase in arousal has been associated with an increase in internal clock speed leading to more pulses accumulated during the same time, hence to overstimation. Gibbon et al. (1984) proposed a more elaborated version (Fig. 1.4), where the link between the pacemaker and the accumulator is controlled by a switch, open or closed, and the accumulated pulses are compared to a reference memory. In this model, attention is considered to control the functionality of the switch, allowing more or less pulses to be accumulated (Gamache et al., 2011). Although this type of pacemaker-accumulator models have been critized, and different architectures proposed (Matell & Meck, 2004), more recent versions (Balcı & Simen, 2014, 2016), capitalizing on the success of accumulation-to-bound models

for decision making, have revived this idea, which remains the prevalent one in the domain.

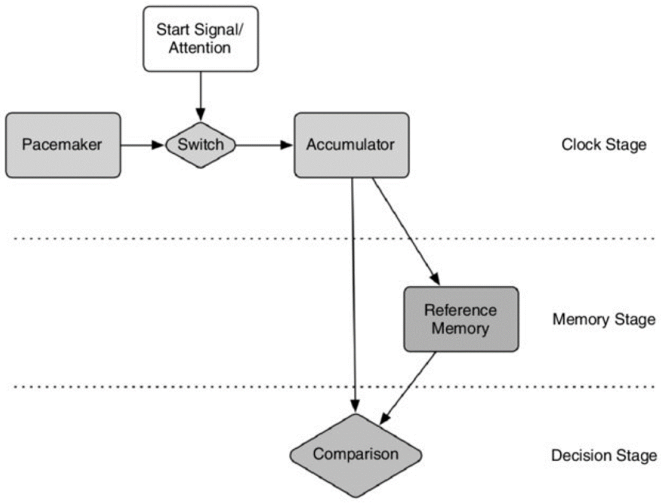


Figure 1.4. *Pacemaker-accumulator model proposed by Gibbon et al., 1984.*

In the brain, timing processes are implemented in a cortico-thalamo-basal ganglia circuit (Merchant et al., 2013). Involved structures include basal ganglia, cerebellum, supplementary motor area (SMA) and PFC (for a review see Coull et al., 2011). In particular, the striatum and dopaminergic system (Coull et al., 2012; Lake & Meck, 2013) appear to have a central role. This is supported by impaired timing function in Parkinson's disease (Koch et al., 2008; J. G. Smith et al., 2007). Neuroimaging studies consistently identified activations of the striatum during timing tasks (Coull et al., 2008; Ferrandez et al., 2003; Pouthas et al., 2005). Striatum activity has been associated with the speed of the internal clock, potentially acting as a pacemaker. On another hand, several evidence suggests that SMA serves as an accumulator of temporal information (for a review see Casini and Vidal, 2011).

Overall, time perception has largely been explored, from both neuropsychological and neurophysiological perspectives. However, whether human subjects are aware of their timing abilities has barely been addressed.

1.2.2. Confidence in time estimation

First studies investigating metacognitive capacities in the time domain used comparison tasks. Methods of comparison, including bisection and temporal generalization, are comparable to the ones used in psychophysics, and involves comparing pairs of intervals or intervals with standard duration(s). This comparison mechanism involves a decision system to be able to choose between same or different, or longer or shorter, when comparing two durations, which can be assimilated to the process involved in 2AFC used in perceptual decision-making. Some authors proposed a two stages diffusion process to account for temporal categorization (Balcı & Simen, 2014, 2016).

In this model, accumulated clock signals during the stimulus constitute first-stage evidence at the basis of a second decision process based on the drift diffusion model.

In this context, very few studies have looked at confidence in temporal decision, in opposition to the vast majority of confidence researches in perceptual decision-making (see section 1.1.4). Some authors have observed that confidence ratings were closely related to accuracy in judging the temporal order of sensory events, with an increase of both as the interval between the two events increased (Allan, 1975; Keane et al., 2015). A similar observation has been made in a temporal comparison task with an 'I don't know' response possibility. The proportion of this response was higher near the mean of the short and long durations, which is the hardest to discriminate (Droit-Volet & Izaute, 2009). A similar finding was observed in rats, which showed a higher number of opt out choices for more difficult comparison when having to choose between two levers associated with two durations, and made more errors when forced to perform the trial (Foote & Crystal, 2007). While partially inconsistent results were observed in a more controlled paradigm with a 'repeat trial' option (Foote & Crystal, 2012), it appears that rats also have some level of metacognitive access to their timing abilities.

A high metacognitive sensitivity to time was later revealed in a temporal generalization task, with a positive relationship between accuracy and confidence judgment, through durations from a few hundred of milliseconds to several seconds (Lamotte et al., 2017). These latter authors also found evidence for an hard-easy effect (overconfidence in difficult conditions and under confidence in easy conditions, Lichtenstein and Fischhoff, 1977; Schneider, 1995), with an overconfidence in temporal judgments of durations slightly longer than the target and underconfidence for comparison durations that were easier to discriminate, i.e much shorter than the target. This result highlights the presence of metacognitive biases concerning temporal decision.

On another side, one of the questions raised by temporal metacognition is whether participants are aware of biases occurring at the level of their temporal performance. One recent study has shown that participants were unaware of the temporal distortion induced by different speeds of a walking stickman, using a bisection task with confidence reports (Öztel & Balcı, 2020). On the contrary, Brocas et al. (2018) have found that participants were aware of biases induced by different reward conditions, and that they could use this information to adjust their temporal decisions. While these contrasting findings indicate that our awareness of temporal biases may depend on the specific context and type of bias, the latter is in accordance with the observation that the declarative knowledge of emotion-related temporal bias decreases the emotional effect on time perception (Droit-Volet et al., 2015). Actually, beliefs about temporal distortions can also be used as a cue for other judgments. For example, based on the famous statement that 'time flies when you're having fun', humans tend to report that they enjoyed an experience more when time subjectively passed faster (Sackett et al., 2010). However, in these types of tasks, differences in durations are experimentally induced and are mainly driven by the stimuli. On another side, the ability to monitor endogenous timing variability in durations that have just been produced has started to be addressed more recently.

1.2.3. Temporal Error monitoring

The two types of temporal errors evoked in section 1.2.1 can also be observed on (re)produced times, in experimental tasks where participants have to (re)produce a target interval. Their distribution shows a certain variability around the mean and this mean can differ to some extent from the target duration, leading to under or over-productions. While the latter biases are most likely not accessible to the participants, otherwise they would be corrected, the question remains for trial-to-trial variability.

Some evidence suggests that humans, and rats, can optimize their temporal decisions based on their level of timing uncertainty (Akdoğan & Balcı, 2016; Balcı et al., 2011; Çavdaroğlu et al., 2014; Jazayeri & Shadlen, 2010; Simen et al., 2011). For example, in the beat the clock task where participants have to emit a timed response as close as possible but not later than a deadline, individuals were able to optimize their behavior based on their level of timing uncertainty (Simen et al., 2011). Similarly, in a task where a reward could appear either with a short latency at one location or with a long latency at another location, humans and mice were accounting for the scalar variability in their estimates to have an optimal switch latency between the short and the long latency locations (Balci et al., 2009). While this suggests that endogenous timing variability is available at some mechanistic level, it does not mean that the subjects are aware of these temporal errors.

More recently, scholars have started to focus on this ability to monitor temporal errors, after the pioneering study of Akdoğan and Balcı (2017). They asked participants to reproduce target durations and subsequently judge i) whether their reproduction was shorter or longer than the target and ii) their confidence in the precision of the reproduction, in each trial. Results showed for the first time that participants were aware of the direction and magnitude of their temporal errors, leading to the field of Temporal Error Monitoring (TEM). The authors proposed a two-stages model (using similar mechanism than their decision model of timing, see Balcı and Simen, 2016). In a first stage, two independent clocks accumulate temporal evidence, until one reaches the threshold, initiating a comparison of threshold crossing times to determine timing error. Upon the slower process's threshold crossing, a second-stage accumulation starts from the earlier crossing point. Longer accumulation in this stage leads to lower confidence, as indicated by greater vertical deviation from the starting point (see Fig. 1.5).

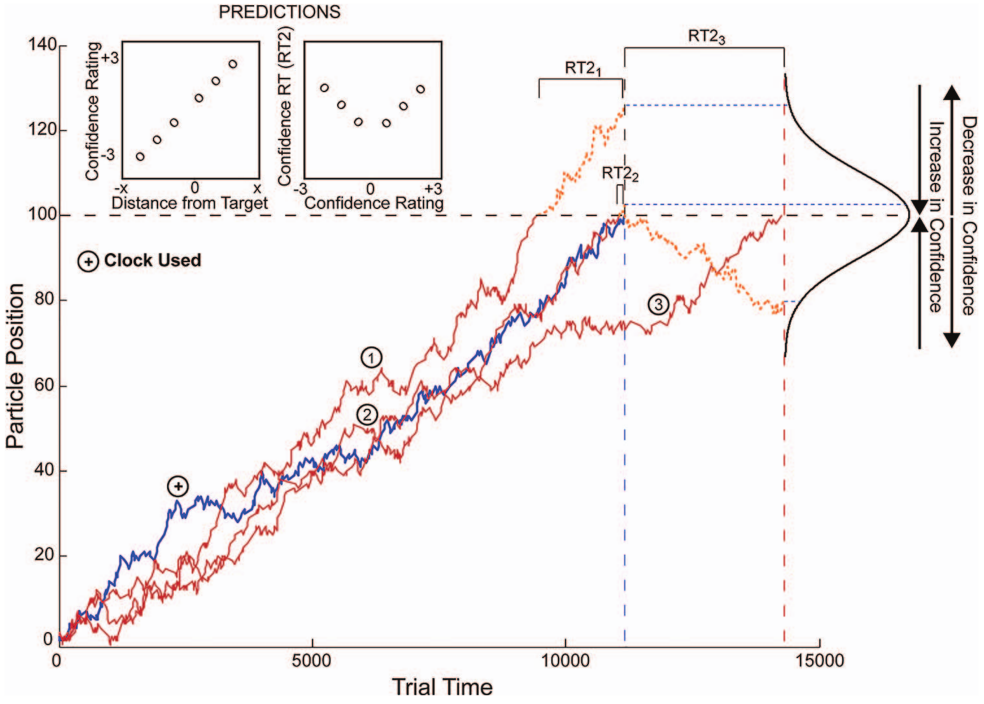


Figure 1.5. *A schematic presentation of the two-stage temporal error monitoring process. In the first-stage, two independent clocks accumulate temporal evidence, whose threshold first crossing times are then compared with each other to determine the direction of timing error. Upon the threshold crossing of the slower timing process, a second-stage evidence accumulation process starts from the earlier threshold (as illustrated with a horizontal dashed line) crossing point, which eventually leads participants to commit to a confidence judgment, with prolonged accumulation processes resulting in lower certainty (due to larger vertical deviation from starting point). Numbers in circles indicate distinct diffusion processes (shown in red), each used as a comparison clock in different trials. The dotted orange lines depict separate second-level conventional diffusion processes, which are in turn used for confidence ratings. Qualitative model predictions regarding the relationship between confidence ratings and timing error as well as the RTs associated with confidence ratings (RT2) are illustrated on the upper left-hand side of the figure. In the words of Akdoğan and Balcı, 2017.*

TEM has been extended to metric error monitoring after similar findings in numerosity (Duyan & Balcı, 2018, 2019) and line length judgments (Duyan & Balcı, 2020). Metacognitive sensitivity to these different metric domains seems to be associated with each other but not with monitoring abilities in perceptual decision tasks in which the error is binary (Yallak & Balcı, 2021).

The results observed in the temporal domain have been replicated several times using direction and confidence judgments (Doenyas et al., 2019; Öztel & Balci, 2022, 2023), judgments on scales ranging from too short to too long (Kononowicz, Roger, & van Wassenhove, 2019; Kononowicz & van Wassenhove, 2019; Öztel et al., 2021) and opt-out option (Yallak & Balcı, 2022), making it a robust phenomenon. Further studies have evidenced that this ability was not affected by the social influence of

being watched by others (Öztel et al., 2021) nor by the working memory load (Öztel & Balci, 2023), but that ownership of errors was necessary for accurate magnitude estimations (Öztel & Balci, 2022).

TEM ability was also observed in rats, in a paradigm in which rodents had to produce a duration by pressing a lever and then estimate their error by choosing between two locations associated with different rewards for different error magnitudes (Kononowicz et al., 2022). High-choice accuracy demonstrated that rats kept track of self-generated durations. Furthermore, the association of this accuracy and size of the temporal production correction between consecutive trials suggested that rats could use estimated errors to guide future behavior. Actually, in humans, it was shown that temporal reproduction was more precise and accurate during re-do trials (Bader & Wiener, 2021), but that non-directional feedback was necessary to increase the precision of the responses, not for the improvement in accuracy. Neuroimaging data evidenced an engagement of both the Default Mode Network (DMN), including the superior frontal gyri, precuneus and posterior cingulate, and the Timing Network, including the SMA, precentral gyrus and right supramarginal gyrus (Bader & Wiener, 2023). Authors hypothesized that DMN may monitor timing performance and transmit information to the motor system when necessary, to execute precisely timed responses. This points to a major role of SMA, involved in monitoring errors (Bonini et al., 2014; Fu et al., 2022, see section 1.3), in monitoring and controlling for timing errors.

One study has looked at the Ne and Pe in this type of task but failed to observe a modulation by temporal errors (Kononowicz & van Wassenhove, 2019). However, the same authors found that the power of β oscillatory activity following the initiation of the time interval was an accurate predictor of produced times, even more so when participants were aware of their time errors (Kononowicz, Roger, & van Wassenhove, 2019). In addition, α power following the response ending the interval was negatively correlated with both produced and estimated durations and was predicted by the initial β power (Kononowicz & van Wassenhove, 2019), in accordance with a process of active metacognitive readout of the first order duration.

Interestingly, one major difference between estimating a sensory interval and a self-produced one is the presence of motor activity in the latter, which certainly adds variability. In fact, when accumulating temporal evidence in order to reproduce a duration, the motor response is only triggered once the threshold is crossed. Whether the time needed to execute the motor response can be accessed when estimating self-produced duration remains unexplored.

The impact of motor activity on timing ability has been mostly investigated with sensorimotor tasks. It has been shown that temporal intervals were more precisely estimated when accompanied by an arm movement (De Kock et al., 2023), yet these estimates can be distorted by various characteristics of the movement such as length (Yon et al., 2017), speed (Yokosaka et al., 2015) or direction (Tomassini & Morrone, 2016). These findings grant insight into the importance of action in the context of timing, and highlights how motor parameters can influence perceptual timing.

In a visuomotor task in which participants had to synchronize a movement with a predictable sensory event, they were shown to be overconfident and underestimated

the magnitude of their motor uncertainty (Mamassian, 2008). In a similar sensorimotor task, humans were also shown to control the timing of their actions by adjusting the time of movement initiation, and based their confidence primarily on this reaction time. However, when explicitly asked to estimate their temporal errors, they were able to take into account both movement onset and movement time information (Jovanovic et al., 2021). This suggests that movement duration can be accessed but is used differently for confidence judgments. Moreover, interval timing can be calibrated by integrating both sensory and motor uncertainties (Schlichting et al., 2023). These results imply that the motor system provides critical information that can be accessed by the system to plan and optimize timing behavior.

However, all these studies have focused on motor activities related to movement in space. The extent to which motor variability associated with isometric key presses, that cannot be cued by visual or proprioceptive information, can be accessed remains an open question.

1.2.4. Reaction times as implicit produced durations

While precise timing is critical and the primary goal in certain everyday tasks, such as punctuality for meetings or anticipating a tennis ball, the timing of many actions is often a by-product of other processes and non-determining to achieve goals. For instance, when my goal is to clean my apartment, the duration of this task can vary due to various factors, yet the goal is accomplished regardless of whether it takes 15 minutes or an hour. In intermediate cases, timing can be crucial but constrained by cognitive processes. For example, a thesis must be done by a certain deadline but is limited by one's scientific and writing capabilities. On a shorter timescale, reacting promptly to environmental stimuli, such as a traffic light turning red, is crucial to avoid potential accidents. This reaction time is determined by the time needed to detect the red light, plan the braking action, and execute it effectively to prevent any adverse consequences, such as colliding with another vehicle crossing the intersection.

These measures of RTs have become of great importance in cognitive psychology and constitute the basis of mental chronometry. It started with the observation that all thoughts take time and that this time can be measured. Just as he measured the speed of nerve conduction by stimulating the sciatic nerve of a frog at different points, Hermann von Helmholtz wanted to measure the propagation speed of stimuli in the sensory nerves of human beings. He described the structure of his experiment: "In a human being, a very weak electric shock is applied to a limited space of skin. When he feels the shock, he is asked to carry out a specific movement with the hand or the teeth, interrupting the time measurement as soon as possible", hence describing a simple RT paradigm. Mental chronometry is rooted in the pioneering work of Franciscus Donders in the 1860s, who formalized the subtractive method, aiming at isolating the duration of mental operations (Donders, 1868). The subtractive method is based on the idea that complex mental tasks can be broken down into a series of simpler cognitive operations. By measuring the RT for tasks that differ by only one cognitive process, the time required for that specific process can be isolated.

To do that, Donders compared three RT tasks, including the simple RT one. In a second task (Go/No-go), participants had to respond to only one of two possible stimuli. Comparing simple RT and Go/No-go would hence allow to measure the decision stage, needed to discriminate both stimuli. In a third task, participants had to produce the specific response assigned to each stimulus, adding a stage of response selection. However, this subtractive method relies on strong assumptions. It first suggests that cognitive processes are strictly serial such as one does not start before the preceding has ended. Additionally, the subtractive method is based on a pure insertion hypothesis, meaning that it can be true only if the addition of a process does not change the others (however, see Vidal et al., 2011).

Development of new methods led to new ways of using RT data to understand cognitive processes. For example, Sternberg's additive factors method (Sternberg, 1984) does not break down cognitive tasks into predefined stages, but looks at the additivity or interactivity of different factors influencing RT. If two factors influence different stages, then their effect on RT should be additive. Analyses of distributions also provide clues for the potential processes underlying the performance of a specific task. For example, the robust observation that the spread of an RT distribution grows with the mean (paralleling time scale invariance, Wagenmakers and Brown, 2007), has been associated with manipulation of the decision difficulty.

Cognitive modeling has become a cornerstone in the study of RTs by providing a structured framework to dissect and understand the underlying cognitive processes. One prominent example, the DDM (Ratcliff and McKoon, 2008), is probably the most used model to make inferences on latent processes in 2AFC tasks. It interprets RT data by modeling the decision-making process as a continuous accumulation of evidence with a certain drift rate until a threshold is reached. The drift rate is associated with the speed at which information is extracted from a stimulus. Hence, the more it is difficult to extract the relevant information from the stimulus, the lower is the drift rate. The threshold is for its part related to response caution. If the goal is to be really accurate, one would increase its response threshold to accumulate a larger amount of evidence. Decision time corresponds to the time taken to reach this threshold. Interestingly, to account for RT data, a residual time must be added to the decision time (=the time during which evidence is accumulated), including an encoding time, and a motor execution time (grouped into non-decision time). Cognitive modeling allows researchers to go beyond simple mean RT comparisons, offering insights into how different factors influence decision making. Additionally, cognitive models can be validated and refined using neuroimaging and neurophysiological data, creating links between behavioral data and neural mechanisms. For example, neurons representing accumulation of evidence have been found in the lateral intraparietal cortex (LIP) of monkeys (Gold & Shadlen, 2007).

Besides modeling, electromyography (EMG) offers a (partial) solution to decompose RT based on objective measures. Its high signal-to-noise ratio allows to record muscle activities linked to the responses of the participants. With these recordings, each RT can be decomposed into two sub-components: the time from the stimulus to the onset of the EMG activity (Premotor time (PMT)) and the time from the EMG

onset to the mechanical response (Motor time (MT)) (Botwinick & Thompson, 1966; Burle, Possamaï, et al., 2002). Studies using this decomposition have shown that MT was undoubtedly related to motor execution processes (Hasbroucq et al., 1999; Tandonnet et al., 2003), and that it was more than a simple constant with its variability contributing significantly to RT duration. While PMT is increased by factors known to impact the decision such as stimulus discriminability, MT is thought to reflect motor execution, as it can be modulated by the force needed to emit a response (Burle, Possamaï, et al., 2002; Weindel, Anders, et al., 2021). Hence, PMT can be associated to encoding time plus decision time and MT to motor execution time of the previously described DDM. However, MT has also been shown to be modulated by perceptual or decisional factors such as stimulus intensity (Servant et al., 2016), speed-accuracy trade-off (Spieser et al., 2017; Weindel, Anders, et al., 2021), but also advance information about the timing (Tandonnet et al., 2003; Tandonnet et al., 2006) or nature (Possamaï et al., 2002) of the stimulus or response to come. These observations suggest that motor processes are not entirely independent of cognitive ones.

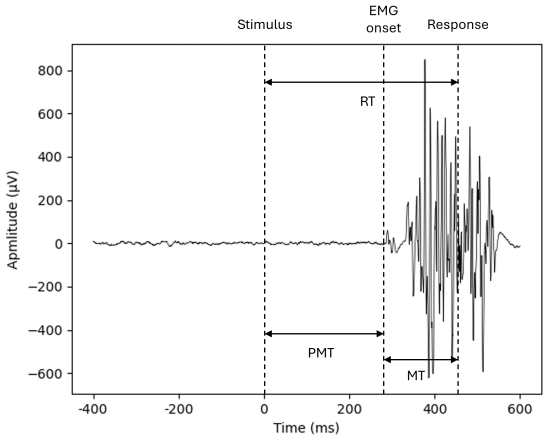


Figure 1.6. *RT fractioning using EMG. The premotor time (PMT) is the time between the stimulus and the EMG onset; the motor time (MT) is the time between the EMG onset and the mechanical response.*

Anyhow, RT measures are fundamental tools for cognitive psychology and neuroscience, and combination of factorial approaches and decomposition methods can help researchers deciphering the organization of mental operations involved in a task and understanding the dynamics of decision-making processes.

RTs have been minimally explored from a metacognitive perspective, unlike the extensive literature on metacognitive judgments regarding decision accuracy (see section 1.1.4). However, the question of whether humans can access the duration of their RTs is intriguing for two reasons. First, just as RT is a measure of the duration of cognitive processes involved in a task, a similar chronometric approach can be applied at the metacognitive level. By examining the awareness of RT duration under different conditions, we can determine which underlying processes are accessible to the subject. Second, RT is a critical component of performance in time-pressured tasks, often emphasized alongside accuracy in task instructions. The speed-accuracy trade-off,

where increasing response speed can lead to more errors, complicates performance assessment. It is challenging to determine who performed better between a fast participant with many errors and an accurate but slower participant, as it depends on the favored dimension. Besides mean speed, RT fluctuates from trial to trial due to variations in attention or fatigue and can be adjusted to optimize behavior, as demonstrated by PES. In RT tasks, participants are usually aware of their errors (see section 1.1.4). Additionally, they can time a wide range of durations and recognize their temporal errors in self-produced durations (see sections 1.2.1 and 1.2.3). Thus, we can question whether participants are aware of their RT when performing choice tasks. The next section delves into introspective estimates of reaction times.

1.2.5. Introspective reaction times

P. Rabbitt (1979) noted that subjects need to know at least three things to control their performance: (1) they must detect errors when they occur; (2) they must be able to monitor and vary response speed so as to recognize, and avoid, faster RT bands at which errors are intolerably frequent; (3) they must be able to adaptively control response speeds so as to select optimal RT bands and avoid others which may be too fast or too slow. Consequently, participants should be able to estimate their RTs. Indeed, in the early 70's, Sanford (1970) showed that participants were able to meaningfully rate their RTs in 4 categories from "fast" to "slow" in a simple RT task with auditory stimuli with two possible intensities. The author observed that the slow category was used more often for quiet signals and concluded that as participants were aware of the intensity effect on RT, it was not an effect on detection time.

More recently, Corallo et al. (2008) developed a new methodology for participants to provide an estimate of their own RT, they called introspective reaction time (iRT). After each choice RT trial, they asked participants to estimate their RT on a visual analog scale (VAS). Note that from the timing point of view, the task is prospective as participants know in advance that they need to estimate an interval, but that the introspective report is retrospective in that it is made after the completion of the task. This enabled the authors to study the relation between RTs and iRTs, aiming to determine which processes contributing to RT are accessible to introspection. In a number comparison task, they manipulated notation and distance and observed the same effect on RT and iRT, leading to a positive correlation between both (see also Bratzke and Hansen, 2024). They concluded that iRT was a sensitive measure, reflecting both perceptual and central processing stages, and showed for the first time that humans had a good representation of their response times.

Interestingly, they set up a second experiment using the psychological refractory period (PRP) paradigm (Sigman & Dehaene, 2005). In this experiment, participants are asked to perform two tasks presented in a quick succession. The main result from this paradigm is that the RT of task 2 is significantly longer when the interval between the presentation of the two stimuli (stimulus onset asynchrony (SOA)) is smaller. This PRP effect demonstrates that the processing of the second task is delayed because the cognitive system is still engaged in processing the first task. It was widely

used to dissociate perceptual and central processing stages, as it is hypothesized that only the central processing stage is delayed at short SOAs (Pashler & Johnston, 1989). Corallo et al. (2008) asked participants to report iRT for both task 1 and 2 and found that iRT of task 2 was blind to the increase of RT2 with shorter SOAs, suggesting that introspection about time spent on a task reflects mostly the duration of the central processing of that task. This absence of PRP effect on iRT gained a great interest and became central in the iRT literature. Marti et al. (2010) reproduced this result and used multiple quantified introspective judgments at each trial, including judgments of iRT of tasks 1 and 2 and SOA, to reconstruct the dual-task phenomenology and test the hypothesis that conscious perception of the second target requires central-stage processing and is therefore delayed during the PRP.

Observation of positive relationships between RT and iRT as well as the blindness to the PRP effect have been reproduced multiple times with other reports methodology such as reproduction (Bryce & Bratzke, 2015), method of constant stimuli (Bratzke & Bryce, 2016), timeline (Bryce & Bratzke, 2017), making it a very robust phenomenon. Although all methods led to similar results, VAS revealed the highest correlations with less variability in the estimates (Bryce & Bratzke, 2015). Some authors challenged the conscious perception bottleneck interpretation of the PRP effect, arguing that introspective limitations were rather due to disturbed timing abilities (Bratzke & Bryce, 2016), and that introspective reports reflect perceived difficulty instead of timing of cognitive processes (Bryce & Bratzke, 2014). iRTs have also been studied in the context of task switching, where RTs are increased after switching from one task to another (switch cost). Contrary to the PRP effect, switch costs were also reflected in iRT (Bratzke & Bryce, 2019, 2022).

Interestingly, the observation that iRT are strongly influenced by the experienced difficulty (Bryce & Bratzke, 2014) points toward a cue-based account of introspective reports, in accordance with the observation that other judgments (confidence, difficulty, urge-to-err, see section 1.1.5) are inferred from different cues. Task sequence, cue-stimulus and response stimulus intervals in task switching were also shown to shift the RT-iRT function, meaning that participants use this information to infer their RT (Bratzke & Bryce, 2019). Interestingly, not all shifts were in accordance with the objective effect on RT, making some cues invalid. Note that it is, at least partly, what can dissociate estimations of reaction times from estimations of time intervals that are entirely defined by external stimuli. Indeed, as RTs are related to mental processes, it is easy to infer them from information thought to change cognitive processes involved in the task. Actually, some authors have used this non-temporal cue argument to explain why RT estimation does not interfere with a concurrent non-temporal task (M. D. Klein & Stolz, 2018), as it is the case with temporal estimation tasks (S. W. Brown, 2006), and why memory search does not interfere with RT estimation (Bratzke & Bryce, 2022), as it does with temporal production (Fortin et al., 2010). In fact, estimating a self-produced interval is already a dual task as you have to perform the task while estimating how much time it takes. Hence, it might be easier to make the retrospective judgment based on different cues.

Another difference between RT and externally defined duration is that the first is

ended by a motor activity. Just as for TEM, not much is known about if and how it can be taken into account for the temporal metacognitive judgment. TEM and RT estimation nevertheless differ in their first order mechanism, because temporal production already involves an explicit timing process.

Participants appear to have an awareness of the time it takes them to perform a task. However, most previous studies have investigated this awareness primarily within the context of dual-task paradigms, which often involves complex methods to infer which cognitive processes are accessible to introspection. To gain a clearer understanding of the mental processes that participants can introspect, it would be valuable to explore iRT within a single-task context. By examining iRT as a function of various factors and employing RT decomposition methods, we can better address which cognitive processes are accessible for introspection in simpler task settings. Moreover, dual-task paradigms inherently demand high levels of attentional resources, which may force participants to rely on non-temporal cues to perform the tasks. Whether the same types of information are utilized in single-task settings remains an open question. This thesis' first experimental chapter will focus on these two aspects (chapter 2).

As stated in the previous section, subjective access to RT provides a means for the experimenter to discern which mental processes are accessible to the participant's metacognitive system. It also allows participants to introspectively evaluate this aspect of their performance. While the neural bases of both temporal estimation (see Section 1.2.1) and performance monitoring are well-established, research on performance monitoring has predominantly focused on accuracy monitoring (see section 1.3). The question of whether the performance monitoring system also supports the monitoring of timing behaviors, particularly iRT, forms the central focus of the second part of this thesis.

1.3. A generic system of performance monitoring

"The brain is a world consisting of a number of unexplored continents and great stretches of unknown territory."
– Santiago Ramon y Cajal

1.3.1. Performance monitoring in the brain

The identification of a performance monitoring system in the brain started with the discovery of error-related activities, notably the Error Negativity (see section 1.1.4.2), which can be observed in both humans (Falkenstein et al., 1991; Gehring et al., 1993) and monkeys (Godlove et al., 2011) EEG recordings. This performance monitoring activity appears to be rather generic as it is independent of the modality of the stimulus, whether visual, auditory or tactile (Falkenstein et al., 1991), and independent of the effector giving the response, whether manual (Falkenstein et al., 1991; Gehring et al., 1993), movement of the eyes (Van 't Ent & Apkarian, 1999), of the foot (Gehring & Fencsik, 2001) or vocal (Masaki et al., 2001; Riès et al., 2013). In addition, it can be observed in a large diversity of experimental tasks, including Go/No-go (Van 't Ent & Apkarian, 1999). Hence, the Ne seems to be generated by errors related to the decision about which action to take. Indeed, the initial account of the functional significance of error-monitoring activities proposes that the Ne is the result of a comparison between the overt response and the representation of the correct response (M. G. Coles et al., 2001; Falkenstein et al., 1991; Gehring et al., 1993).

Main focus of research on performance monitoring has been on how the brain monitors errors and implements subsequent adaptative behaviors (for a review see Ullsperger et al., 2014). The discovery of the Ne launched a multitude of studies and evidence from neuroimaging, EEG, and intracerebral recordings in humans and non-human primates converges on the currently widely accepted view that this performance monitoring function is implemented by a cortico-subcortical network with a central role of the medial prefrontal cortex (mPFC), including the ACC, the pre-supplementary motor area (pre-SMA), and the dorso-medial prefrontal cortex (dmPFC) (see Fig. 1.7).

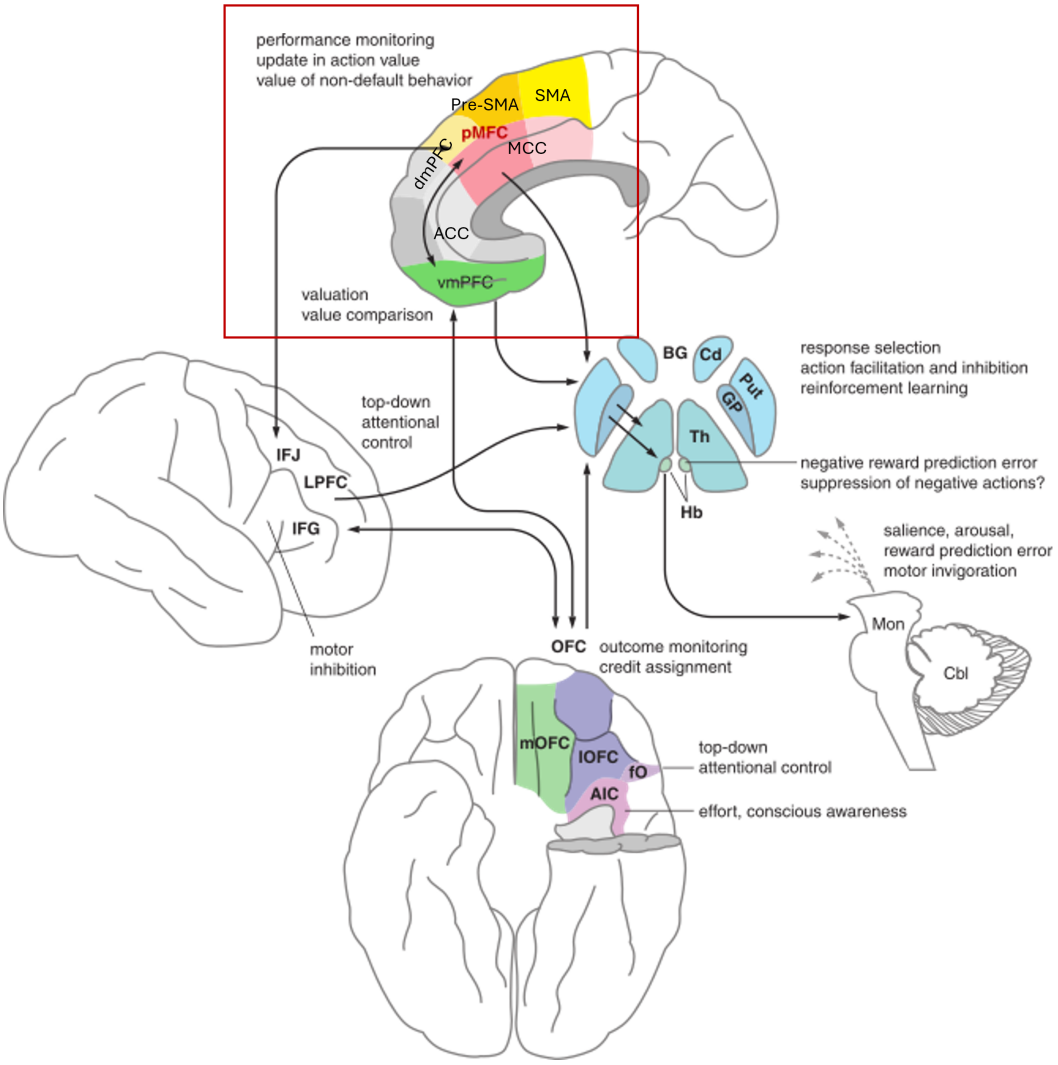


Figure 1.7. *Important brain structures involved in performance monitoring and adaptive behavior, their function, and interaction. pMFC, posterior medial frontal cortex; SMA, supplementary motor area; MCC, midcingulate cortex; ACC, anterior cingulate cortex; dmPFC, dorsomedial prefrontal cortex; vmPFC, ventromedial prefrontal cortex; LPFC, lateral prefrontal cortex; IFJ, inferior frontal junction; IFG, inferior frontal gyrus; OFC, orbitofrontal cortex; m/lOFC, medial/lateral OFC; AIC, anterior insular cortex; fO, frontal operculum; BG, basal ganglia; Cd, caudate nucleus; Put, putamen; GP, globus pallidus; Th, thalamus; Hb, habenula; Mon, mesencephalic and pontine monoaminergic nuclei; Cbl, cerebellum. Main structures evoked in this section are outlined in the red rectangle. Adapted from Ullsperger et al. (2014).*

A central role in performance monitoring has especially been granted to ACC. A first set of evidence comes from intracerebral recordings in both humans and monkeys. Intracranial local field potentials (LFPs) recordings in human epileptic patients revealed a biphasic event-related-potential (ERP) component resembling scalp Ne/Pe potentials after incorrect trials in a visual oddball paradigm, observed in several cortical locations including the ACC (Brázdil et al., 2002). Error-related LFPs were also observed after false alarms in Go/No-go tasks, in the ACC (Brázdil et al., 2005; Pourtois

et al., 2010), as well as in the pre-SMA and in parts of the medial frontal gyrus adjacent to the ACC (Brázdil et al., 2005). Error-related LFPs were also recorded in ACC of monkeys (Emeric et al., 2008; Gemba et al., 1986), paralleling the activity observed at the scalp level in EEG. Furthermore, intracerebral recordings in monkeys ACC have revealed a subset of neurons showing an increased firing activity after errors in various behavioral tasks (Isomura et al., 2003; Ito et al., 2003; Nakamura et al., 2005; Niki & Watanabe, 1979).

A second theoretical line was developped after the observation of activity in the human ACC, measured with fMRI, during erroneous responses as well as during correct responses with high response competition in a modified AX-CPT task (Carter et al., 1998). This led to the conflict monitoring hypothesis (M. M. Botvinick et al., 2001). This model proposes that cognitive control is recruited based on the level of conflict in information processing, conflict being defined as the coactivation of mutually incompatible response processes, and that the Ne originates from this conflict detection (Yeung et al., 2004). Multiple neuroimaging studies have shown activity in ACC in situations of high response conflict (Barch et al., 2000; M. Botvinick et al., 1999; Carter et al., 2000; Lütcke & Frahm, 2008; van Veen et al., 2001), and electrophysiological recordings in patients undergoing cingulotomy revealed a group of cells responding preferentially to high conflict tasks (Davis et al., 2005). In addition, lesion to the rostral ACC impaired regulation of behavior following conflict trials in a variant of the Simon task (Di Pellegrino et al., 2007). However, invasive studies in monkeys failed to observe neurons coding for response conflict in this region (Ito et al., 2003; Nakamura et al., 2005). A case study showed that a focal lesion in the dorsal ACC region induced a reduced Ne amplitude and deficits in error correction, but preserved conflict detection (Swick & Turken, 2002). In fact, in a research combining simulation and experimental studies, Burle et al. (2008) showed that while conflict increased with temporal overlap between response activations, the amplitude of the Ne decreased, hence invalidating the conflict account of the Ne.

Interestingly, ACC neurons increasing their firing rate after errors also increase their activity after omission of reinforcement on correct trials (Emeric et al., 2008; Ito et al., 2003; Niki & Watanabe, 1979). Ito et al. (2003) found a different pool of neurons signaling earned or unexpected reinforcement. Indeed, a negative scalp potential resembling the Ne is observed after feedback about task performance (Miltner et al., 1997), with a higher amplitude for negative feedback (Gehring and Willoughby, 2002; Holroyd and Coles, 2002; Miltner et al., 1997; Yeung et al., 2005, for reviews see Holroyd et al., 2004; Nieuwenhuis et al., 2004). Miltner et al. (1997) first reported that this Feedback-Related Negativity (FRN) was elicited after feedback indicating inaccurate performance in a time estimation task. Source localization studies situated the generator of this activity in medial prefrontal regions, including ACC (Luu et al., 2003). More importantly, the location of this FRN could be predicted from indivividual morphological features of the cingulate/paracingulate cortex (Amiez et al., 2013).

The fact that the Ne and the FRN are generated by the same or at least highly overlapping regions of the mPFC led to the assumption that both activities are functionally equivalent (Holroyd & Coles, 2002; Miltner et al., 1997). Holroyd and Coles (2002)

developed a reinforcement learning theory in which they argued that the Ne/FRN is generated when an action outcome is worse than expected. Through learning, FRN is thought to decrease and Ne to increase during errors as action-outcomes become more predictable (Krigolson et al., 2009). This follows the observation that after learning, phasic dopaminergic response recorded in monkeys midbrain neurons is no longer elicited by the reward but by the conditioned stimulus (Ljungberg et al., 1992; Schultz et al., 1993). Following this idea, Ne/FRN are thought to reflect a decrease in dopaminergic signals coming from the basal ganglia in case of negative outcomes, leading to a disinhibition of the pyramidal neurons of the ACC. The implication of subcortical regions, especially basal ganglia, in performance monitoring is evidenced by a disrupted Ne in patients with focal striatal lesions (Ullsperger & Von Cramon, 2006), as well as in Parkinson's disease patients (Falkenstein et al., 2001; Stemmer et al., 2007). ACC activity after erroneous action would hence reflect a reward prediction error signal. In a stimulus-reward association task, error-related unitary extracellular activity in the dorsal anterior cingulate cortex (dACC) of monkey was modulated by the amount of reward or proximity to reward delivery (Amiez et al., 2005) and might represent a negative deviation from a predicted goal. Other authors found separate groups of neurons responding to positive and negative feedback in the mPFC and bigger responses in learning blocks in which feedback was less predictable (Matsumoto et al., 2007), suggesting that the mPFC cells could contribute to behavioral adaptation by representing the direction and amount of error in action value prediction.

Other theories have built upon similar principles of reinforcement learning. J. W. Brown and Braver (2005) developed the error-likelihood model which posits that the dACC computes the likelihood of making an error in a given context, and that the level of this signal is used to assess the need to recruit cognitive control. In this framework, response conflict is only a special condition in which error likelihood is high.

While previous models can account for some but not all effects observed on ACC activity, later ones attempted to develop a unified view, still based on reinforcement learning principles. The Predicted-Response Outcome (PRO) model (Alexander and Brown, 2011) suggests that ACC activity might not depend on the valence of events, but only on their expectedness (see Fig.1.8). In this model, mPFC activity is assumed to code the learned prediction of the probability of possible action outcomes at hand. When an expected outcome occurs, the corresponding prediction signal is inhibited. Consequently, non-occurrence of predicted outcomes results in maximal mPFC activity. Under this simple computational principle, the PRO model is able to simulate a wide variety of empirical findings, including sensitivity of ACC to conflict, errors, and reward prediction errors (Alexander & Brown, 2014). The Reward Value and Prediction Model (RVPM), independently proposed by Silvetti et al. (2011), shares properties with the PRO model and implements a prediction mechanism of the value of future outcomes in ACC, when reward is at stake. This model has also received empirical support (Silvetti et al., 2014). With these accounts, Ne would reflect unsigned prediction error, i.e surprise, which is large in case of unexpected events, whether positive or negative. The observed error effects are hence most of the time confounded with unexpectedness, as most studies use tasks in which errors are rare events. This idea

is supported by fMRI data in humans where, in a gambling task, the error effect is reversed when losses are more likely than gains (Jessup et al., 2010). In a similar line of idea, introducing a response probability bias in a choice RT task led to a decrease in Ne amplitude for unexpected conditions, in which errors are more predictable (Meckler et al., 2011).

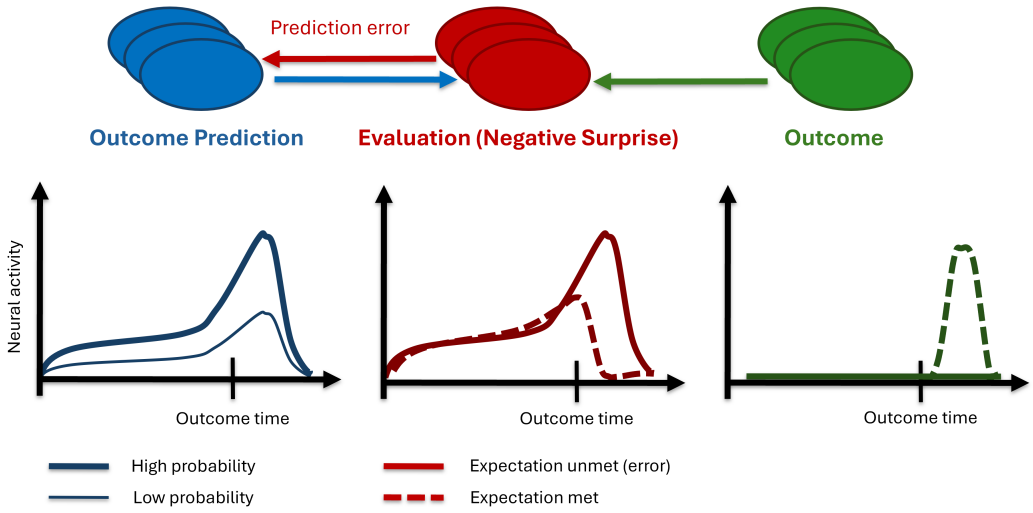


Figure 1.8. *The Predicted Response Outcome (PRO) model. The prediction units (in blue) generate a timed prediction of what outcomes are expected, with what probability, and when. Greater probabilities are associated with greater prediction activity, and the activity peaks at the time the outcome is expected to occur. The evaluation units (in red) compute negative surprise—that is, they detect when an expected outcome fails to occur. Negative surprise is computed simply by subtracting the actual outcome signal from the prediction-unit signal at each point in time. Events that are predicted yield greater surprise signals when they fail to occur (solid line) than when they do occur as expected (dashed line). Outcome units (in green) signal when an actual outcome occurs. An outcome unit is transiently activated at the moment a corresponding outcome actually occurs (dashed line); if no outcome occurs, the outcome unit remains inactive (solid line). Adapted from J. W. Brown (2013).*

Altogether, the link between error-related activities and ACC, as well as the different computational models, have put the emphasis on a central role of the ACC in error monitoring. However, intracerebral recordings have also revealed error-related activities in other structures of the mPFC. In a saccade countermanding task, Stuphorn et al. (2000) recorded different groups of neurons activated after errors, successful witholding of movement, or with reinforcement in the supplementary eye field (SEF) of macaque monkeys (see also Sajad et al., 2019; Schall et al., 2002). Error-related LFPs were also observed in SEF of monkeys in saccade countermanding task (Emeric et al., 2010), and visual search task (Purcell et al., 2012). SEF is involved in the control of eye movements and is thought of having a homologous function as SMA for general motor control. In fact, in an arm countermanding task, Scangos et al. (2013) found highly similar neuronal activity in SMA and pre-SMA coding for expected outcome, the actual

outcome, and the mismatch between actual and intended outcome. More recently, Fu et al. (2019) recorded LFPs and unit activities in the pre-SMA and the ACC in humans performing in a manual Stroop task. The authors identified error-sensitive neurons and error-related LFPs in both structures, in accordance with previous observation of higher LFP activity for errors in human SMA (Bonini et al., 2014).

Overall, these studies suggest that both SMA (or SEF for eye movements) and ACC are likely generators of the Ne. However, the precise computational mechanisms underlying this activity remain a topic of debate. Despite this uncertainty, there is consensus that the Ne serves as a reliable index of performance monitoring, that could extend beyond errors. In fact, while this activity may not exclusively reflect ACC function, it may also not be specific to errors (see below).

1.3.2. Performance monitoring beyond errors

While the literature emphasizes this particular situation, the Ne might not be specific to errors. Indeed, activities with the same polarity and latency are observed both after partial errors and correct responses (Vidal et al., 2000), although with smaller amplitudes. The amplitudes of the Ne observed in partial errors lies between the one of errors and of correct trials, much smaller. The negativity in correct trials (Correct Negativity, Nc, Falkenstein, 2004 or Correct-Related-Negativity, CRN, Ford, 1999) was initially revealed by Laplacian transform of the EEG data, which acts as a high-pass spatial filter (Nunez et al., 1994). Its existence was first contested, as many studies failed to observe it in monopolar data, which is probably due to large recovering of different signals (Luu & Tucker, 2001). While its existence is no longer questioned, its functional significance, and whether it has the same status as Ne, remain debated.

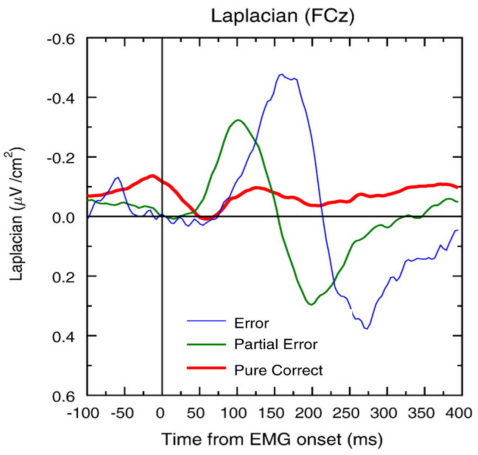


Figure 1.9. *Laplacian-transformed grand-averages of the Ne for correct trials (red), partial errors (green) and errors (blue). From Roger et al. (2010).*

In fact, this component challenges the majority of existing models of performance monitoring, except if we consider that it is a different component than the Ne (see Vidal et al., 2022). Ne and Correct Negativity (Nc) still share several properties. Actually, Nc is also modulated by performance in some specific conditions. Notably, the Nc of a trial preceding an error has a smaller amplitude than the Nc of a trial preceding another correct trial (Allain et al., 2004), suggesting that it also has an important role in action monitoring. In addition, Ne and Nc have comparable amplitudes in schizophrenic (Alain et al., 2002; Mathalon et al., 2002) and prefrontal patients (Gehring & Knight, 2000). A disappearance of the difference in amplitude between error and correct could also be observed in healthy patients in the unexpected condition of a task with a response probability bias (Meckler et al., 2011). These behavioral cues and the gradual sensitivity to partial and overt errors point to a common performance monitoring function of the Ne and Nc. If so, they should also share a common cerebral source.

The scalp distributions of the Ne and the Nc are different in monopolar data (Vocat et al., 2008), which could be taken as evidence that both activities have different brain generators. However, this is not warranted because large parietal positive component(s) overlap(s) with the fronto-central negativity, and more easily when the latter is smaller. Indeed, simulations with a medial frontal negative dipole and a positive parietal dipole have shown that decreasing the amplitude of the first one while keeping the second constant modified the scalp distribution (Vidal et al., 2015). Indeed, the topographies become highly similar after Laplacian transform (Roger et al., 2010; Vidal et al., 2000), strongly attenuating overlapping effects (Vidal et al., 2015). Similar topographies are still not strong evidence for similar neural generators, as the inverse problem indicates that one scalp distribution can be explained by an unlimited number of sources configurations. Using two different localization methods in EGG, Roger et al. (2010) found a same generator for Nc and Ne (in errors and partial errors) in the rostral cingulate zone (RCZ), extending to the pre-SMA/SMA.

As for the Ne, intracerebral data bring valuable insights into the sources of potentials recorded at the scalp level. Some authors who observed error-related LFPs in monkeys SEF with tasks involving saccadic movements also noticed smaller activity following correct saccades in a latency range corresponding to that of the Nc (Emeric et al., 2010). In line with this, spiking activity of error-sensitive neurons of SEF, lower on correct saccades predicted the amplitude of the scalp Nc (Sajad et al., 2019). This indicates that SEF plays a role in performance monitoring of both correct and error trials. Studies using intracerebral recordings in humans SMA with task requiring hand-responses revealed findings that closely parallel what was observed in SEF for saccadic responses. Bonini et al. (2014), in addition of LFP having a high amplitude after error responses, clearly observed a similar activity on correct responses with a much smaller amplitude. LFPs exhibited an intermediate amplitude in case of partial errors. Interestingly, they found error-related LFPs in ACC, delayed compared to the ones of SMA, but no correct-related activity in this structure. This aligns with more recent results showing error-related spiking activity and LFPs in both pre-SMA and ACC, with the latter structure displaying a delay of approximately 50 ms (Fu et al., 2019). A similar delay between error activities in SEF and ACC was also observed

in monkeys (Emeric et al., 2010). However, Fu et al. (2019) did not examine correct-related activities, but some reported results from the ACC did not show any sign of such activity.

These differences in latencies and the fact that no correct-related activity seems to be present in ACC suggest a primary role of SMA/pre-SMA in performance monitoring. This structure might initially evaluate performance gradually with its output being relayed to more rostral structures like the ACC. A threshold mechanism between these regions could lead to a secondary all-or-none error response in the ACC (Bonini et al., 2014). This also implies that Ne and Nc reflect, at least partially, the same process implemented by the SMA. Consequently, in this thesis, I will consider that both Ne and Nc are the same component and will refer to Ne for all these negativities evoked by the responses.

1.3.3. Subjective performance monitoring?

Majority of Ne studies have been done in tasks where participants are not explicitly asked to monitor their performance, but certainly detect most of their errors (P. M. Rabbitt, 1966, see section 1.1.4.2). A part of the literature on error monitoring specifically addressed the question of error awareness, and of its specific relation with error-related EEG activities. To this aim, participants are asked to report whether they think they made a mistake on the current trial, allowing to dissociate 'aware' from 'unaware' errors.

Interestingly, one group has investigated whether the implementation of a response accuracy rating itself influences error processing (Grützmann et al., 2014) and found higher Ne amplitude in rating condition for both error and correct trials. This result can be interpreted as a general increase in performance monitoring with more attention being given to response accuracy when it has to be rated. This is somehow in line with observations of increased Ne amplitude with enhanced error salience (Endrass et al., 2010; Hajcak et al., 2005; Riesel et al., 2012). Similarly, the Ne of errors is larger when accuracy is emphasized over speed (Falkenstein et al., 2000; Gehring et al., 1993). While similar effects on Ne of correct trials was far less explored, Grützmann et al. (2014)'s result reproduces what is observed in obsessive-compulsive patients, with larger negativities in both type of trials (Endrass et al., 2008, 2010). These findings are in agreement with the view of a generally overactive performance monitoring system, with greater attention on the correctness of the responses. Hence, asking participants to judge their performance can elicit a similar enhancement of performance monitoring.

If error significance seems to influence the Ne (M. E. Maier & Steinhauser, 2016), its relationship with error awareness remains controversial. The first study which explicitly looked at the Ne in relation to subjective judgments of accuracy found an increasing amplitude from 'sure correct' to 'sure incorrect', with the highest amplitudes for errors perceived as errors (Scheffers & Coles, 2000). One year later, Nieuwenhuis et al. (2001), found no difference in Ne amplitude between aware and unaware errors in an antisaccade task with an error awareness button. These two contradicting results

were followed by a multitude of other studies backing-up either view. Indeed, an important number of studies have found no relationship between Ne amplitude and error awareness (Endrass et al., 2005, 2007; Hoonakker et al., 2016; Niessen et al., 2017; Nieuwenhuis et al., 2001; R. G. O'Connell et al., 2007; Shalgi et al., 2009). On another hand, some other studies did find the Ne to be dependent on aware error perception, supporting Scheffers and Coles (2000) initial result (Dhar et al., 2011; Hewig et al., 2011; Kirschner et al., 2021; M. Maier et al., 2008; Shalgi & Deouell, 2012; Steinhauser & Yeung, 2010; J. Wessel et al., 2011; Woodman, 2010). All of these studies used various tasks including flanker, antisaccade, stop signal and Go/No-go tasks (for a review see J. R. Wessel, 2012). An error awareness task (EAT), a Go/No-go task using stroop-like stimuli, has also been developed to ensure a sufficient number of non-reported errors, tackling an important methodological problem in error awareness studies. Methods of assessing aware errors also differ between studies. Most of them use binary judgment with either a forced choice between error and correct, or error signaling only, where participants must press a button only when they think they made an error. Actually, this latter method is most likely suboptimal, with a bias towards non-reporting errors, possibly contaminating 'unaware errors' with unsure trials. Some other studies have used parametric scales with 'not sure' options (Hewig et al., 2011; Scheffers & Coles, 2000). However, the differences in methods cannot explain the divergence in observed results, as both positive and null findings can come from the same methodology.

Similarly, whereas some neuroimaging studies reported that the medio-frontal cortex (MFC) was unrelated to error awareness (Hester et al., 2005; T. A. Klein et al., 2007), other studies did find greater MFC activation when participants were aware of their errors (Hester et al., 2012; T. A. Klein et al., 2014; Orr & Hester, 2012). Dali et al. (2022) sought to assess the neural network of error awareness in a large fMRI study using the EAT. They found that aware errors were accompanied by significantly greater activity in several regions including the insula, supramarginal gyrus (SMG), and structures known as the generators of the Ne: the ACC and SMA.

Although results concerning the Ne and error awareness is not unequivocal, they do not seem completely unrelated. On the contrary, findings concerning the Pe appear more consistent, with the component being observed only for aware errors (Endrass et al., 2005, 2007; Murphy et al., 2012; Nieuwenhuis et al., 2001; R. G. O'Connell et al., 2007; Overbeek et al., 2005), although a more gradual relationship has been highlighted (Boldt & Yeung, 2015).

Interestingly, one study has looked into 'partial error' awareness in a variant of the Simon task with judgment about partial errors given on a scale from 'sure I did not' to 'sure I did' commit a partial error (Ficarella et al., 2019). The Ne was larger for aware than for unaware partial errors, and a Pe was observed only after the corrective EMG activity of aware partial errors trials, paralleling what can be observed on overt errors.

Altogether, those findings about Ne and Pe relationships with error awareness have led to the idea that the Ne, representing the first signal that an error has occurred, might serve as a feedforward input into a system more responsible for awareness (Murphy et al., 2012; J. R. Wessel, 2012), whereas the Pe would reflect the accumulation of information leading to error awareness (Desender et al., 2021; T. A. Klein et al., 2014;

Steinhauser & Yeung, 2010).

1.3.4. Performance monitoring beyond action selection

The great majority of studies on performance monitoring and Ne have focused on the difference between correct and error responses in terms of action selection. Indeed, studied errors were mainly related to making incorrect choices in 2AFC tasks. This raises two questions. First, choice errors in these tasks are by construction binary, as most of the time only two choices are possible. But as partly evidenced with partial errors, it is possible that the performance monitoring system operates in a more gradual manner. Second, much less is known about the monitoring of other parameters than action selection.

Error monitoring has been, to a lesser extent, studied in the context of visuo-motor tasks. It allows to have different error magnitudes, defined by the spatial deviation between the target and the position of the hand at the end of the pointing movement. In such tasks, Ne amplitude was higher for larger deviations, i.e higher error magnitude (Anguera et al., 2009; MacLean et al., 2015; Vocat et al., 2011). These results indicate that the Ne can encode the degree of mismatch between the actual and expected motor outcome and does not represent error in an all-or-none manner (Vocat et al., 2011).

In 2AFC tasks, without any arm displacement required, motor execution can be measured by the force produced by participants during responses. Few attempts have been made to explore if this execution of the response could be monitored. Some authors have asked participants to perform a force production task, in which they had to produce either a weak or a strong force (Armbrecht et al., 2012; Armbrecht et al., 2013; de Bruijn et al., 2003), to see whether the Ne could be sensitive to errors either in force selection or in force execution. In the work of de Bruijn et al. (2003), a Ne was evoked by selection errors for the weak condition only and was not sensitive to force execution errors. However, they argued that the constant change in force might have obstructed the comparison between the executed force and a clear representation of the required force. Armbrecht et al. (2012) did not find a Ne sensitivity to errors of execution either and concluded that this parameter was not monitored. However, in typical choice RT task, while some trials contain a small EMG activation on the incorrect side (partial errors), some others can also contain a small EMG activation on the correct side, not strong enough to trigger the mechanical response (Meckler et al., 2017). These 'partial correct' activities might correspond to responses for which the required force was not produced correctly, which is supported by the fact that the produced force measured in the response correcting partial corrects is higher than that of 'pure' correct or of responses correcting partial errors (Meckler, 2010). It has been shown that the Ne of these partial correct activations had a larger amplitude than in pure correct trials suggesting that the monitoring system can evaluate execution errors as well as selection errors (Meckler et al., 2017; Ramdani et al., 2024). In accordance with this latter result, ACC error-related LFPs of monkeys were recorded after both execution and selection errors in a stimulus-reward association task (Amiez et al.,

2005), even though execution referred to the breaking of eye fixation in this case.

It is not excluded that the monitoring system could also be sensitive to the temporal dimension of action. Indeed, Ne amplitude was shown to increase with increasingly late responses given after a deadline in a modified flanker task (Luu et al., 2000, see also Stahl, 2010), highlighting a sensitivity to temporal errors. Yet, the above results have been obtained in frameworks in which there is a clear and explicit distinction between correct and erroneous dimensions, although in various and continuous dimensions. Some further research has looked at the relationship between Ne and RTs in correct trials, in typical choice RT tasks without any deadline. In one of the initial papers reporting a Ne in correct trials, Vidal, Burle, et al. (2003) did not observe any differences in amplitude between trials with short RT and long RT. Conversely, more recent studies could observe a modulation of the Ne with RT in conflict tasks (Valt & Stürmer, 2017, 2021). In a first study, Valt and Stürmer (2017) observed a smaller Ne amplitude for slow RT compared to fast and average ones and interpreted this finding in terms of positive emotional arousal. They argued that the optimal nature of fast correct responses might generate positive arousal, as errors generate negative arousal. In line with this result, the same authors later observed a larger Ne for correct fast trials, with an amplitude equal to that of errors, in a flanker task with a monetary punishment for errors and slow responses (Valt & Stürmer, 2021). They concluded that it did not reflect an aversive reaction to inappropriate performance as correct fast reflected optimal responses. They stipulated that this increase in amplitude might come from positive reinforcement, where correct fast responses could represent better than expected outcomes, the same way errors could reflect negative reinforcement. However, in a Go/No-go task with speed feedback, Ne was found to have a higher amplitude for slower responses, even more so when a gain was associated with fast responses (Files et al., 2021). Authors concluded that RT can be monitored when it is motivationally relevant and that the relation between Ne and RT could reflect the degree of subjective motivational relevance. Indeed, these few studies which revealed a link between Ne and RT all gave feedback on RT and punishment or reinforcement based on RT (Files et al., 2021; Valt & Stürmer, 2021), or asked participants to subjectively judge each of their RT into categories from fast to slow (Valt & Stürmer, 2017). These contexts possibly oriented the monitoring system towards RT.

Although the specific nature of the relationship has yielded diverging results, the evidence suggests that the Ne can indeed play a role in monitoring timing performance, at least in certain contexts. Given that SMA is a key player in both performance monitoring and interval timing (Coull et al., 2011; Wiener et al., 2010), it emerges as a strong candidate for the monitoring of temporal performance. Considering the SMA's involvement in these processes, the role of the Ne in overseeing various aspects of action, as well as its potential connection to subjective judgments, this cerebral activity appears relevant for temporal metacognition. The second part of this thesis (chapter 3) is thus dedicated to exploring the contribution of the Ne to the subjective evaluation of RT.

1.4. Outline of the thesis

The present thesis aims to investigate the sources of information underlying iRT judgments and the role of the cerebral performance monitoring system in these metacognitive assessments. The introduction highlighted that most metacognition studies in decision-making have primarily focused on accuracy judgments. Here, we emphasize that response speed is also a valuable index of performance, and, drawing inspiration from the chronometric approach, we argue that its subjective evaluation offers important insights into the cognitive processes accessible to the metacognitive system. Additionally, determining whether introspective RT judgments, like other metacognitive assessments, are based on mechanisms involved in the initial decision-making process or on various external cues would significantly enhance our understanding of metacognition on a larger scale. At the cerebral level, we focused on the medial prefrontal cortex (MPFC) due to its established role in performance monitoring. Identifying a relationship between this system and RT estimation would not only deepen our understanding of the MPFC's functions but also provide critical insights into how metacognitive judgments are implemented in the brain.

These questions are addressed in three experimental chapters based on the results of three experiments. Chapter 2 incorporates results from two experiments: the first was developed to be run online during COVID times, and the second was conducted in the lab. This first experimental chapter evaluates whether introspective reaction times (iRT) directly reflect first-order reaction times (RT) or whether they are inferred from other cues. Additionally, it explores which processes underlying RT can be accessed by participants. Chapter 3 focuses on the potential cerebral basis of iRT judgments, with particular emphasis on the contribution of the MPFC. This chapter is divided into two parts and is based on the results of a third experiment, involving two groups of participants, conducted using EEG. Chapter 4 investigates the global and sequential influences of iRT judgments on first-order behavior, seeking to identify metacognitive control mechanisms that may be established based on iRT. A final methodological chapter presents a new method for online detection of the onset of EMG activity, which was developed with the perspective of a fourth experiment that could not be carried out during the thesis period.

2. Investigating the building blocks of iRT

Summary

2.1. Note on the present chapter 49
2.2. Article: Temporal metacognition: direct readout or mental construct? The case of introspective reaction time 50
2.2.1. Abstract 51
2.2.2. Introduction 51
2.2.3. Experiment 1 54
2.2.3.1. Methods 54
2.2.3.2. Results 55
2.2.3.3. Discussion 58
2.2.4. Experiment 2 58
2.2.4.1. Methods 59
2.2.4.2. Results-Pure correct trials 62
2.2.4.3. Results-Beyond pure correct trials 65
2.2.4.4. Discussion 65
2.2.5. General discussion 67
2.2.5.1. Of which processes are the durations evaluated: contribution of decision and motor execution 67
2.2.5.2. iRT inference from multiple cues 68
2.2.5.3. Why use non-temporal cues? 69
2.2.6. Conclusion 69
2.2.7. References 70

2.1. Note on the present chapter

The first objective of the thesis was to assess the mental processes that contribute to subjective estimates of reaction times (iRT). Previous studies, which used psychological refractory period (PRP) paradigms (see section 1.2.5 of the General Introduction), have concluded that iRT primarily reflects central processing stages (Corallo et al., 2008; Marti et al., 2010). Interestingly, one notable observation from these studies, is that iRT judgments consistently underestimate actual RT (Corallo et al., 2008). Although this seems to be a rather general phenomenon (Sackur, personal communication), this phenomenon lacks a clear explanation.

RTs comprise the duration of various processes leading to a response. Motor execution of this response is typically assumed to start after the decision to act has been made. Individuals accumulate information about the stimulus until a predetermined threshold is reached, signaling the decision point. Once this threshold is crossed, motor execution begins, introducing additional variance into the total RT. It is possible that participants base their iRT judgments solely on the decision-making time. If this is the case, it could explain why iRTs tend to underestimate actual RTs.

This hypothesis can be directly tested by using electromyography (EMG) to separate RTs into premotor times (PMT), associated with decision, and motor times (MT), associated with motor execution. If iRT is indeed based only on the decision time, two predictions can be made: first, iRT should correlate more closely with PMT than with the combined PMT+MT (i.e. RT); second, iRT should be unaffected by experimental manipulations that influence MT.

These assumptions were directly tested in the present chapter, which presents results from two experiments. The first experiment, conducted online during the COVID-19 pandemic, served as a pilot study for the experimental design and allowed for the recruitment of a large number of participants to more accurately assess their ability to estimate their RTs. The second experiment was designed to specifically test the aforementioned hypotheses. By manipulating factors that affect either PMT or MT and measuring these components with EMG recordings, we aimed to determine their contribution to iRT.

Additionally, both experiments provided an opportunity to explore whether other types of information, beyond RT, could influence subjective judgments, testing a cue-based account of iRT (Bryce & Bratzke, 2014) within a simple task framework.

Results of these two experiments are the subject of the first article related to this thesis, currently in revision in Journal of Experimental Psychology: General.

2.2. Article: Temporal metacognition: direct readout or mental construct? The case of introspective reaction time

Journal of Experimental Psychology: General
in revision

©the authors

Temporal metacognition: direct readout or mental construct? The case of introspective reaction time.

Nathalie Pavailler[1], Wim Gevers[2], and Boris Burle[1]
[1]Aix-Marseille Université, CNRS
[2]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 read-out of the duration of the underlying processes as participants may instead rely on inferences based on other salient, non temporal, 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 read-out 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 sub-process 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 read-out can be biased by non-temporal cues. The proposed methodology allows to dissociate the two.

Keywords: Reaction times, Introspection, Performance monitoring, Electromyography.

Public significance statement

Do we have direct access to the content of our mental life? Or do we merely re-construct 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 participant 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.

Introduction

In many aspects of life, understanding the cognitive processes that led to our behavior is essential, especially for learning and academic achievement (Stanton, Sebesta, & Dunlosky, 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 Times (RT) measures, that allowed to reveal the temporal sequencing and relative durations of mental processes (Donders, 1868; Luce, 1986; Meyer, Osman, Irwin, & Yantis, 1988; Posner, 1978; Sternberg, 1969). Recent evidences suggest that such a chronometric approach can also be useful to reveal core aspects of metacognition (Corallo, Sackur, Dehaene, & Sigman, 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, Roger, & van Wassenhove, 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 non-temporal decision tasks, potentially using it to set an optimal response threshold (Balci et al., 2011). Indeed, in the early 70's, Sanford (1970) showed that participants were able to meaningfully rate their RTs in 4 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) reasonned 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, Bryce, & Seifried-Dübon, 2014; Bryce & Bratzke, 2015, 2017; Corallo et al., 2008; Marti, Sackur, Sigman, & Dehaene, 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): 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 underly 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, Frankenhuis, Waldorp, & Borsboom, 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, Van Opstal, & Van den Bussche, 2017; Questienne, Atas, Burle, & Gevers, 2018 for comparable use in different contexts). If one considers iRT as the dependent variable, and both RT and conditions as 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

N.P. and B.B., Centre de Recherche en Psychologie et Neurosciences, CNRS, Aix-Marseille Univ., and W.G., Center for Research in Cognition and Neurosciences, ULB Neurosciences Institute, Université Libre de Bruxelles. N.P. is funded by a PhD grant from the french ministery 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 ANR fundings (AMX-19-IET-004 and ANR-17-EURE-0029), and by the ANR (grant number: ANR-22-CE28-0027). A CC-BY public copyright license has been applied by the authors to the present document and will be applied to all subsequent versions up to the Author Accepted Manuscript arising from this submission, in accordance with the grant's open access conditions. The authors wish to thank D. Louber and D. Paleressompoulé for developping the response device and F.-X. Alario for fruitful discussions and advices. Data and codes used in this study are available at https://osf.io/3rzak/. Parts of the data of this manuscript 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. Correspondence concerning this article should be addressed to Boris Burle, Centre de Recherche en Psychologie et Neurosciences, Case C, Aix-Marseille Université, 3 Place Victor Hugo, 13331 Marseille, cedex 3. E-mail: boris.burle@univ-amu.fr

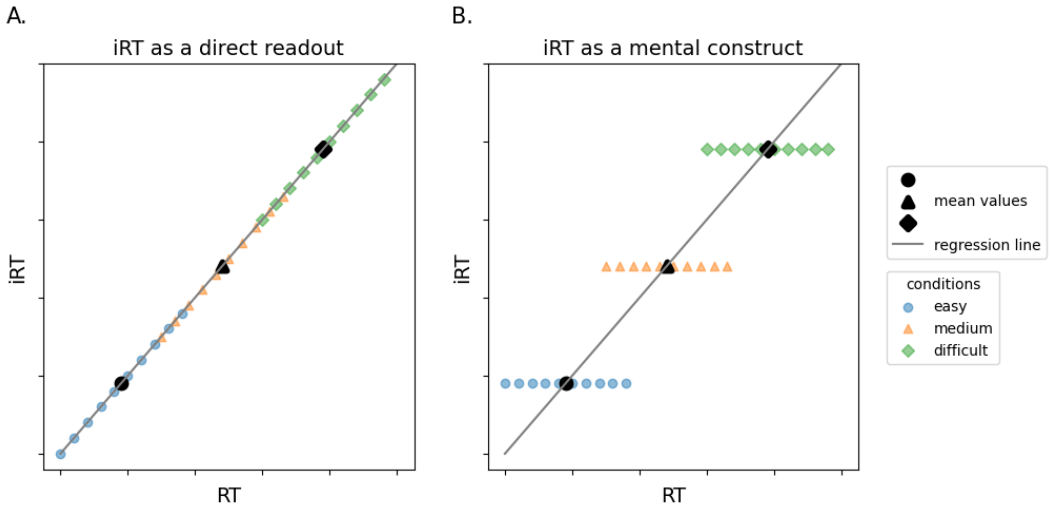


Figure 1

Linear association between iRT and RT. Panel A: iRT is purely based on duration of RT or panel B: iRT is purely inferred from task conditions. Each colored (grey) point represents an individual trial, black points represent the means. If performed across means, correlations are identical between A. and B.

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, 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 non-decision times (Luce, 1986; Ratcliff, 1978; Ratcliff, Smith, Brown, & McKoon, 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 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 non-decisional processes in introspective judgement. Miller, Vieweg, Kruize, and McLea (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 non-decision processes, they inferred them based on unverifiable and unwarranted psychological and physiological assumptions. Note that even outside the context of introspection, non-decision stages have been much less studied. As a matter of fact, non-decision 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, Possamaï, Vidal, Bonnet, & Hasbroucq, 2002; Weindel, Anders, Alario, & Burle, 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 electromyography (EMG) pro-

vides 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 pre-motor 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, Burle, Osman, & Hasbroucq, 2002; Weindel et al., 2021). Note that, per construction, RT = PMT + MT. Extending Corallo et al. (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:

$$iRT \sim RT + exp_factors \quad (1)$$

$$iRT \sim PMT + exp_factors \quad (2)$$

$$iRT \sim PMT + MT + exp_factors \quad (3)$$

Depending on which model better predicts the data, different conclusions will be drawn. If model 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 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. Functionnally, 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 deserves a bit of 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 i) both PMT and MT are accessible, ii) they have different weights and iii), for the reasons explained for model 2, this would be incompatible with an external duration estimation process.

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 i) refining the relation between individual RTs and iRTs in a simple choice reaction time task and ii) 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 robustess 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).

Methods

This study was programmed with Psychopy (Peirce et al., 2019) and exported to Pavlovia.org (v2020.2) to be run online.

Participants

To recruit participants, the link to the online experiment was transmitted via various means of communications within the university and lab networks, as well as social medias. The experiment was also distributed to the volunteer base of the Cognitive Science Information Network (RISC, unit of the French CNRS). We recruited 148 participants. Fifty-eight were excluded because of high error rates (>50%) or non-respect of the instructions [1]. Thus, 90 participants (mean age=28.1 years) were included in the analyses. Sixty-one of them reported their gender as females, and 29 as male. Fourteen were left-handed. All participants performed the experiment online, on their own computer. The experiment was self-paced by the participants and took approximately 30 minutes to complete. Before the beginning of the experiment, consent form, information notice and instructions were presented on the screen and participants had to validate them. The study was approved by Aix-Marseille University ethical committee (#2020-12-03-008).

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

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 5 cycles/stimulus. Their contrast was set to 0.8 on a scale from 0 (uniform grey) to 1 (black and white). The stimulus could be tilted either clockwise or counterclockwise (factor "orientation") by either 1 or 7 degrees (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-seconds presentation of a fixation cross, a Gabor patch was presented at the center of the screen. The task consisted in 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, 1868; Ulrich, Mattes, & Miller, 1999; Vidal, Burle, Grapperon, & Hasbroucq, 2011). Participants were explicitly asked to be as fast and accurate as possible. Responses had to be given in less than 1 second. The stimulus remained on the screen until the response. After each response, a visual analog scale was presented on the screen. The scale ranged from 0 to 1 second and was labelled every 500 ms. Participants had to click on the scale with their mouse or touchpad to evaluate their reaction time, 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 (>1s or No-go trials), the scale appeared in grey for 2 seconds without the possibility to interact with it. During the training blocks (12 trials), feedback concerning the reaction time was given to participants on the same scale in the form of a green marker presented for 2 seconds. 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 minus 2 mad and superior to the median plus 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 (ANOVA), correlations and mixed models, using Python packages 'pingouin' (Vallat, 2018) and 'statsmodels' (Seabold & Perktold, 2010), and R package 'lme4' (Bates, Mächler, Bolker, & Walker, 2015). In the linear mixed model analysis, iRT was used as the dependent variable, with RT, angle and orientation as fixed effects, participants as random effect, with both random intercepts and random slopes for each predictor. Preditors were chosen based on comparisons of the full model with models including a subset of predictors, with likelihood ratio tests and comparison of "Akaike Information Criterion" (AIC, Akaike, 1974). AIC incorporates the likelihood of a model while also penalizing for the number of parameters, the lowest AIC being preferred. The full model always outperfomed the reduced ones (see Appendix A for the AIC of all tested models). 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=0.2253). Furthermore, given the question of interest (temporal vs other cues), fixed effects were judged appropriate enough to address it powerfully.

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 manuscript 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.

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=0.26), nor on iRTs (t(89)=0.19, p=0.85). 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

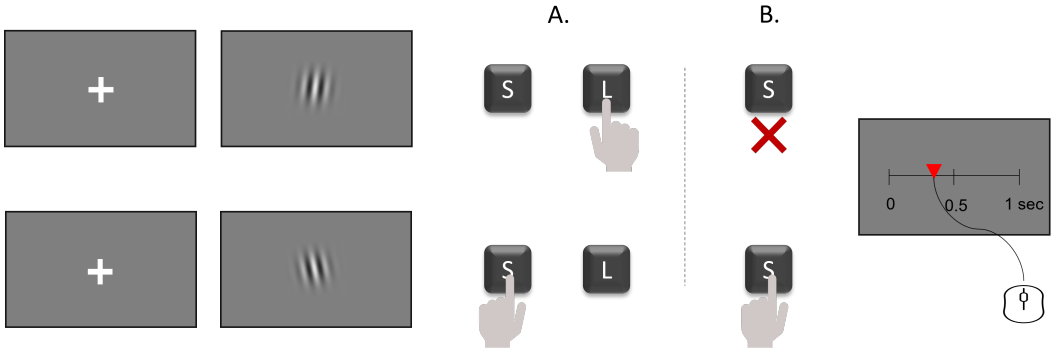


Figure 2

Experimental paradigms of A. Choice B. Go/No-go blocks of experiment 1.

results from the choice task will be presented below[2]. Analysis of omission are reported in Appendix B. In a first set of analyses, only correct trials were included.

Effect of experimental conditions on RTs and iRTs

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 ($F(1, 89)=546.95$, $p<0.001$, $\eta_p^2=0.86$), and when stimuli were counterclockwise ($F(1, 89)=11.16$, $p=0.001$, $\eta_p^2=0.11$)[3]. No interaction was found between these two factors ($F(1, 89)=2.28$, $p=0.13$). iRTs were also affected by angle: they were higher when angle was smaller ($F(1, 89)=132.49$, $p<0.001$, $\eta_p^2=0.60$). Contrary to RT, no effect of orientation was observed ($F(1, 89)=1.0$, $p=0.32$), but the two factors interacted ($F(1, 89)=5.91$, $p=0.017$, $\eta_p^2=0.06$). 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 ($r_{rm}(7371)=0.56$, $p<0.001$). Correlation coefficients were also computed for each participant using Pearson's correlation (see Figure 3). Most subjects (77/90, considering an $\alpha = 0.05$) presented a positive correlation between RTs and iRTs (range=-0.06-0.89, mean=0.56, std=0.23, see Appendix L for examples of individual data).

Dissociating RT and experimental factors contributions to iRT

We evaluated whether participants based their evaluation solely on time information or whether non-timing information could also impact the subjective evaluation of their timing performance. Paralleling the correlation observed above, mixed model showed a main effect of RT (β=0.54, t=17.71, p<0.001). More importantly, effects of angle (β=-0.002, t=-3.5, p<0.001, see Figure 4A) and orientation (β=-0.015, t=-4.47, p<0.001, 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 temporal evaluation as it reveals that, for the same objective duration, RT for smaller angles, or for clockwise stimuli, were judged longer. This is further supported by a mediation analysis provided in Appendix D.

We so far concentrated on correct trials. Finally, to assess if RT of errors were estimated in the same way as correct ones, we used a mixed model with iRT as dependent variable, and RT and response accuracy as predictors. In addition to the main effect of RT (β=0.57, t=17.67, p<0.001), an effect of accuracy (β=0.014, t=2.38, p=0.017) was observed, with smaller iRT for error trials, even though RTs were shorter on correct trials (t(87)=6.7, p<0.001).

[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 response side. It is hence impossible to know whether it is a perceptual - stimulus orientation - or a motor - response side - effect

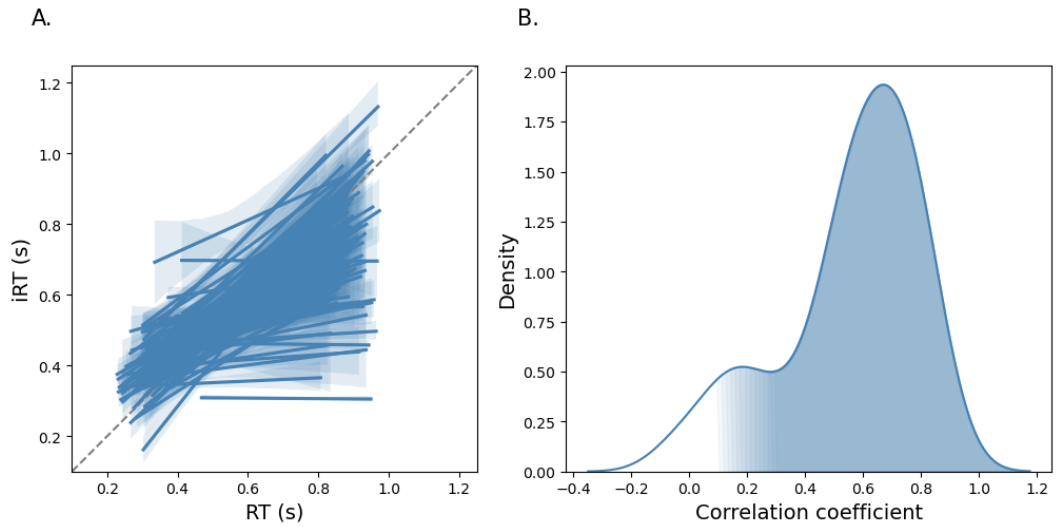


Figure 3

Experiment 1. A. Individual trial-by-trial correlations between RT and iRT. Each regression line represents a participant. Grey dashed line is the x=y line. B. Distribution of individual correlation coefficients (r). Blue (grey) zone corresponds to positive correlations and white zone to non significant correlations (considering an $\alpha = 0.05$). The gradient between the two is the zone of uncertainty concerning correlation significance.

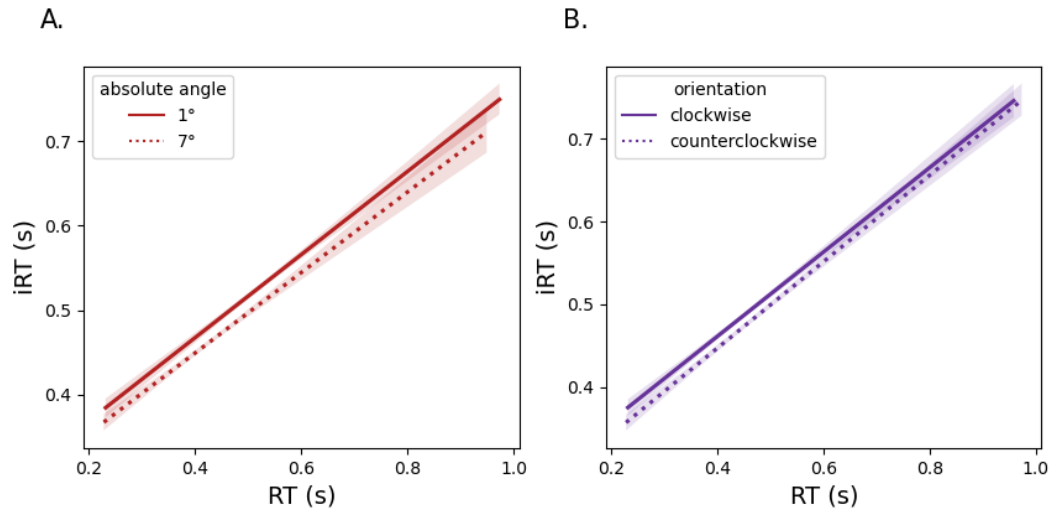


Figure 4

Experiment 1. Linear regression between trial-by-trial RTs and iRTs for A. the two stimulus angles, B. the two orientations.

Table 1

Means of the different variables for each experimental condition of experiment 1. CI-95% indicates the confidence interval around the mean. Statistics on omissions and accuracy are reported in Appendix B

	Orientations							
	-7		-1		1		7	
	Mean	*CI-95%*	*Mean*	*CI-95%*	*Mean*	*CI-95%*	*Mean*	*CI-95%*
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]

Discussion

Experiment 1 reproduced 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 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, *i.e.* 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, RT of error trials were 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 5 (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 which (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 premotor (PMT, from stimulus onset to EMG onset) and motor times (MT, from EMG onset to recorded mechanical response, see figure 6). 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.

Methods

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" (n°2021-A01548-33). 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 (mean age=25.07) were included in the analyses. Fourteen of them reported their gender as male and sixteen as female. Eight were left-handed. Sample size was determined using G*Power (Faul, Erdfelder, Buchner, & Lang, 2009) with effect sizes available from correlation analyses in the existing literature and experiment 1 (r=0.5). Based on this effect, an alpha of 0.05 and an 80% power, the required number of participants was 21. However, we assumed smaller effect sizes given RT fractioning, so 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 one-meter distance, with a frame rate of 120Hz. 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 3ms sound feedback at 400 Hz (resembling a small click). Left button was fixed in a vertical position by a magnet located at the bottom end of the cylinder (see Figure 5A). When the magnet was off, the button could be freely tilted to the right (towards the center of the device) up to 45 degrees, 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 electromyographic (EMG) activity of the *flexor pollicis brevis* of both hands with two electrodes placed on each thenar eminences. 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-hour session including a short duration estimation task of approximately 15 minutes and a reaction time 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 degrees 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 reaction time variations. Reports of durations and reaction times were collected via a horizontally presented visual analog scale. The scale was marked with vertical lines at 0 (left end), 0.33, 0.66, 1 and 1.33s (right end). Only the 1 second tick was labeled as a cue for participants. We set the right end of the scale at a value superior to one second 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 3 left- and 3 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. 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).

[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.

Externally defined duration estimation task

To assess participants' general timing abilities, they first performed an externally-defined duration estimation task (see Figure 5B), 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 second, 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 second. Possible durations were defined beforehand from a lognormal distribution (μ=0.45 and std=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 visual analog scale 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 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 6N (600g). 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 seconds 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 reaction times after each trial (see Figure 5C). Each trial began with the presentation of a fixation cross in the center of the screen. After 1 second, 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 3 blocks of each force (to avoid 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 reaction time, i.e., 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 reaction time and to press the left button to validate their estimation. The force threshold for this validation was set constant at 6N (600g). 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 second, 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. Inter-trial 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.

EMG processing

EMG signal was processed with MYOnset (Spieser & Burle, in preparation) a custom-made python program to detect EMG onsets (released under open-source license, available at https://github.com/lspieser/myonset). Briefly, the signal was first high pass filtered at 10Hz, then for each trial, the baseline mean (μ) and standard deviation (σ) of the signal was computed from 0.5 s preceding stimulus onset, for each hand separately. On the post-stimulus 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, trial was not marked. In trials where a single EMG burst was detected, motor time (MT) was defined as the time between the onset of EMG burst and the force threshold crossing recorded. Pre-motor time (PMT) was defined as the time between stimulus onset and the EMG burst onset (see

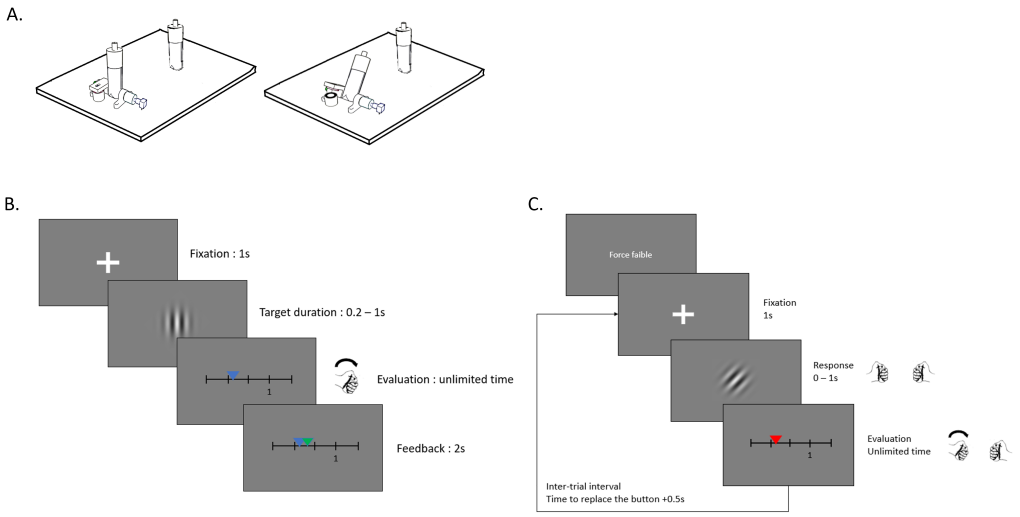


Figure 5

Experiment 2. A. Response device. Response buttons are mounted on force sensors. 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.

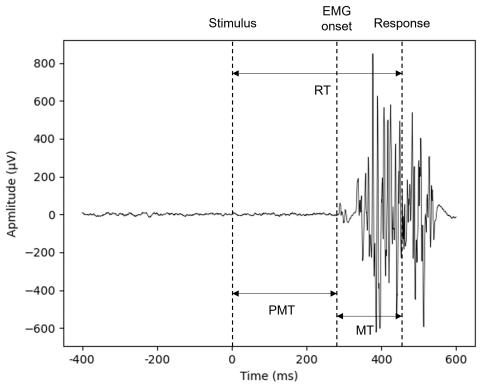


Figure 6

Experiment 2. RT fractioning into premotor (PMT) and motor times (MT).

Figure 6).

Data analysis

The first trial of each block was removed. As in the first experiment, 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 minus 2 mad and superior to the median plus 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 analyses of variance (ANOVA), 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, 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, participants as random effect, with both random intercepts and random slopes for each predictor. Preditors 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 over-complex. In accordance with experiment 1, model without interaction was judge 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 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 trials has already been reported before (Servant, White, Montagnini, & Burle, 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, *i.e.* correct trials with a unique EMG burst, were analysed. The number of the remaining trials per experimental condition is provided in Appendix H. The impact of those multiple bursts was analysed in a second step.

Pure correct trials

Effect of experimental conditions on RT and iRT

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 3 factors. RTs increased as stimulus angle decreased ($F(4, 116)=107.46$, $p<0.001$, $\eta_p^2=0.79$, $\epsilon=0.35$), and were slower in the strong force condition ($F(1, 29)=32.45$, $p<0.001$, $\eta_p^2=0.53$), and for counterclockwise stimuli ($F(1, 29)=22.92$, $p<0.001$, $\eta_p^2=0.44$, see Figure 7A). No first nor second-order interactions were observed (all $p>0.1$).

These effects on RT were underlied by effects on its subcomponents: PMT and MT. As expected, PMT was increased when stimulus angle decreased ($F(4, 116)=94.68$, $p<0.001$, $\eta_p^2=0.77$, $\epsilon=0.36$, see Figure 7C). It was also smaller for clockwise stimuli ($F(1, 29)=7.80$, $p=0.009$, $\eta_p^2=0.21$). No effect of force ($F(1, 29)=0.16$, $p=0.69$) nor interaction was found (all $p>0.1$). MT was longer in the strong force condition ($F(1, 29)=114.29$, $p<0.001$, $\eta_p^2=0.8$), and for counter-clockwise stimuli ($F(1, 29)=26.76$, $p<0.001$, $\eta_p^2=0.5$, see Fig. 7D). MT was also increased when stimulus angle decreased ($F(4, 116)=15.06$, $p<0.001$, $\eta_p^2=0.34$, $\epsilon=0.41$). No interaction was found (all $p>0.1$).

Same effects of angle ($F(4, 116)=17.04$, $p<0.001$, $\eta_p^2=0.37$, $\epsilon=0.38$) and force ($F(1, 29)=9.96$, $p=0.004$, $\eta_p^2=0.26$) as those observed on RTs were observed on iRTs. In contrast, unlike RTs, iRTs were higher for clockwise stimuli ($F(1, 29)=11.78$, $p=0.002$, $\eta_p^2=0.29$, see Figure 7B). Angle and orientation interacted ($F(4, 116)=3.70$, $p=0.013$, $\eta_p^2=0.11$, $\epsilon=0.79$). No other first-order interaction was observed (all $p>0.1$). A second-order interaction was also present ($F(4,116)=3.91$, $p=0.014$, $\eta_p^2=0.12$, $\epsilon=0.70$)

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 trials level, we performed correlation analyses. Repeated measures correlation analysis revealed a positive correlation between trial-by-trial RTs and iRTs ($r_{rm}(15806)=0.39$, $p<0.001$). 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, mean=0.42, std=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 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 ($r_{rm}(2369)=0.72$, $p<001$). Individual Pearson's correlations showed that all subjects presented a positive correlation between target and estimated durations (range=0.27-0.9, mean=0.73, std=0.11, see Figure 9). Not surprisingly, correlation coefficients were larger than for the RT evaluation task ($t(29)=-10.97$, $p<0.001$).

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, Kriegel, Ng, & Sander, 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 ($\beta=0.92$, $t=2.61$, $p=0.015$).

Dissociating RT and experimental factors contributions to iRT

As in Exp. 1, we used mixed models with iRT as dependant variable and RT and the experimental factors as regres-

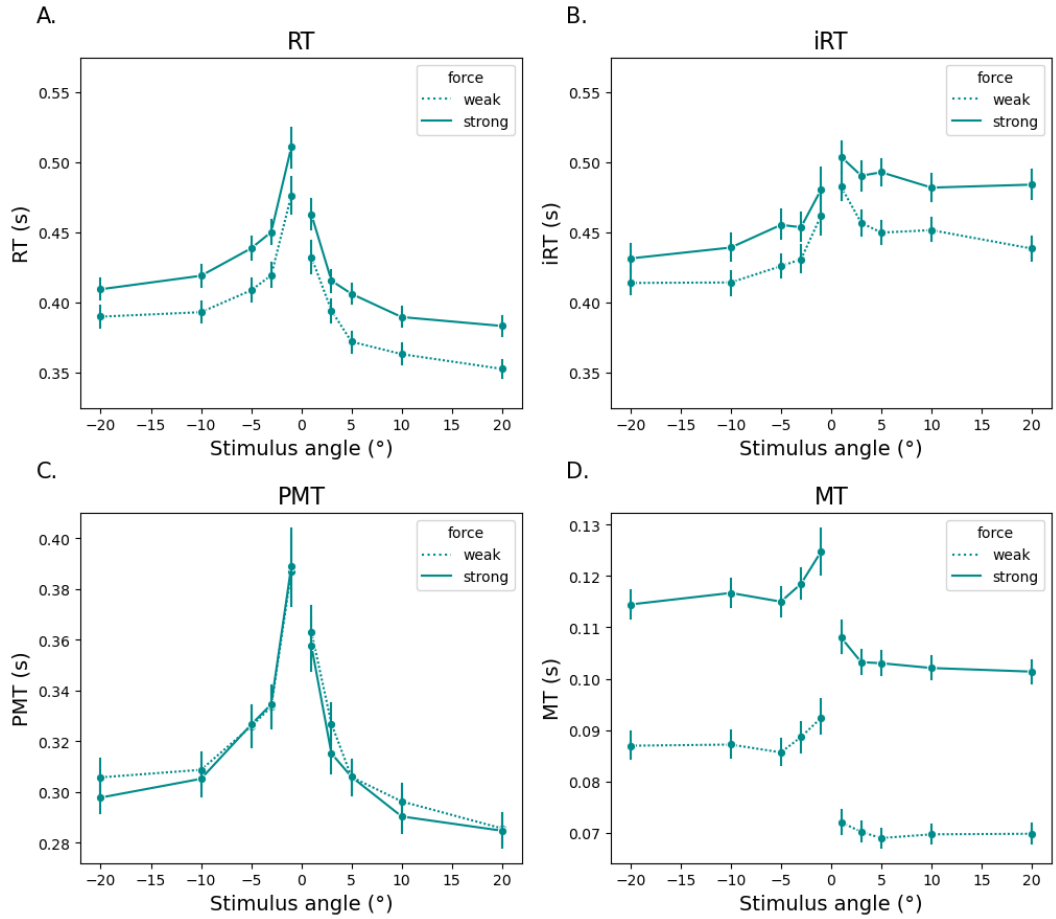


Figure 7

Experiment 2. Mean A. RT, B. iRT, C. PMT and D. MT as a function of stimulus angle and response force. Negative and positive angles represent counter- and clockwise orientations, respectively.

sors to assess if non-temporal information could also participate to RT evaluation. This model showed main effects of RT (β=0.48, t=14.49, p<0.001), force (β=-0.018, t=-1.99, p=0.047) and orientation (β=0.043, t=6.78, p<0.001), but no effect of stimulus angle (β<0.001, t=0.021, p=0.98). It means that for the same RT, iRT were higher in the strong force condition and for clockwise stimuli. Multiple regression performed on means showed similar results with main effects of RT (β=0.44, t=6.47, p<0.001), force (β=-0.02, t=-4.99, p<0.001, see Figure 11A) and orientation (β=0.04, t=13.47, p<0.001, see Figure 11B), but no effect of stimulus angle (β<-0.001, t=-0.37, p=0.72). All results are confirmed by mediation analyses provided in Appendix I.

Impact of PMT and MT on iRT

To assess to what extend premotor and motor components of RT contribue to iRT, we compared three models (models 1,2 and 3 exposed in 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 factors. Model 3 provided a better fit than model 2

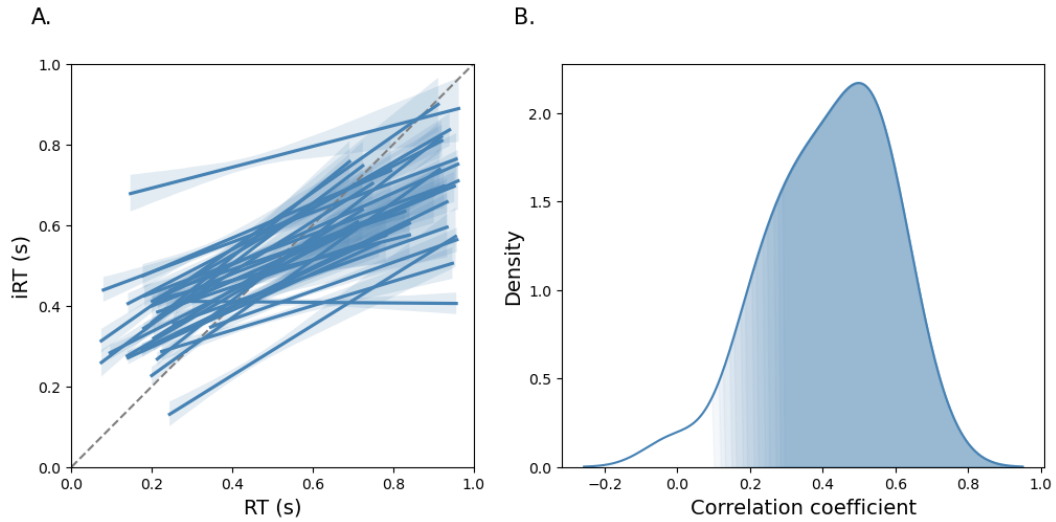


Figure 8

Experiment 2. A. Individual trial-by-trial correlations between RT and iRT. Each regression line represents a participant. Grey dashed line is the x=y line. B. Distribution of individual correlation coefficients (r). Blue (grey) zone represents positive correlations and white zone represents null correlations (considering an $\alpha = 0.05$). The gradient between the two is the zone of uncertainty concerning correlation significance.

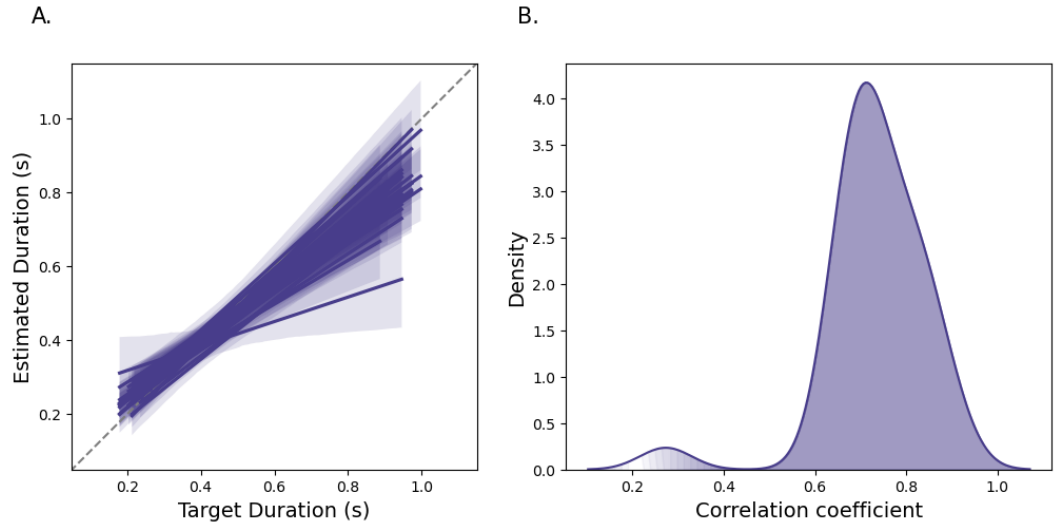


Figure 9

Experiment 2. A. Individual trial-by-trial correlations between target and estimated durations. Each regression cline represents a participant. Grey dashed line is the x=y line. B. Distribution of individual correlation coefficients (r). Purple (grey) zone represents positive correlations and white zone represents null correlations (considering an $\alpha = 0.05$). The gradient between the two is the zone of uncertainty concerning correlation significance.

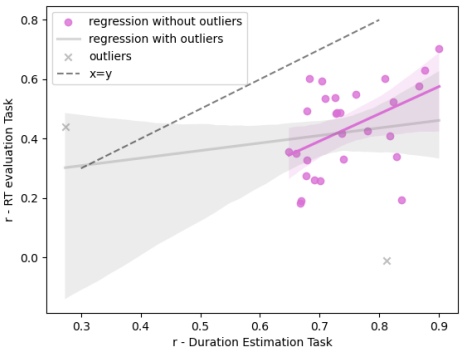


Figure 10

Experiment 2. Linear regression between individual duration estimation task and RT evaluation task correlation coefficients. Each dot is a subject, the crosses are outliers. Regressions are presented with (in light grey) and without (in dark grey) the two outliers (respectively in grey and pink in the colored version). Dashed line is the x=y line.

(χ^2(6)=482.68, p<0.001, see Table. 2). Both PMT (β=0.47, t=13.88, p<0.001) and MT (β=0.57, t=7.68, p<0.001) 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 introduction for the rationale). Model 3 (PMT+MT, different weights) better accounts for the data (χ^2(6)=124.93, p<0.001) 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 random splits of RT (N = 1000) and compared these random splits to model 1 and model 3. Details and results are reported in Appendix J: briefly, model 3 largely outperformed random-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 non-pure 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 dependent variable and RT and the number of EMG bursts as 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 (β=0.48, t=13.64, p<0.001) and force (β=-0.018, t=-1.96, p=0.05), an effect of the number of bursts (β=0.019, t=2.79, p=0.005) 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 dependent variable, and RT and response accuracy (correct vs. errors) as predictors. In addition to the main effect of RT (β=0.52, t=17.03, p<0.001), an effect of accuracy (β=0.034, t=5.46, p<0.001) 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=0.49), replicating results obtained in Experiment 1.

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 artefacts linked to the way this evaluation was provided. In experiment 1, we speculated that this asymmetry might be due to a bias towards the right part of the evaluation scale induced by the clockwise stimulus or the right response. The present data do not provide any new argument for or against such explanation, that certainly requires more investigation. Anyways, whatever the reason of 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,

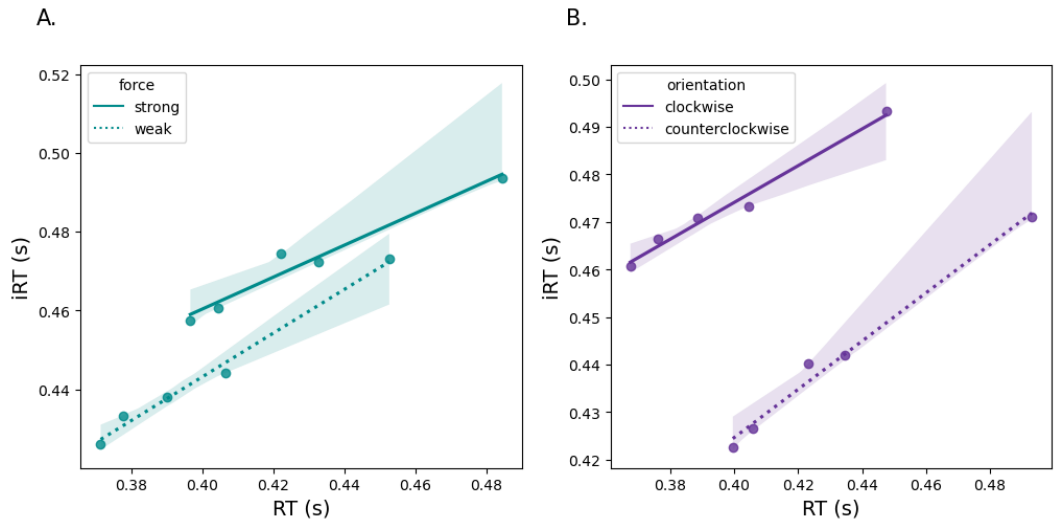


Figure 11

Experiment 2. Linear regression between mean iRTs and RTs for A. the two force levels and B. the two orientations.

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***	-28279
PMT + MT	0.242***		0.475***	0.57***	-0.016*	0.046***	-28392
PMT	0.314***		0.462***		-0.033***	0.037***	-27921

***p<.001, *p<0.1

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 motor time.

On another 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, Gratton, Bashore, Eriksen, & Donchin, 1985), called "partial errors" (Burle et al., 2002), index automatic incorrect response activation. In non-conflicting 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., 2023, 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 activations likely remain unconscious (Ficarella, Rochet, & Burle, 2019; Rochet, Spieser, Casini, Hasbroucq, & Burle, 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 sub-processes are considered to estimate RT, with different weights though. The implications of all those results will be discussed in the General Discussion.

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 non-temporal 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 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.

Although the implication of motor processes in metacognition has been observed in confidence studies (Fleming et al., 2015; Gajdos, Fleming, Saez Garcia, Weindel, & Davranche, 2019), the present result brings the first evidence that participants can access the variability of their motor execution times when assessing their RT duration. Few studies have assessed the contribution of motor variability to metacognition in visuomotor synchronization tasks (Jovanovic, López-Moliner, & Mamassian, 2021; Mamassian, 2008), with both confidence judgments and temporal error monitoring. However, this paradigm requires first-order temporal estimation, and involves ample movements with large arm displacement in space. In this case, movement time could be estimated as the duration between two visible external events: the start and the end of the arm movement. Conversely, reaction time tasks often involve isometric presses (which is the case in the present studies) without any actual movement (or very limited ones), preventing participants to base their estimation on a visual cue of the end of their motor activity.

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 on the sum of the estimations of the two intervals, an introspective PMT (iPMT) and an introspective MT (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, Gleason, Wright, & Pearl, 1983). Such conclusion, however, relies on very strong assumptions concerning the non-decision time and the lack of 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, Zhou, Datta, Joiner, & Wiener, 2023; De Kock, Zhou, Joiner, & Wiener, 2021), no studies have directly asked participants to estimate the duration of their motor execution time in non timing 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, non-temporal, information is also impacting it. This is in accordance with previous work demonstrating influence of non-temporal 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 effects of experimental factors on iRT, that 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 on 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, i.e. 'It was difficult/effortful, I must have been slow', remains an open question.

Such an account opens the question as to whether both 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 perception, 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) showed that response conflict, response time and response repetition all influenced experience of difficulty and concluded that this subjective perception relies on multiple cues. As it stands, whether "difficulty/effort" determines iRT or the other way around remains a chicken-and-egg problem, that 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 account, however, is challenged by some of the current results. First, error iRT are under-estimated. 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 of the angle effect on experiment 2 is also a challenge, as there is no reason why

the feeling of difficulty would have vanished 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 reaction times, 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 2 were present on 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 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 non-temporal 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, Grondin, & Casini, 1994), and when judging their own reaction times, 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 can not 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 is not fixed and varies 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, results showed that subjective evaluation of RT emerges from a combination of temporal and non-temporal 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, 1979’s terminology). This observation supports the idea that internally defined durations can be approximated from non-temporal 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 judgements are continuous and can be regressed one onto the other. One hence cannot be sure that the present conclusions generalize to non-continuous 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 heterogenous 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. doi: 10.1007/978-1-4612-1694-0_16

Akdoğan, B., & Balcı, F. (2017, March). Are you early or late?: Temporal error monitoring. *Journal of Experimental Psychology: General*, *146*(3), 347–361. Retrieved 2022-10-20, from http://doi.apa.org/getdoi.cfm?doi=10.1037/xge0000265 doi: 10.1037/xge0000265

Bakdash, J. Z., & Marusich, L. R. (2017, April). Repeated Measures Correlation. *Frontiers in Psychology*, *8*, 456. Retrieved 2022-10-20, from https://www.frontiersin.org/article/10.3389/fpsyg.2017.00456/full doi: 10.3389/fpsyg.2017.00456

Balci, F., Simen, P., Niyogi, R., Saxe, A., Hughes, J. A., Holmes, P., & Cohen, J. D. (2011, February). Acquisition of decision making criteria: Reward rate ultimately beats accuracy. *Attention, perception & psychophysics*, *73*(2), 640–657. Retrieved 2024-03-04, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3383845/ doi: 10.3758/s13414-010-0049-7

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015, October). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*, 1–48. Retrieved 2022-11-09, from https://doi.org/10.18637/jss.v067.i01 doi: 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. Retrieved 2022-11-16, from http://doi.apa.org/getdoi.cfm?doi=10.1037/h0022634 doi: 10.1037/h0022634

Bratzke, D., & Bryce, D. (2016, August). Temporal discrimination of one's own reaction times in dual-task performance: Context effects and methodological constraints. *Attention, Perception, & Psychophysics*, *78*(6), 1806–1816. Retrieved 2022-10-20, from http://link.springer.com/10.3758/s13414-016-1161-0 doi: 10.3758/s13414-016-1161-0

Bratzke, D., & Bryce, D. (2019, June). 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. Retrieved 2022-10-20, from http://doi.apa.org/getdoi.cfm?doi=10.1037/xlm0000635 doi: 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. Retrieved 2024-05-21, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9232406/ doi: 10.3758/s13414-022-02510-6

Bratzke, D., Bryce, D., & Seifried-Dübon, T. (2014, November). Distorted subjective reports of stimulus onsets under dual-task conditions: Delayed conscious perception or estimation bias? *Consciousness and Cognition*, *30*, 36–47. Retrieved 2022-10-20, from https://linkinghub.elsevier.com/retrieve/pii/S1053810014001305 doi: 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. Retrieved 2024-05-21, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10805817/ doi: 10.1007/s00426-023-01854-4

Breunig, M. M., Kriegel, H.-P., Ng, R. T., & Sander, J. (2000, May). LOF: identifying density-based local outliers. *ACM SIGMOD Record*, *29*(2), 93–104. Retrieved 2023-01-02, from https://doi.org/10.1145/335191.335388 doi: 10.1145/335191.335388

Brown, S. W. (1997, January). Attentional resources in timing: Interference effects in concurrent temporal and nontemporal working memory tasks. *Perception & Psychophysics*, *59*(7), 1118–1140. Retrieved 2023-02-27, from https://doi.org/10.3758/BF03205526 doi: 10.3758/BF03205526

Bryce, D., & Bratzke, D. (2014, July). Introspective reports of reaction times in dual-tasks reflect experienced difficulty rather than timing of cognitive processes. *Consciousness and Cognition*, *27*, 254–267. Retrieved 2022-10-20, from https://linkinghub.elsevier.com/retrieve/pii/S1053810014000853 doi: 10.1016/j.concog.2014.05.011

Bryce, D., & Bratzke, D. (2015, April). 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. Retrieved 2022-10-20, from http://link.springer.com/10.3758/s13414-014-0804-2 doi: 10.3758/s13414-014-0804-2

Bryce, D., & Bratzke, D. (2017, January). 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. Retrieved 2022-10-20, from https://linkinghub.elsevier.com/retrieve/pii/S0001691816303031 doi: 10.1016/j.actpsy.2016.10.007

Burle, B., & Casini, L. (2001, February). 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. doi: 10.1037//0096-1523.27.1.195

Burle, B., Possamaï, C.-A., Vidal, F., Bonnet, M., & Hasbroucq, T. (2002, November). Executive control in the Simon effect: an electromyographic and distributional analysis. *Psychological Research*, *66*(4), 324–336. Retrieved 2022-10-20, from http://link.springer.com/10.1007/s00426-002-0105-6 doi: 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*, 529–553. (Place: US Publisher: American Psychological Association) doi: 10.1037/0096-1523.11.5.529

Corallo, G., Sackur, J., Dehaene, S., & Sigman, M. (2008, November). Limits on Introspection: Distorted Subjective Time During the Dual-Task Bottleneck. *Psychological Science*, *19*(11), 1110–1117. Retrieved 2022-10-20, from http://journals.sagepub.com/doi/10.1111/j.1467-9280.2008.02211.x doi: 10.1111/j.1467-9280.2008.02211.x

De Kock, R., Zhou, W., Datta, P., Joiner, M. W., & Wiener, M. (2023, February). The role of consciously timed movements in shaping and improving auditory timing. *Proceedings of the Royal Society B: Biological Sciences*, *290*(1992), 20222060. Retrieved 2023-11-29, from https://royalsocietypublishing.org/doi/10.1098/rspb.2022.2060 doi: 10.1098/rspb.2022.2060

De Kock, R., Zhou, W., Joiner, W. M., & Wiener, M. (2021, April). Slowing the body slows down time perception. *eLife*, *10*, e63607. Retrieved 2023-11-29, from https://elifesciences.org/articles/63607 doi: 10.7554/eLife.63607

Dendauw, E., Evans, N., Logan, G., Haffen, E., Bennabi, D., Gajdos, T., & Servant, M. (2023, November). The Gated Cascade Diffusion Model: An Integrated Theory of Decision Making, Motor Preparation, and Motor Execution. *Psychological Review*, *in press*. Retrieved from https://hal.science/hal-03924556 doi: 10.1037/rev0000464

Desender, K., Van Opstal, F., & Van den Bussche, E. (2017, March). Subjective experience of difficulty depends on multiple cues. *Scientific Reports*, *7*, 44222. Retrieved 2022-11-24, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5347002/ doi: 10.1038/srep44222

Donders, F. C. (1868). Over de snelheid van psychische processen [on the speed of mental processes]. *Onderzoekingen gedaan in het Physiologisch Laboratorium der Utrechtsche Hoogeschool, 1868–1869, Tweede reeks,*, *II*, 92–120 (Transl. by W. G. Koster (1969), In W. G. Koster, Attention and performance II. Acta Psychologica, 30, 412–431).

Droit-Volet, S., & Coull, J. (2015, July). The Developmental Emergence of the Mental Time-Line: Spatial and Numerical Distortion of Time Judgement. *PLoS ONE*, *10*(7), e0130465. Retrieved 2022-11-08, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4489723/ doi: 10.1371/journal.pone.0130465

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009, November). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*(4), 1149–1160. Retrieved 2022-11-16, from http://link.springer.com/10.3758/BRM.41.4.1149 doi: 10.3758/BRM.41.4.1149

Ficarella, S. C., Rochet, N., & Burle, B. (2019, November). Becoming aware of subliminal responses: An EEG/EMG study on partial error detection and correction in humans. *Cortex*, *120*, 443–456. Retrieved 2022-10-20, from https://linkinghub.elsevier.com/retrieve/pii/S0010945219302631 doi: 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*, 906–911. (Place: US Publisher: American Psychological Association) doi: 10.1037/0003-066X.34.10.906

Fleming, S. M., Maniscalco, B., Ko, Y., Amendi, N., Ro, T., & Lau, H. (2015, January). Action-Specific Disruption of Perceptual Confidence. *Psychological Science*, *26*(1), 89–98. Retrieved 2022-10-20, from http://journals.sagepub.com/doi/10.1177/0956797614557697 doi: 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), niz001. doi: 10.1093/nc/niz001

Jovanovic, L., López-Moliner, J., & Mamassian, P. (2021, August). Contrasting contributions of movement onset and duration to self-evaluation of sensorimotor timing performance. *European Journal of Neuroscience*, *54*(3), 5092–5111. Retrieved 2022-11-02, from https://onlinelibrary.wiley.com/doi/10.1111/ejn.15378 doi: 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*. Retrieved 2023-03-07, from https://www.frontiersin.org/articles/10.3389/fpsyg.2013.00513

Klein, M. D., & Stolz, J. A. (2018, January). Making time: Estimation of internally versus externally defined durations. *Attention, Perception, & Psychophysics*, *80*(1), 292–306. Retrieved 2023-06-25, from http://link.springer.com/10.3758/s13414-017-1414-6 doi: 10.3758/s13414-017-1414-6

Kononowicz, T. W., Roger, C., & van Wassenhove, V. (2019, September). Temporal Metacognition as the Decoding of Self-Generated Brain Dynamics. *Cerebral Cortex*, *29*(10), 4366–4380. Retrieved 2022-10-20, from https://academic.oup.com/cercor/article/29/10/4366/5257447 doi: 10.1093/cercor/bhy318

Kononowicz, T. W., & van Wassenhove, V. (2019, November). Evaluation of Self-generated Behavior: Untangling Metacognitive Readout and Error Detection. *Journal of Cognitive Neuroscience*, *31*(11), 1641–1657. Retrieved 2022-10-20, from https://direct.mit.edu/jocn/article/31/11/1641-1657/95339 doi: 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*, 1121–1134. (Place: US Publisher: American Psychological Association) doi: 10.1037/0022-3514.77.6.1121

Libet, B., Gleason, C. A., Wright, E. W., & Pearl, D. K. (1983, September). 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 (Pt 3)*, 623–642. doi: 10.1093/brain/106.3.623

Liu, J., & Liu, Q. (2016, February). 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.

Luce, R. (1986). Response Times: Their Role in Inferring Elementary Mental Organization.. Retrieved 2023-03-01, from https://www.semanticscholar.org/paper/Response-Times%3A-Their-Role-in-Inferring-Elementary-Luce/4f714b26b7ceb23ccad0e47d19a6fd94ec10bf4f

Macar, F., Grondin, S., & Casini, L. (1994, November). Controlled attention sharing influences time estimation. *Memory & Cognition*, *22*(6), 673–686. Retrieved 2023-02-27, from https://doi.org/10.3758/BF03209252 doi: 10.3758/BF03209252

Mamassian, P. (2008, June). Overconfidence in an Objective Anticipatory Motor Task. *Psychological Science*, *19*(6), 601–606. Retrieved 2022-11-10, from http://journals.sagepub.com/doi/10.1111/j.1467-9280.2008.02129.x doi: 10.1111/j.1467-9280.2008.02129.x

Marti, S., Sackur, J., Sigman, M., & Dehaene, S. (2010, May). Mapping introspection's blind spot: Reconstruction of dual-task phenomenology using quantified introspection. *Cognition*, *115*(2), 303–313. Retrieved 2022-10-20, from https://linkinghub.elsevier.com/retrieve/pii/S0010027710000168 doi: 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. doi: 10.1016/0301-0511(88)90013-0

Miller, J., Vieweg, P., Kruize, N., & McLea, B. (2010, December). Subjective reports of stimulus, response, and decision times in speeded tasks: How accurate are decision time reports? *Consciousness and Cognition*, *19*(4), 1013–1036. Retrieved 2022-10-20, from https://linkinghub.elsevier.com/retrieve/pii/S1053810010001182 doi: 10.1016/j.concog.2010.06.001

Morsella, E., Wilson, L. E., Berger, C. C., Honhongva, M., Gazzaley, A., & Bargh, J. A. (2009, November). Subjective aspects of cognitive control at different stages of processing. *Attention, Perception & Psychophysics*, *71*(8), 1807–1824. Retrieved 2022-10-20, from http://link.springer.com/10.3758/APP.71.8.1807 doi: 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. doi: 10.1037/0033-295X.84.3.231

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., … Lindeløv, J. K. (2019, February). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195–203. Retrieved 2022-10-31, from https://doi.org/10.3758/s13428-018-01193-y doi: 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, Jul). Partial advance information, number of alternatives, and motor processes: an electromyographic study. *Acta Psychol (Amst)*, *111*(1), 125–139.

Questienne, L., Atas, A., Burle, B., & Gevers, W. (2018, January). Objectifying the subjective: Building blocks of metacognitive experiences in conflict tasks. *Journal of Experimental Psychology: General*, *147*(1), 125–131. Retrieved 2022-10-20, from http://doi.apa.org/getdoi.cfm?doi=10.1037/xge0000370 doi: 10.1037/xge0000370

Rabbitt, P. M. (1966). Errors and error correction in choice-response tasks. *Journal of Experimental Psychology*, *71*(2), 264–272. Retrieved 2023-01-02, from http://doi.apa.org/getdoi.cfm?doi=10.1037/h0022853 doi: 10.1037/h0022853

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

Ratcliff, R., Smith, P. L., Brown, S. D., & McKoon, G. (2016, April). Diffusion Decision Model: Current Issues and History. *Trends in cognitive sciences*, *20*(4), 260–281. Retrieved 2022-11-22, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4928591/ doi: 10.1016/j.tics.2016.01.007

Reyes, G., & Sackur, J. (2014, October). Introspection during visual search. *Consciousness and Cognition*, *29*, 212–229. Retrieved 2022-10-20, from https://linkinghub.elsevier.com/retrieve/pii/S1053810014001470 doi: 10.1016/j.concog.2014.08.009

Rochet, N., Spieser, L., Casini, L., Hasbroucq, T., & Burle, B. (2014, September). Detecting and correcting partial errors: Evidence for efficient control without conscious access. *Cognitive, Affective, & Behavioral Neuroscience*, *14*(3), 970–982. Retrieved 2022-10-20, from http://link.springer.com/10.3758/s13415-013-0232-0 doi: 10.3758/s13415-013-0232-0

Sanford, A. J. (1970, December). Rating the speed of a simple reaction. *Psychonomic Science*, *21*(6), 333–334. Retrieved 2022-10-20, from http://link.springer.com/10.3758/BF03335809 doi: 10.3758/BF03335809

Seabold, S., & Perktold, J. (2010). Statsmodels: Econometric and Statistical Modeling with Python. In (pp. 92–96). Austin, Texas. Retrieved 2022-11-09,

from https://conference.scipy.org/proceedings/scipy2010/seabold.html doi: 10.25080/Majora-92bf1922-011

Servant, M., White, C., Montagnini, A., & Burle, B. (2016, October). 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. Retrieved 2022-11-25, from https://direct.mit.edu/jocn/article/28/10/1501/28581/Linking-Theoretical-Decision-making-Mechanisms-in doi: 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. Retrieved from http://www.jstor.org/stable/2984065

Stanton, J. D., Sebesta, A. J., & Dunlosky, J. (2021). Fostering Metacognition to Support Student Learning and Performance. *CBE Life Sciences Education*, *20*(2), fe3. Retrieved 2023-01-27, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8734377/ doi: 10.1187/cbe.20-12-0289

Sternberg, S. (1969). The discovery of processing stages: Extensions of Donders' method. *Acta Psychologica*, *30*, 276-315. Retrieved from https://api.semanticscholar.org/CorpusID:143389711

Ulrich, R., Mattes, S., & Miller, J. (1999, July). Donders's assumption of pure insertion: an evaluation on the basis of response dynamics. *Acta Psychologica*, *102*(1), 43–76. Retrieved 2024-03-04, from https://linkinghub.elsevier.com/retrieve/pii/S0001691899000190 doi: 10.1016/S0001-6918(99)00019-0

Vallat, R. (2018, November). Pingouin: statistics in Python. *Journal of Open Source Software*, *3*(31), 1026. Retrieved 2022-11-09, from https://joss.theoj.org/papers/10.21105/joss.01026 doi: 10.21105/joss.01026

Vidal, F., Burle, B., Grapperon, J., & Hasbroucq, T. (2011, September). An ERP study of cognitive architecture and the insertion of mental processes: Donders revisited: ERPs and the insertion assumption. *Psychophysiology*, *48*(9), 1242–1251. Retrieved 2022-10-20, from https://onlinelibrary.wiley.com/doi/10.1111/j.1469-8986.2011.01186.x doi: 10.1111/j.1469-8986.2011.01186.x

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

Yeung, N., & Summerfield, C. (2012, May). Metacognition in human decision-making: confidence and error monitoring. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1594), 1310–1321. Retrieved 2022-10-20, from https://royalsocietypublishing.org/doi/10.1098/rstb.2011.0416 doi: 10.1098/rstb.2011.0416

Table A1

Experiment 1. Different models tested and their corresponding AIC.

Model	AIC
RT	-14327
RT+angle	-14395
RT+orientation	-14433
RT+angle+orientation	-14492

Appendix

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 RT only, RT and a subset of experimental factors or RT and all experimental factors as fixed effects. The different models tested and their correponding AIC are presented in Table A1. Likelihood ratio tests revealed that the full model with all experimental factors outperformed the others (all p<0.001).

Appendix B
Mean omission rate and accuracy, experiment 1

Trials without any recorded response, e.g., when no response was given in less than 1 second, were considered as omissions (3.9% of trials). Omission and error rates were arcsine transformed and two-way ANOVAs were performed to assess the effect of experimental conditions. The omission rate was higher with smaller angles (F(1,89)=91.9, p<0.001, η_p^2=0.51), but no effect of orientation (F(1,89)=0.62, p=0.43) nor any interaction between these two factors (F(1,89)=0.68, p=0.41) were observed. Angle and orientation also affected accuracy, with higher error rate with smaller angles (F(1,89)=444.38, p<0.001, η_p^2=0.83), and for counter-clockwise stimuli (F(1,89)=4.6, p=0.035, η_p^2=0.05).

Appendix C
Results of the Go/NoGo 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<0.001). This was also the case for iRT (t(89)=-9.32, p<0.001). Repeated measures correlation analysis revealed a positive correlation between the two variables (r_{rm}(3807)=0.61, p<0.001). Correlation coefficients were also computed for each participant using Pearson's correlation (range=-0.16-0.92, mean=0.59, std=0.23). Mixed model with RT and angle as predictors of iRT showed main effects of RT (β=0.59, t=42.22, p<0.001) and angle (β=-0.001, t=-2.5,p=0.013). This means that, as in the choice task, for the same value of RT, iRT was higher for smaller angles.

Table D1

Mediation analysis results with data from Experiment 1. Stimulus angle was used as predictor, RT as mediator and iRT as the dependant variable. ACME=Average Causal Mediation Effect, ADE=Average Direct Effect, Total Effect=combined direct and indirect effects and Prop. Mediated=the ratio of these estimates. CI-95% indicates the confidence intervals.

	Estimate	*CI-95%*	*p-value*
ACME	-0.009	[-0.0103, 0.0076]	<0.001
ADE	-0.002	[-0.0035, -0.0009]	<0.001
Total Effect	-0.011	[-0.013, -0.009]	<0.001
Prop. Mediated	0.79	[0.72, 0.90]	<0.001

Table E1

Range of response forces used. Forces are presented in Newtons. MVC = Maximum voluntary force

	MVC	**Weak force**	**Strong force**
min	39.07	1.95	7.81
max	98.80	4.94	19.76
mean	66.94	3.35	13.39
std	18.41	0.92	3.68

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 predictor, RT as mediator and iRT as the dependant variable. Results, presented in Table D1, showed that even after accounting for the mediating role of RT (ACME=-0.009, p<0.001) angle still had a direct effect on iRT (ADE=-0.002, p<0.001).

Appendix E
Range of used forces, experiment 2

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

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 correponding AIC are presented in Table F1. Likelihood ratio tests revealed that the full model with all experimental factors outperformed the others (all p<0.001).

Appendix G

Table F1

Experiment 2. Different models tested and their corresponding AIC.

Model	**AIC**
RT	-26351
RT+angle	-26406
RT+force	-27143
RT+force+angle	-27273
RT+angle+force+orientation	-28338

Mean omission rate and accuracy, experiment 2

Trials without any recorded response, e.g., when no response was given in less than 1 second, 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 ($F(4,116)=24.51$, $p<0.001$, $\eta_p^2=0.46$, $\epsilon=0.84$), was higher in the strong force condition ($F(1,29)=19.32$, $p<0.001$, $\eta_p^2=0.40$) and for counterclockwise stimuli ($F(1,29)=5.52$, $p=0.026$, $\eta_p^2=0.16$, see Figure G1A). No first nor second order interactions were observed (all p>0.1). Accuracy increased with stimulus angle ($F(4,116)=126.55$, $p<0.001$, $\eta_p^2=0.81$, $\epsilon=0.53$), was higher in the strong force condition ($F(1,29)=22.89$, $p<0.001$, $\eta_p^2=0.44$) and for clockwise stimuli ($F(1,29)=12.19$, $p=0.002$, $\eta_p^2=0.30$, see Figure G1B). An interaction between angle and orientation was observed ($F(4,116)=7.11$, $p<0.001$, $\eta_p^2=0.20$, $\epsilon=0.49$). No other interaction was present (p>0.1).

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 measure ANOVA revealed effects of angle ($F(4, 116)=34.29$, $p<0.001$, $\eta_p^2=0.54$, $\epsilon=0.77$) and force ($F(1, 29)=47.84$, $p<0.001$, $\eta_p^2=0.62$) on the percentage of trials.

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 predictor, RT as mediator and iRT as the dependant variable. Results, presented in Table I1, showed that even after accounting for the mediating role of RT (ACME=-0.0127, p<0.001) force still had a direct effect on iRT (ADE=-0.0175, p<0.001). On the contrary, all the angle effect was mediated by RT (ACME=-0.001325, p<0.001), with no direct effect remaining (ADE=-0.000233, p=0.054).

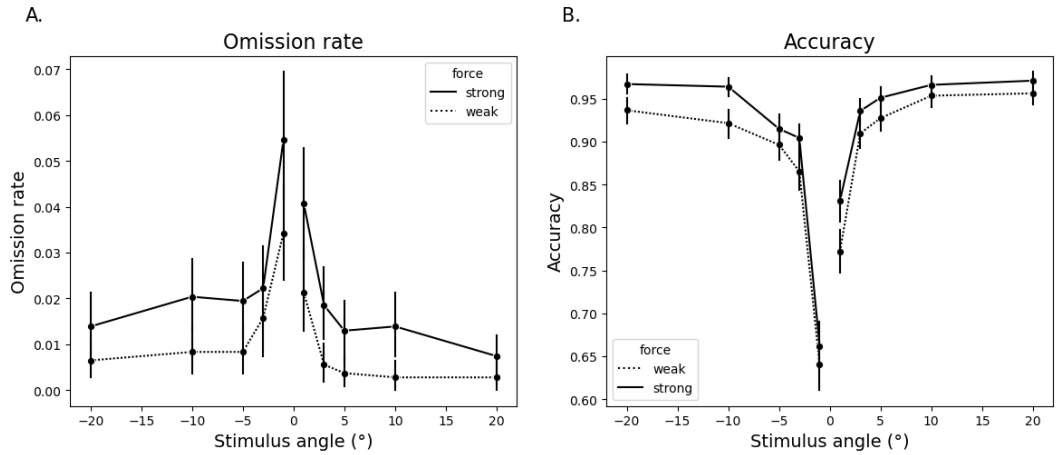


Figure G1

Experiment 2. A. Omission rate and B. mean accuracy as a function of stimulus angle and response force. Negative and positive angles represent counter- and clockwise orientations, respectively.

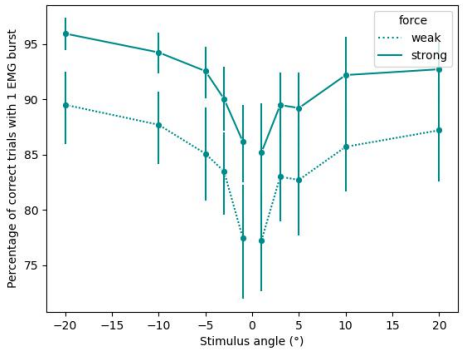


Figure H1

Percentage of correct trials included, i.e. with one EMG burst, as a function of stimulus angle and response force. Negative and positive angles represent counter- and clockwise orientations, respectively.

Table I1

Mediation analysis results with data from Experiment 2. Stimulus angle or response force was used as predictor, RT as mediator and iRT as the dependant variable. ACME=Average Causal Mediation Effect, ADE=Average Direct Effect, Total Effect=combined direct and indirect effects and Prop. Med.=the ratio of these estimates. CI-95% indicates the confidence intervals.

	Estimate	*CI-95%*	*p-value*
Response force			
ACME	-0.013	[-0.014, 0.011]	<0.001
ADE	-0.018	[-0.021, -0.014]	<0.001
Total Effect	-0.030	[-0.034, -0.027]	<0.001
Prop. Med.	0.42	[0.37, 0.48]	<0.001
Stimulus angle			
ACME	-0.0013	[-0.0014, -0.0012]	<0.001
ADE	-0.00023	[-4.9e-04, -7.3e-07]	0.054
Total Effect	-0.0016	[-0.0018, -0.0013]	<0.001
Prop. Med.	0.849	[0.732, 1.001]	<0.001

Appendix J
Random split of RTs, experiment 2

To ensure that the prediction of iRT by PMT and MT was not equivalent to its prediction by any random subdivisions of RT, we generated random intervals summing to the RT (random PMTs and MTs) and used them as predictors of iRT. To do that, empiric MT/PMT ratios distribution across all participant was plotted, and for each RT values, random PMTs and MTs were generated by taking a random ratio from this distribution. Random PMTs and MTs were finally entered as predictors of iRT in mixed models. We ran 1000 iRT~random PMT + random MT models. The distribution of their AIC 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 (zscore=-1.1), the AIC of the iRT~PMT+MT model was clearly lower (zscore=-25.76, see Fig. J1).

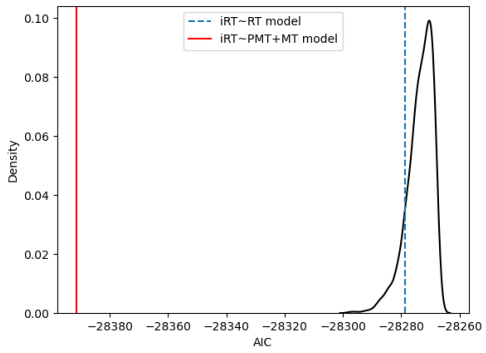


Figure J1

Distribution of AICs of models with randomly generated PMTs and MTs as predictors. 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

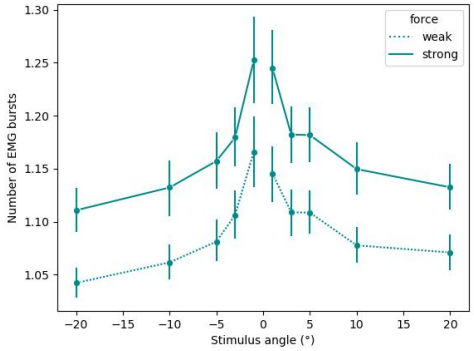


Figure K1

Mean number of EMG bursts as a function of stimulus angle and response force. Negative and positive angles represent counter- and clockwise orientations, respectively.

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 Fig. K1). Number of bursts was increased when stimulus angle decreased ($F(4, 116)=36.0$, $p<0.001$, $\eta_p^2=0.55$, $\epsilon=0.61$). It was also smaller in the weak force condition ($F(1, 29)=49.29$, $p<0.001$, $\eta_p^2=0.63$).

Appendix L

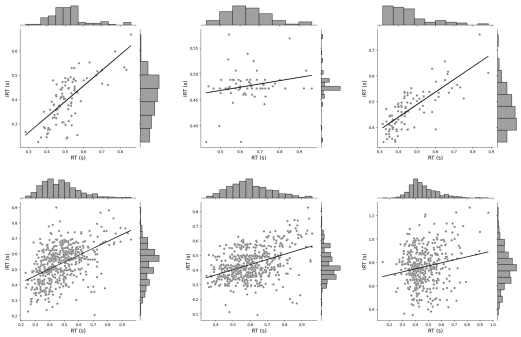


Figure L1

Scatter plots of individual iRT as a function of RT for the 3 first participants of experiment 1 (top row) and experiment 2 (bottom row). Each scatter plot is associated with histograms of both variables and the linear regression line.

Examples of individual RT-iRT correlations

To provide a sense of intraindividual variability in the relationship between RT and iRT, we present the scatter plots of the 3 first participants of experiment 1 and 2 in Fig. L1.

3. Assessing MPFC contribution to subjective evaluation of RT

Summary

3.1. Note on the present chapter . . . 78
3.2. Article: Metacognitive context tunes MPFC performance monitoring activity . . . 79
3.2.1. Introduction . . . 80
3.2.2. Results . . . 82
3.2.3. Discussion . . . 84
3.2.4. Methods . . . 87
3.2.4.1. Participants . . . 87
3.2.4.2. Apparatus . . . 87
3.2.4.3. Stimuli . . . 88
3.2.4.4. Maximum voluntary force . . . 88
3.2.4.5. Choice reaction time task . . . 88
3.2.4.6. EMG and EEG preprocessing . . . 89
3.2.4.7. Data analysis . . . 91
3.2.5. Extended Data Figures . . . 94
3.2.6. References . . . 98
3.2.7. Supplementary Information . . . 103
3.2.7.1. Supplementary Notes . . . 103
3.2.7.2. Supplementary Figures . . . 106
3.3. From objective to subjective performance monitoring: MPFC activity and beyond . . . 110
3.3.1. Introduction . . . 110
3.3.2. Methods . . . 111
3.3.3. Results . . . 112
3.3.3.1. MPFC activity as a function of iRT . . . 112
3.3.3.2. MPFC activity as a function of PMT and MT . . . 113
3.3.3.3. Primary motor cortex activation . . . 115
3.3.3.4. Activity following the response . . . 116
3.3.4. Discussion . . . 117
3.3.4.1. Contribution of MPFC activity to iRT . . . 117
3.3.4.2. iRT likely arises from multiple cerebral activities . . . 119
3.3.4.3. A need to dissociate first and second order measures . 120

3.1. Note on the present chapter

To further advance on our understanding of the processes leading to iRT, we investigated the cerebral activities that could underly this subjective experience. Neurophysiological measures provide objective data that can be correlated with subjective measures (Frith et al., 1999), allowing for the quantification and qualification of subjective experiences through observable physiological activities.

Our working hypothesis (see General Introduction) was that subjective RT estimation is taken care by a general performance monitoring system, of which Ne is a direct index. Crucially, several conditions have to be met to consider this activity as a basis for the subjective representation of the time taken to respond.

First, if it has such a role, Ne should vary parametrically with first-order RT, that is, it should be sensitive to this temporal dimension. However, many activities could vary with RT without being involved in its subjective judgment. Hence, a second condition must hold: this sensitivity should be observed only when participants experience this subjective representation of RT, meaning that it should be specific to RT judgments.

We designed an EEG experiment, capitalizing on the one of chapter 2, in which participants had to evaluate their RT after each trial, to assess the two above properties: sensitivity and specificity. To assess the specificity to RT judgments, we had to build up a control task without such judgment. However, a mere comparison with a task without any metacognitive evaluation would not be a good comparison as i) it would provide a very different timeline of the task and ii) being involved in a subjective evaluation in itself increases Ne amplitude (Grützmann et al., 2014). We hence developped a "control" task in which participants had to evaluate another dimension of their performance, rather than RT. The force produced on the response button was chosen as this second dimension because it met several requirements. First, force and RT are, a priori, uncorrelated, so the evaluation of the first one cannot be confounded with the second. Second, force is internally produced by the participant, so that its evaluation is also metacognitive. Finally, produced force is continuous and accessible at the moment of the response, as RT. To avoid interference between the two judgments, two different groups of participant performed either one of the two tasks (see method in section 3.2).

Interestingly, besides the questions of the present thesis concerning RT judgments, this force estimation experiment allowed to highlight metacognitive abilities concerning another dimension, i.e. produced force, and allowed to replicate some of the findings of chapter 2, both on RT and force (see Appendices A.1 and A.2). This also led to the hypothesis that a similar sensitivity of the Ne should be observed with force when it is the dimension to evaluate. This first set of results, that revealed interesting properties of the MPFC performance monitoring activity, as well its specificity to the dimension being evaluated, constitutes the first part of the present chapter, and the second article related to this thesis, which has been submitted for publication.

Although the analysis reported on the first part clearly establish that MPFC/Ne has a very generic role in performance monitoring, they say few about its direct relationship with the metacognitive representation of RT. Note that, we do not have access to this

subjective representation, only to its behavioral expression, iRT. As developped above, our initial hypothesis was thus that Ne supports iRT. If true, it should also vary with iRT, more closely than with RT. However, based on the results reported on chapter 2, we can cast doubt on the hypothesis of a strong link between Ne and iRT. Indeed, chapter 2 showed that iRT is a complex representation of RT, built not only from an estimation of its duration, but also by inferences drawn from task context. Furthermore, it evidenced that participants can access both decision- and motor execution-related processes (measured through PMT and MT) to report their iRT. The Ne occuring at the end of PMT, during MT, it is unlikely that it can directly have access to MT duration. However it appears as a perfect candidate to code for a subjective representation of PMT, iPMT. In the second part of this chapter, we hence first investigated Ne sensitivity to iRT and to RT subcomponents, PMT and MT. In a second set of analyses, much more exploratory, we investigated some other brain activities that could contribute to iRT or to one of its subcomponents.

3.2. Article: Metacognitive context tunes MPFC performance monitoring activity

Metacognitive context tunes medial prefrontal cortex performance monitoring activity

Nathalie Pavailler[1], Franck Vidal[1], Boris Burle[1*]

[1]CRPN, Aix-Marseille Univ., CNRS, Marseille, France.

*Corresponding author(s). E-mail(s): boris.burle@univ-amu.fr;

Abstract

Cognitive processes and behavioral performance are inherently variable. While this variability is essential for adaptation, it can have deleterious effects, and must hence be internally monitored. Here we show that medial pre-frontal cortex (MPFC) activity can selectively track various sources of performance variability, when contextually relevant. Participants had to either evaluate the time they took or the force they produced to respond in a choice reaction time task. While both dimensions were highly variable in both cases, MPFC activity was selectively tuned by the evaluated dimension, but was insensitive to the non-evaluated one. This double dissociation demonstrates the richness of performance monitoring during correct actions, besides error processing, and shows that it can be configured to selectively track relevant variations, defined by the ongoing metacognitive context.

Keywords: Performance Monitoring, Medial prefrontal cortex, Electroencephalography

Introduction

In daily activities, such as typing on a keyboard, or reaching our cup of coffee, one never performs the same task twice the same way. Our behavior is variable in many different dimensions such as space, time, smoothness, etc. While this variability, both in performance and in the underlying processes, is a key aspect for efficient adaptation [1], it can also lead to catastrophes, often caused by the co-occurrence of low probability events [2]. This variability has hence to be monitored, both externally by the goal-setters, and internally by the agent performing the task. However, depending on the context, all those sources of variability might not be equally relevant and may not deserve the same amount of monitoring.

How the human brain evaluates intrinsic variability and how it copes with its different sources remains poorly understood. The present study aims at clarifying these points.

The medial pre-frontal cortex (MPFC), especially the Supplementary Motor Areas (SMA) and the Anterior Cingulate Cortex (ACC), has been shown to play a critical role in action monitoring [3–6]. In

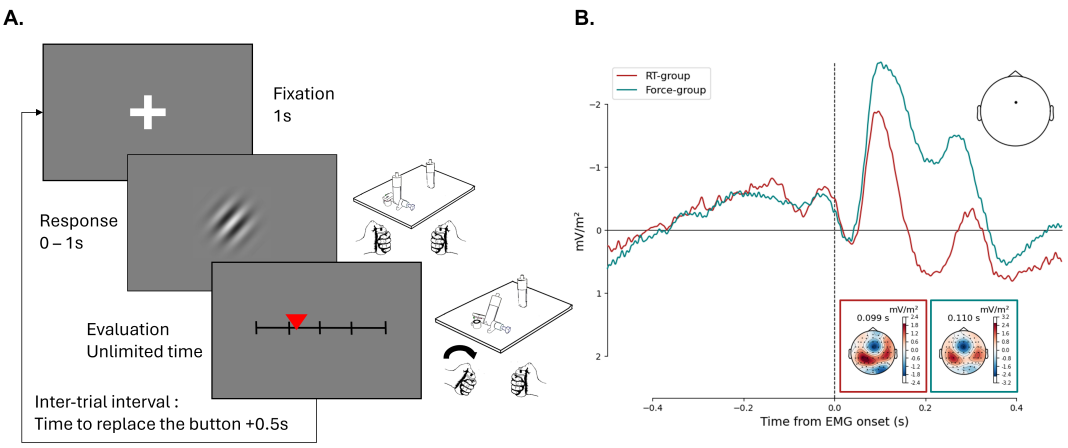


Fig. 1 Experimental task and response device. (A) Participants had to respond according to the orientation of a Gabor patch by pressing a button mounted on a force sensor. They subsequently evaluated their RT or produced force on a visual analog scale. (B) Grand averages of all trials at electrode FCz for RT (red) and Force (blue) groups and their corresponding topographies at peak latencies. Ne of both groups had similar fronto-central topographies. Ne peaked slightly later in the Force-group (mean=110 ms) compared to the RT-group (mean=98 ms, $t(34) = 1.92, p = .06$), and had a marginally higher amplitude ($t(37) = 1.92, p = .06$).

humans, an EEG neural marker of this monitoring can be found in the so-called "Error Negativity" (Ne, [7], also termed "Error Related Negativity", ERN, [8]). While this activity was initially thought to reflect "error detection" [7–9], some evidence suggests that it could subserve a more generic "action monitoring" process [4, 10]. First, the presence of this activity, although of smaller amplitude, on correct trials [11–14] challenged the idea of an error detection mechanism. Second, the observations, among other evidence, that Ne amplitude varies with spatial error magnitude in a reaching task [15, 16], or with RT when a response deadline is imposed [17], suggest that the Ne tracks not only choice errors but also some more continuous dimensions such as timing [17–19] or motor execution [20–22]. Furthermore, in contexts where the monitoring system was oriented towards the temporal aspect of performance, a relationship between Ne and RT in correct trials was observed (when feedback or reinforcement was provided based on RT [19, 23], or when participants were required to subjectively rate each of their RT [18]). This contrasts with an earlier study, that did not include specific instructions regarding RT, where no significant link between Ne amplitude and RT was observed [14]. Along with the fact that Ne amplitude is increased in both correct and error trials when participants are asked to subjectively rate their performance [24], this suggests that the MPFC performance monitoring activity might depend on task demand and context.

To adress those questions, we used MPFC monitoring activity (measured through the Ne[1]), as a marker of online performance monitoring to assess its scope and selectivity. Two groups of 20 participants

[1] As will be evident later, the terms "Error Negativity" or "Error Related Negativity" are clearly inappropriate. We so far keep these terms as they are the ones used in the literature.

performed exactly the same between-hand two-choice Reaction Time (RT) task (orientation discrimination of tilted Gabor patches) with right or left thumb key presses (on force transducers) for clockwise and counterclockwise tilts, respectively. After every trial, participants had to provide a metacognitive evaluation of their performance (Fig. 1). However, the nature of this metacognitive evaluation differed between the two groups: one group was asked to evaluate how long they took to provide their response (*i.e.* estimate the duration of their Reaction Time – RT-group), while the other one was asked to evaluate the force they produced on the response button to provide their response (Force-group). While clear variability was observed on both RT and produced force in both groups (see Supplementary Figure 1), metacognitive instruction made one of these variabilities relevant, enabling to study how the performance monitoring system responds to this relevance. Electroencephalographic activity was recorded with 64 electrodes (see method), and Current Source Density was computed to improve the spatial [25, 26] and temporal [27, 28] resolutions of the EEG signal. Electromyographic activity of the prime mover muscles involved in the thumb press (*flexor pollicis brevis*) was also recorded to better time-lock the evoked EEG activity.

Results

Behavioral results essentially replicate data already observed [29], and are reported in Extended Data Figures 1 and 2.

Sensitivity and selectivity of MPFC activity

ERP analysis

In the following analyses, we focused on correct trials (see Extended Data Figure 3 for the classical effect of errors vs corrects). For EEG analyses, for each participant of the two groups, correct trials were sorted both by RT (from fast to slow) and by the force produced (from weak to strong), only one being relevant for the ongoing subjective judgment.

In a first analysis, data were binned in 5 categories and standard averaged ERP, time-locked to the response EMG onset, were computed for each bin (Fig. 2) to extract Ne amplitude. In the Force-group, Ne amplitude varied as a function of force ($F(4,72) = 9.23, p < .001, \eta_p^2 = .34$), but not as a function of RT ($F(4,72) = .76, p = .55$) while the opposite effect was obtained in the RT-group: Ne amplitude did not vary as a function of force ($F(4,76) = .11, p = .98$) but did as a function of RT ($F(4,76) = 2.82, p = .08, \epsilon = .46, \eta_p^2 = .13$), albeit marginally. Close inspection of Fig. 2 suggests different relationships between Ne amplitude and magnitude of the evaluated dimension (fast-slow *vs.* weak-strong): Ne amplitude and force seem to be linearly related in the Force-group, while Ne amplitude seems to follow a U-shape with RT in the RT-group.

To clarify the marginal effect of RT in the RT-group, and to better describe the relationships between Ne amplitude and evaluated dimensions, we ran mixed models with mean Ne amplitude as a dependent variable and either mean force or mean RT and mean RT-squared (to account for the potential nonlinear relationship with RT) as predictors for both groups. In the Force-group, force predicted Ne amplitude

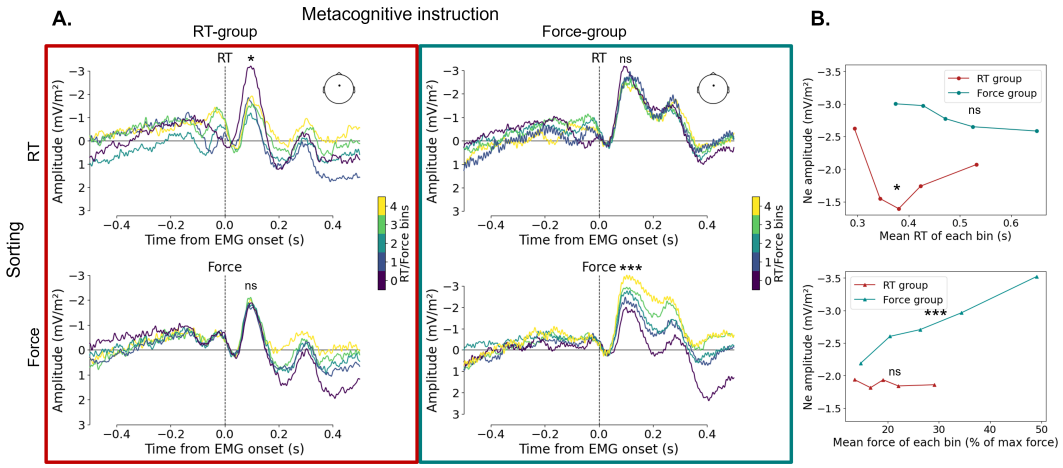


Fig. 2 (A) Grand averages per RT (top row) and force (bottom row) bins for RT-group (red border) and Force-group (blue border). (B) Calculated amplitude is plotted as a function of mean RT per bin (top) and mean force per bin (bottom). Red line represents data of RT group and blue line data of force group. $^{*}p < 0.1$, $^{***}p < 0.001$.

($\beta = -.029, t = -5.3, p < .001$), but not RT ($\beta = -.39, t = -.062, p = .95$) nor RT-squared ($\beta = 1.96, t = .34, p = .74$). Conversely, in the RT-group, both RT ($\beta = 22.38, t = 2.17, p = .03$) and RT-squared ($\beta = -26.23, t = -2.27, p = .023$) predicted Ne amplitude, but not force ($\beta = -.004, t = -.36, p = .72$). [2]

Grand averaged ERP image

The above results are based on a somehow arbitrary categorization of data (five bins), and on an averaging across trials, known to potentially distort the signal [30]. To better assess the relationships between Ne amplitude, force and RT, and to consolidate the results beyond five-class averaging, we resorted to a much finer grain analysis [31]: For each participant of each group, we built two ERP-images [32], where each line corresponds to a trial (see below) and the CSD activity is color-coded; in one of them, trials were sorted by increasing RT, while they were sorted by increasing force in the other. The individual ERP-images were then combined to get a "grand-averaged ERP-image" of around 380 lines for each group and dimension, each line being a grand-averaged 'pseudo trial' (see methods for details).

For each grand-average, Ne amplitude was then measured, and we ran regression models with Ne amplitude as a dependent variable and RT or force as a predictor for both groups. This finer grain analysis confirmed and extended the results observed on 5 bins (Fig. 3). In the Force-group, force linearly predicted MPFC activity, with increased Ne amplitude for stronger forces ($\beta = -.04, t = -8.03, p < .001$); no relationship with force was observed in the RT-group ($\beta = .003, t = .33, p = .74$). In contrast, in the RT-group, both RT ($\beta = 27.43, t = 8.09, p < .001$) and RT-squared ($\beta = -28.59, t = -7.67, p < .001$) predicted MPFC activity, indicating a decrease in Ne amplitude with increasing RT, although non-linear,

[2] The significant β coefficients of the force model and the RT model have different orders of magnitude due to: 1) the force model being linear and the RT model being quadratic, and 2) the regressors having different scales (< 1s for RT, < 100% for force)

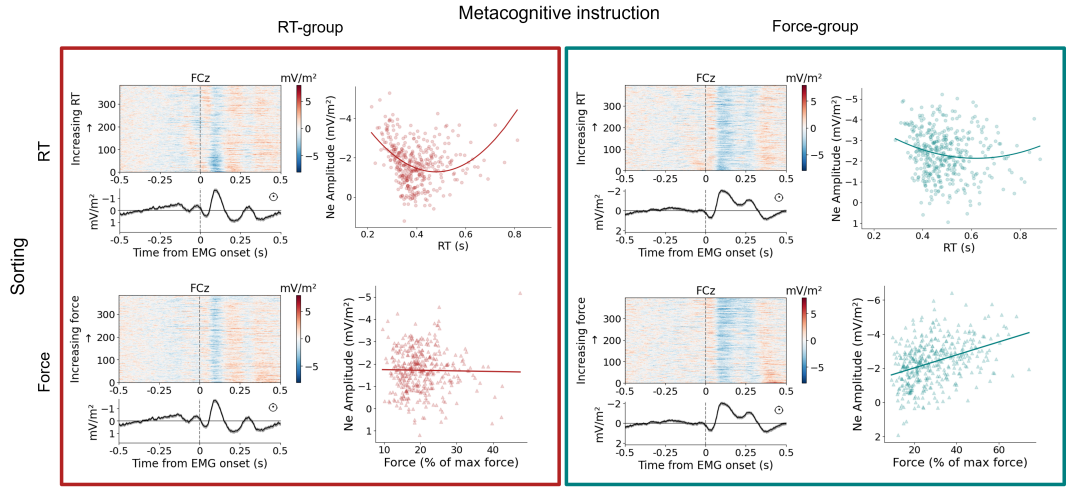


Fig. 3 Grand average ERP image sorted by RT (top) and force (bottom) of RT-group (left) and Force-group (right). Epoch 0 corresponds to shortest RT/weakest force. For each group, extracted Ne amplitude is plotted as a function of RT with fitted quadratic function (top) and force with fitted linear function (bottom).

with higher amplitude for faster and slower RTs. In the Force-group, a similar pattern was observed, although largely reduced ($\beta = 10.66, t = 2.66, p = .008$ for RT, and $\beta = -8.6, t = -2.35, p = .02$ for RT-squared).

Overall, although the two groups were performing exactly the same first-order task inducing important variability in both RT and force, MPFC monitoring activity was largely modulated by variations in the dimension relevant for the metacognitive judgment, but was not sensitive (or weakly) to the irrelevant one.

Discussion

Since the pioneering works of P. Rabbit on error detection [33], performance monitoring literature has largely been focused on the contrast between correct and erroneous responses. This was further reinforced by the discovery of the Ne, initially considered to be specific to (choice) errors [7, 8] and interpreted as an error detection mechanism. This led to undermine, or even ignore, any possible performance monitoring on correct trials, almost confining those trials to mere controls for errors. However, some pieces of evidence already suggested that erring is not as binary as previously thought [34, 35] and that MPFC is sensitive to error gradation [4, 13, 36]. Few studies even suggested that it can be modulated by more continuous forms of errors, besides response-choice errors [15–17].

The present results go far beyond by highlighting the richness of performance monitoring on **correct** trials, prompting a comprehensive re-interpretation of both performance monitoring in general and the role of MPFC in this process. They reveal parametric modulation of MPFC activity by the intrinsic

timing and execution variability of correct actions, extending previous observation of modulation by RT [18, 19, 23]. Such modulation, however, was only present when the evaluated dimension was relevant for the current context, demonstrating the versatility of the monitoring system. We show that both time and force execution are trackable by MPFC and, although speculative, we can bet that other dimensions of action would also be. Probing which dimensions are and are not trackable by MPFC activity, along with the conditions under which they are tracked, will provide crucial insights into the role of this structure in performance monitoring and the computational processes that underlie it.

Although MPFC activity showed a clear sensitivity to the two analyzed dimensions, a linear relationship was observed between MPFC activity and force, while it had a quadratic one with time. This U-shape indicates that there is a RT value for which MPFC activity is minimal, and that deviations from this value increase MPFC activity. Although the precise location of this minimum is difficult to determine, it seems to be close to the maximal density of the RT distribution, which may serve as a reference point to which all RTs are compared. Measuring the deviation from a predicted reference value is at the core of "prediction error" concept, that has been proposed as a general principle of brain functioning [37]. Models of MPFC role in performance monitoring based on this principle generally focus on reward prediction error [38–40], comparing received outcome to a predicted one. Some extentions forecast a similar prediction error for expected actions occurring at an unexpected time [41]. The current RT findings are in line with this view, if one assumes that MPFC measures the difference between the current RT and an internally built "expected one". In such case, this means that "prediction error" is a core aspect of performance monitoring that extends far beyond reward and error processing.

However, if this extended unsigned prediction error is the core computation performed by MPFC, why is the relationship with force linear instead of U-shaped? Note that a U-shape should be observed if the produced forces are centered on the expected value. In contrast, if they are all above (or below) this expected value, a monotonic, linear relationship is to be expected. An obvious choice for the expected value, to which all produced forces should be compared, is the threshold for the response to be recorded (see Methods for details). In such case, the forces actually produced during correct trials are always above this expected force, leaving only the right, linear, part of the relationship. One prediction, though, is that MPFC activity should also be larger when the force produced is below the threshold. Although the present dataset does not allow to evaluate this possibility (too few trials in which the force did not reach the threshold), previous studies [21, 22] reported this exact result, supporting this hypothesis. Hence, MPFC appears to monitor actions variability by comparing the actual performance to a predicted value.

The present data reveal interesting properties of this general purpose mechanism. First, the performance monitoring competence of the MPFC extends far beyond mere "failure" detection and can provide a much richer information. This leads to a new vista on the sensitivity of MPFC to errors: while for long considered as a key feature, it might only be one case among others. It might, however, be based on the same computational principles as on correct trials [40], with errors representing deviations from the expected correct response [38]. Second, the present data demonstrate the capacity of MPFC to selectively

track some sources of variance and ignore others. Indeed, while both time and force **can** be tracked by the monitoring system, they only are when relevant for the task. This could be the reason why errors have received so much interest: likely because accuracy is most of the time a dimension emphasized in experimental protocols, making errors particularly relevant. Hence, much like the attentional system, the performance monitoring system appears to possess both a sustained component, with a global increase in activity under high monitoring demand (*i.e.* aspecific increased efficiency, [24]), and a focused component, enabling it to target specific, relevant, and salient aspects of performance that require monitoring. Third, the way the expected value is built is dependent on the ongoing context as well and hence appears very versatile. For the Force-group, the expected (force) value is explicitely provided by the instruction; although of very different nature, this is also likely the case for choice errors, where an explicit instruction defines the correct response. In contrast, for the RT-group, it is built internally based on experience, which is also the case in reinforcement-learning protocols. Importantly, in all those cases, MPFC performs exactly the same operation (computing prediction error between an expected value and an actual one), but applied to very different inputs, which endows it with a highly generic performance monitoring function.

Overall, the present results also provide a new view on the so called Ne/ERN/CRN: it is neither related to errors, nor to correct responses, but appears as a generic performance monitoring activity, and should hence better be termed "Performance Related Negativity", PRN. This also has direct impact on the use of this activity as a biomarker in psychopathology (see *e.g.* [42]): if it is, it might not reflect exactly, or not only, what has been assumed so far (*i.e.* error processing deficit), maybe requiring a revision of the nature of the degraded processes in such pathologies.

Methods

Participants

Forty-four participants were recruited from the Aix-Marseille University community. Ne/ERN is a very robust signal, that allows to acheive a high statistical power (.8) with relatively few trials (down to 8 trials) and rather low sample (down to 16 participants, see [43, 44]). Although we have a much larger number of trials (above 300, see below), we increased the sample size to 20 per group. 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" (n° 2021-A01548-33). Participants gave their written consent and were compensated at a rate of 15 € per hour. Three participants were excluded due to a high error rate (>50%), and one because of bad quality of EEG recordings. Note that this exclusion was decided before any EEG data processing, to avoid biasing results. Another participant was excluded since in about 26% of trials, the produced force exceeded the maximum voluntary force (MVF), indicating that this participant did not follow the instruction to press as strong as possible during the measure of this MVF (see below). In the end, data of 39 participants (mean age = 25.33) were included in the analyses. Twenty-four reported their gender as female and fifteen as male. Eight were left-handed. They were divided into two groups who performed the same choice reaction time task with two different metacognitive instructions. Participants in the RT-group (N=20, mean age = 23.95, 13 females and 7 males, 3 left-handed) had to evaluate their reaction times, and the ones in the Force-group (N=19, mean age = 26.79, 11 females and 8 males, 5 left-handed) had to evaluate their produced forces (see below).

Apparatus

Participants were comfortably seated in front of a computer screen (frame rate: 120Hz) placed at a one-meter distance, within a Faraday cage. All signals (force, EMG, EEG) were recorded using an Active-2 Biosemi system (BioSemi Instrumentation, Amsterdam, the Netherlands) with a sampling frequency set at 1024 Hz (analog low pass bandwidth limit: -3 dB at 1/5th of the sampling rate). Responses were given by pressing either a left or a right button with the corresponding thumb. Buttons were placed on two cylinders mounted on force sensors allowing to continuously measure the force produced and to adjust the response threshold (*i.e.* the amount of force needed for a response to be recorded, see below). Crossing this threshold determined the behavioral response (button press) and defined the RT. At button press, participants heard a 3ms sound feedback at 400 Hz (resembling a small click). The left cylinder was mounted on a rotating device allowing to tilt the cylinder to the right (towards the center of the device) up to 45 degrees (see Fig. 1). By default, the cylinder was locked in a vertical position by a magnet located at the bottom end of the cylinder, so that it was identical to the right one. When the magnet was off, the button could be freely tilted to move 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 electromyographic (EMG)

activity of the flexor pollicis brevis of both hands with two flat Ag/AgCl electrodes placed on each thenar eminences. 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. The EEG activity was recorded using a Biosemi cap including 64 Ag/AgCl electrodes (pin-type active electrodes). A 65th electrode was placed on the right mastoid to reference the signal offline. Electrodes for vertical electrooculogram (VEOG) were at Fp2 and below the right eye, and electrodes for horizontal electrooculogram (HOEG) were at the outer canthi of the left and right eyes. All participants performed one single 2-hours session including one training block and 10 experimental blocks of a choice reaction time task, separated by self-paced breaks.

Stimuli

Stimuli presentation was controlled by the Psychopy software [45]. Stimuli were a single Gabor patch presented in the middle of the screen. It had a spatial frequency of 0.5 cycle/visual angle (in degree) and its size was 15° visual angle. Its contrast was set to 0.5 on a scale from 0 (uniform gray) to 1 (black and white). The Gabor patches could be tilted, either to the left (counterclockwise) or to the right (clockwise), by either 1 or 7° from vertical, to increase variability in RT and force. Subjective reports of reaction times (iRT) and produced force (iForce) were collected via a horizontally presented visual analog scale (VAS), with the cursor moved using the left response button once released (see above). Five ticks were present one this scale: two at each end and 3 separating it into 4 equal segments.

Maximum voluntary force

Before the experimental tasks, "Maximum Voluntary Force" (MVF) was measured for each participant, to individually adapt the response threshold. Participants were requested to press as strong as possible on the force sensors. They were asked to alternate 3 left- and 3 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 of the three presses was retained for each hand separately as their hand_MVF. To avoid tiring the weakest hand, the lowest hand_MVF of the two hands was retained as the participant MVF. The force defining the response threshold was set individually at 10% of the participant MVF.

Choice reaction time task

Participants performed the orientation detection task and had to estimate their reaction times or their produced force, depending on their group, after each trial (see Fig. 1). Each trial began with the presentation of a fixation cross in the center of the screen. After 1 second, the stimulus replaced the cross. Stimuli were presented for a constant duration of 150 ms. Stimulus angle was pseudo-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. Within each trial, when a response was given, the VAS 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 the RT-group, participants were asked to precisely evaluate their reaction time, *i.e.*, 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 reaction time (iRT). Left and right end of the scale corresponded to 0 and 1.33 seconds respectively (so that the fourth tick corresponded to 1 second), but this information was not given to the participants. In the Force-group, participants were asked to precisely evaluate the force applied on the button to give the response on the current trial. For that, participants were asked to place the marker at the position corresponding to their estimated produced force (iForce). The scale was visually the same as for RT evaluation. Left end of the scale corresponded to the force threshold (*i.e.* 10% of their MVF) and the equivalent of the one second tick was set to 60% of MVF, but this information was not given to the participants. During the instruction, they were told to judge by how much they thought they had exceeded the threshold for each response. In both cases, they had to validate the evaluation with a press on the left button. The force threshold for this validation was set constant at 6N (600g). 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 second, the same visual scale, but greyed, was presented and the left button remained locked so that no interaction was possible. The scale was displayed for 1.25 s before the next trial. Inter-trial 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. Before the experimental blocks, participants performed 24 training trials, to make sure that they understood the task properly. During this training block, feedback on produced RT or force was presented on the scale after each evaluation in the form of a green triangle marker. Once trained, participants performed 10 blocks of 60 trials.

EMG and EEG preprocessing

EMG signal was processed with MYOnset (Spieser & Burle, in preparation) a custom-made python program to detect EMG onsets (released under open-source license, available at https://github.com/lspieser/myonset). The signal was first high pass filtered at 10Hz, then for each trial, the mean (μ) and standard deviation (σ) of the signal baseline was computed from 0.5 s preceding stimulus onset, for each hand separately. On the post-stimulus 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 optimized for each participant). 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 [46]. 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, trial was not marked and excluded from analyses.

EEG signal was processed using MNE python [47]. First, data were re-referenced to the right mastoid and signal was low pass filtered at 100 Hz. Blinks were corrected using spatial projector computed from epochs around eyeblinks detected on VEOG. Other artifacts were rejected with visual inspection of the signal. The response device described above, for unknown reasons, induced some systematic noise on some participants around the CMS/DRL electrodes (the noise was completely absent when the device was unplugged). The noise was characterized by a 50Hz frequency and a parieto-occipital topography. For some participants, one or two electrodes (among POz, P2, PO3 and PO4) were too noisy and had to be interpolated. POz was interpolated for 5 participants, P2 for 7, PO3 for 1 and PO4 for 1. Both P2 and POz had to be interpolated for 3 participants. This noise being spatially well defined, and with a very stable frequency (see Fig. 4), ICA proved very efficient to capture it, and hence to remove it. Time course, topography and frequency of components were visually inspected by the experimenter to identify the one that best captured the noise.

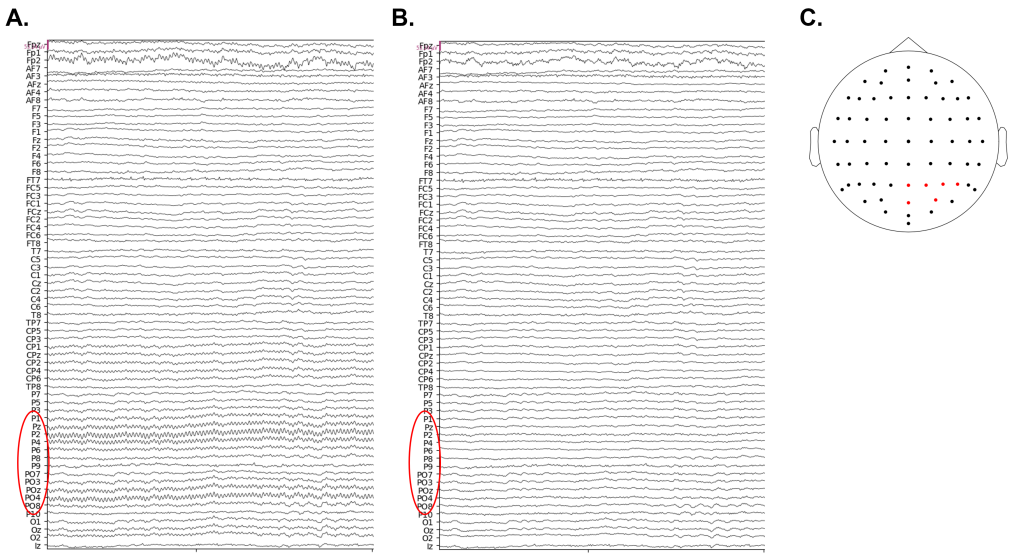


Fig. 4 (A) Raw EEG signal with noise related to the response device. (B) EEG signal after removal of the noise with ICA. (C) Parieto-occipital noisy electrodes are indicated in red.

Epochs from -0.5 to 0.5 seconds around EMG onset were extracted. Current Source Density (CSD) was then estimated by computing the Surface Laplacian of the scalp potentials [48]. We used the default parameters of the CSD computation of MNE-python: the smoothness parameter was set to $1e^{-5}$, the

degree of spline to 4 and the interpolation was computed with a maximum of 50 degrees for the Legendre polynomial. Compared to scalp potentials, CSD dramatically improves spatial [25, 26] but also, secondarily, temporal resolution [27, 28].

Data analysis

Force peaks were extracted for each trial and values were normalized by the MVF of each participant and are hence expressed in percentage of this MVF. The first trial of each block was removed. 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 minus 2 mad and superior to the median plus 5 mad were then excluded. In total, 2.93% of trials were excluded in the RT group and 2.89% in the force group (see [29]).

The main analyses were restricted to correct trials with only one EMG burst. Among all non-rejected trials, those single EMG burst trials represented 84% in the RT-group and 90% in the Force-group. Those trials will simply be termed "correct trials" herein. EEG epochs containing artifacts were rejected (9.5% in RT-group and 15% in Force-group).

For both groups, we analyzed MPFC activity (measured through Error Negativity amplitude) as a function of both RT duration and produced force. All EEG activities were time-locked to the onset of the EMG that led to the behavioral response. To do so, for each participant, all included epochs were averaged time-locked to the EMG onset to assess the Ne characteristics. We identified the electrode that best captured the activity in terms of amplitude and signal-to-noise ratio. FCz was the chosen electrode for most participants. For 4 of them, 2 neighboring electrodes were preferred, namely Cz and Fz. Note that this choice was performed on an average of all trials, hence without any reference to any sensitivity to RT or Force. Onset and peak indices were extracted from these averages, as depicted on Fig. 5.

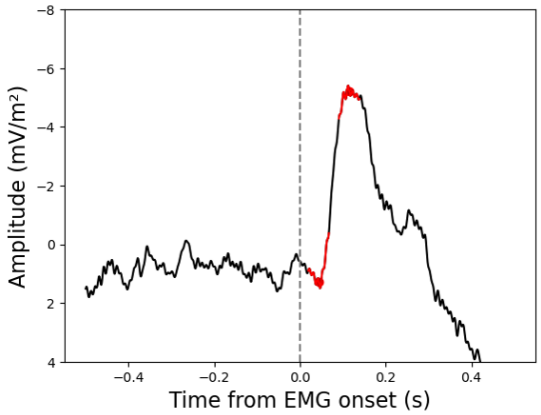


Fig. 5 Example of a single participant ERP at electrode FCz. Red points represent onset and peak of the negativity and red lines are the windows of analysis for peak-to-peak amplitude calculation.

In a first set of analyses, RT and force individual distributions were divided into 5 bins from shortest RT/weakest force (bin 0) to longest RT/strongest force (bin 4), each containing the same number of trials (see Fig. 6). On average, each bin comprised 76 epochs in the RT-group and 80 epochs in the Force-group. Epochs within each bin were averaged and Ne amplitude was calculated as the difference between the average value of 50 points around peak and 50 points around onset (peak-to-peak amplitude), which indices were determined for each individual (see above). All statistical analyses were done using Python packages Pingouin [49] and Statsmodel [50], and R package lme4 [51].

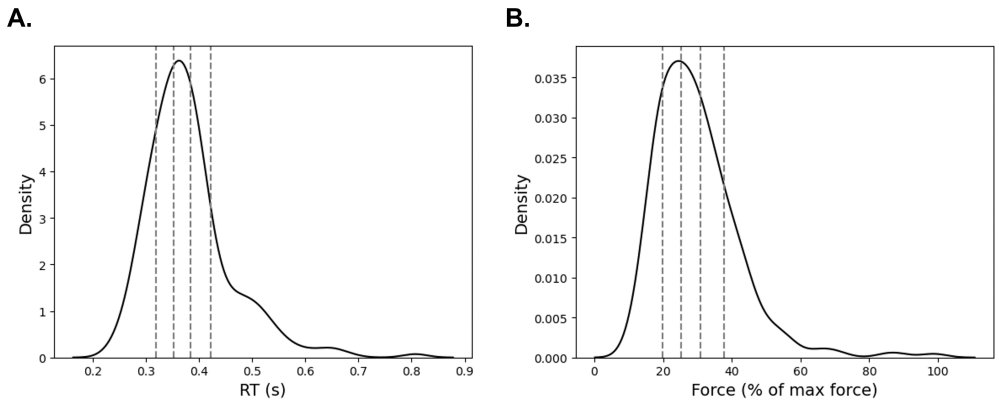


Fig. 6 Examples of a single participant's distributions of RT (A) and force (B). Dashed vertical lines represent bins limits, ensuring equal trials number within each five bins. The curve is normalized so that the integral over all possible values is 1, meaning that the scale of the density axis depends on the data values.

Analyses of variance (ANOVAs) were used to compare mean Ne amplitudes of the different bins for each dimension (RT and force). When the condition of sphericity was violated, Greenhouse-Geisser correction was applied. When needed, as the sample sizes were unequal between groups, the Welch–Satterthwaite equation was used to approximate the adjusted degrees of freedom. In addition, to better take into account the ordered nature of bins, we ran mixed models with mean RT or force values of each bin as regressors and used random intercepts for participants.

However, averaging results in a significant loss in capturing the dynamics of the studied process, as it implies categorizing continuous data with arbitrary boundaries. In a subsequent step, we sought to explore the relationship between Ne amplitude and RT/force with a much finer grain analysis. To this aim, we extracted Ne amplitudes from a "grand-average ERP-images" [31]. This was performed in different steps (see Fig. 7):

1. Constructing the single participant's ERP-image (Fig. 7, row 1): individual EEG epochs were sorted based on the relevant measure (RT or force) and arranged in ascending order in a matrix, so that each line corresponds to a trial. Those matrices can be represented as images (ERP-images, [32]), with

the amplitude of the CSD being color coded. For each participant p, the number of trial included N_p differs. Hence the matrices representing these ERP-image are of dimension 1024 columns $\times$ N_p lines, with 1024 corresponding to the number of samples in a 1s-epoch centered on EMG onset.

2. Interpolating the individual ERP-image to normalize their size (Fig. 7, row 2): in order to be able to average all individual ERP-images, they must have the same dimensions. To reach this condition, we interpolated (sub- or over-sampling) the individual ERP-images, using the function `interp1d` from the `scipy.interpolate` Python package (version 1.9.1) so that all new individual ERP-images are of dimension $1024 \times N_{ga}$, with N_{ga} being the new number of "pseudo trials". We opted to align N_{ga} to the median of the trials number across all participants, to globally minimize the individual ERP-image modifications, resulting in 384 trials for the RT-group and 397 trials for the Force-group.
3. Grand-averaging of individual ERP-images (Fig. 7, row 3): all ERP-images for all participants in a given condition having the same size, they can be directly averaged on a point-by-point manner. We hence obtained N_{ga} grand averaged "pseudo trials", organized in a matrix of dimension $1024 \times N_{ga}$ that forms a grand-averaged ERP-image. For each line ("pseudo trials"), the amplitude of the Ne was extracted as described above for the standard averages.

To quantify the link between these measures and Ne amplitude, we performed linear regressions with (interpolated) RT or force as predictors. To test whether a non-linear association would improve the models' fit, we first compared these models with models containing additional quadratic terms (RT- or force-squared), with likelihood ratio test. For both groups, quadratic model was preferred for RT (RT-group:$\chi^2(1) = 55.15, p < 0.001$, Force-group:$\chi^2(1) = 5.5, p = 0.02$) but not for force (RT-group:$\chi^2(1) = 0.008, p < 0.93$, Force-group:$\chi^2(1) = 2.4, p = 0.12$).

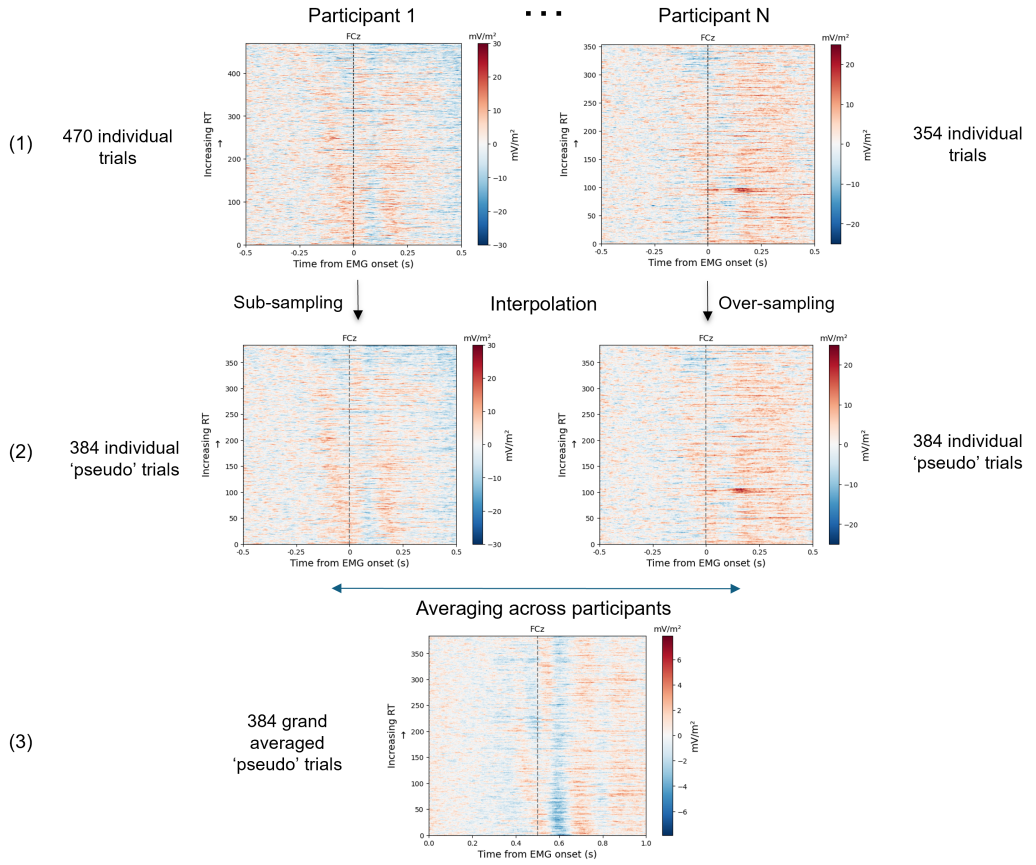


Fig. 7 Method to obtain a grand averaged ERP image. Individual ERP-images are first built by sorting epochs by increasing RT or force (first row). Each individual ERP-image is then interpolated to get the same number of 'pseudo trials' for each participant (second row). Finally, these individual ERP-images are averaged across participants (third row).

Extended Data Figures

Effect of task difficulty

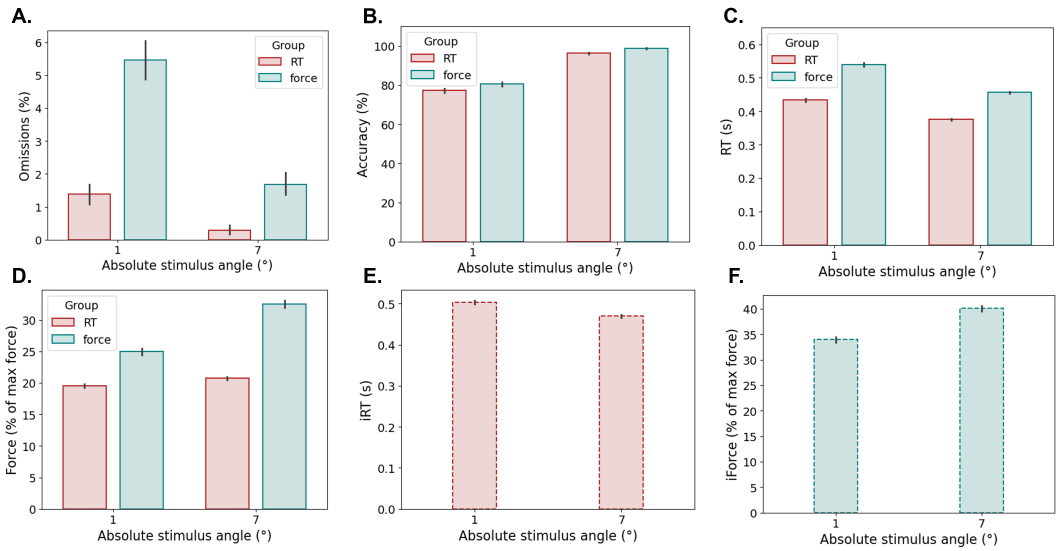


Fig. 1 (A) Omissions were observed on 0.8% of trials in the RT-group and 3.6% of trials in the Force-group. Omission rate was higher for lower stimulus angle (*i.e.* highest difficulty) in both groups ($t(19) = -5.55, p < .001$ in RT-group and $t(18) = -7.59, p < .001$ in Force-group). (B) Accuracy was comparable between the two groups (86.6% in the RT-group, 89.7% in the Force-group). It was lower in the difficult condition in both of them ($t(19) = 17.28, p < .001$ in RT-group and $t(18) = 14.2, p < .001$ in Force-group). (C) Correct trials' RT were longer for higher task difficulty ($t(19) = -7.56, p < .001$ in RT-group and $t(18) = -12.4, p < .001$ in Force-group). (D) Correct trials' produced forces were lower for higher difficulty ($t(19) = 3.41, p = .003$ in RT-group and $t(18) = 4.9, p < .001$ in Force-group). (E) In the RT-group, iRT was higher in the difficult condition ($t(19) = -3.97, p < .001$). (F) In the Force-group, iForce was higher in the easy condition ($t(18) = 4.42, p < .001$). The error bars indicate the 95% confidence intervals.

Correlations between objective and subjective measures

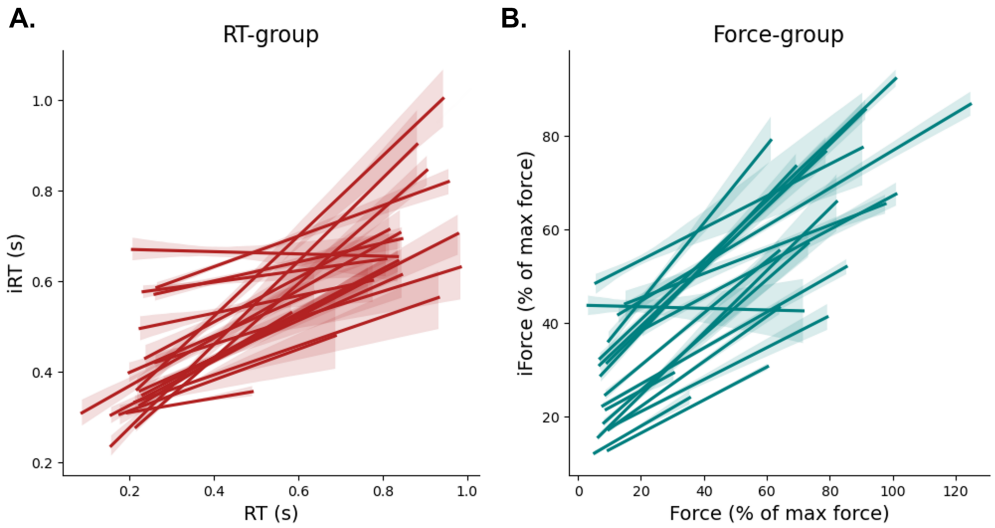


Fig. 2 Regression lines for each participant, illustrating the relationships between RT and iRT (A) and Force and iForce (B). iRT correspond to the subjective reports of RT on the VAS and iForce to the subjective reports of produced force. The shaded areas indicate the 95% confidence intervals around the regression lines. In the RT-group, a repeated measure correlation [52] revealed a positive correlation between RT and iRT ($r_{rm}(8320) = .44, p < .001$). Pearson's correlation coefficients were also computed for each participant individually, and all but one presented a positive correlation (range=-.02–.82, mean=.44, std=.21). In the Force-group, Force and iForce were also positively correlated ($r_{rm}(8671) = .65, p < .001$). As in RT-group, all participants but one presented a positive Pearson correlation (range=-.012–.88, mean=.62, std=.23).

Effect of accuracy on Ne amplitude

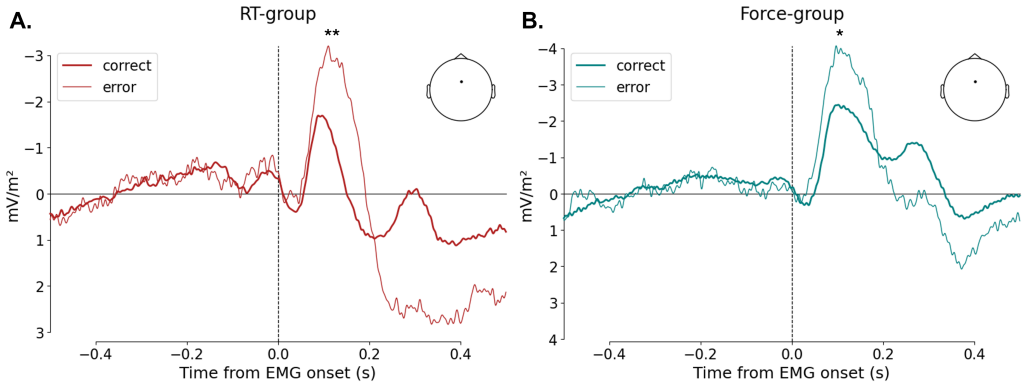


Fig. 3 Grand averages for correct (thin lines) and error (thick lines) trials at electrode FCz for (A) RT-group and (B) Force-group. Amplitude for error trials was higher than correct trials in both RT-group ($t(19) = 2.59, p = 0.02$) and Force-group ($t(18) = 2.04, p = 0.056$). *$p < 0.1$, **$p < 0.05$

Acknowledgments

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

The authors wish to thank D. Louber and D. Parellesompoulé for developping the response device and F.-X. Alario for fruitful discussions and advices.

A CC-BY public copyright license has been applied by the authors to the present document and will be applied to all subsequent versions up to the Author Accepted Manuscript arising from this submission, in accordance with the grant's open access conditions.

Funding statement

N.P. is funded by a PhD grant from the french ministery for research. This work has received support from the French government under the Program "Investissements d'Avenir", Initiative d'Excellence d'Aix-Marseille Université via A*Midex and ANR fundings (AMX-19-IET-004 and ANR-17-EURE-0029), and by the ANR (grant number: ANR-22-CE28-0027).

Authors contributions:

- NP: conceptualization, data curation, formal analysis, investigation, methodology, software, vizualisation, writing - original draft, writing - review & editing
- FV: methodology, validation, writing - review & editing
- BB: conceptualization, data curation, formal analysis, funding acquisition, methodology, projet administration, software, validation, vizualisation, writing - original draft, writing - review & editing

Competing interests statement

The authors declare no competing interests.

References

[1] Bruijning, M., Metcalf, C. J. E., Jongejans, E. & Ayroles, J. F. The evolution of variance control. *Trends in Ecology & Evolution* **35**, 22–33 (2020).

[2] Reason, J. *Human Error* (Cambridge University Press, 1990).

[3] Ridderinkhof, K. R., Ullsperger, M., Crone, E. A. & Nieuwenhuis, S. The role of the medial frontal cortex in cognitive control. *Science* **306**, 443–447 (2004).

[4] Bonini, F. *et al.* Action monitoring and medial frontal cortex: leading role of supplementary motor area. *Science* **343**, 888–891 (2014).

[5] Carter, C. S. *et al.* Anterior cingulate cortex, error detection, and the online monitoring of performance. *Science* **280**, 747–749 (1998).

[6] Fu, Z. *et al.* Single-Neuron Correlates of Error Monitoring and Post-Error Adjustments in Human Medial Frontal Cortex. *Neuron* **101**, 165–177.e5 (2019). URL https://linkinghub.elsevier.com/retrieve/pii/S0896627318310079.

[7] Falkenstein, M., Hohnsbein, J., Hoormann, J. & Blanke, L. Effects of crossmodal divided attention on late ERP components. II. Error processing in choice reaction tasks. *Electroencephalography and Clinical Neurophysiology* **78**, 447–455 (1991).

[8] Gehring, W. J., Goss, B., Coles, M. G. H., Meyer, D. E. & Donchin, E. A neural system for error detection and compensation. *Psychological Science* **4**, 385–390 (1993).

[9] Coles, M. G., Scheffers, M. K. & Holroyd, C. B. Why is there an ERN/Ne on correct trials? Response representations, stimulus-related components, and the theory of error-processing. *Biological Psychology* **56**, 173–189 (2001).

[10] Vidal, F., Burle, B. & Hasbroucq, T. On the comparison between the Nc/CRN and the Ne/ERN. *Frontiers in Human Neuroscience* **15** (2022).

[11] Endrass, T., Klawohn, J., Schuster, F. & Kathmann, N. Overactive performance monitoring in obsessive-compulsive disorder: ERP evidence from correct and erroneous reactions. *Neuropsychologia* **46**, 1877–1887 (2008).

[12] Roger, C., Bénar, C. G., Vidal, F., Hasbroucq, T. & Burle, B. Rostral Cingulate Zone and correct response monitoring: ICA and source localization evidences for the unicity of correct- and error-negativities. *Neuroimage* **51**, 391–403 (2010).

[13] Vidal, F., Hasbroucq, T., Grapperon, J. & Bonnet, M. Is the 'error negativity' specific to errors? *Biological Psychology* **51**, 109–128 (2000).

[14] Vidal, F., Burle, B., Bonnet, M., Grapperon, J. & Hasbroucq, T. Error negativity on correct trials: a reexamination of available data. *Biological Psychology* **64**, 265–282 (2003).

[15] Anguera, J. A., Seidler, R. D. & Gehring, W. J. Changes in Performance Monitoring During Sensorimotor Adaptation. *Journal of Neurophysiology* **102**, 1868–1879 (2009). URL https://www.physiology.org/doi/10.1152/jn.00063.2009.

[16] Vocat, R., Pourtois, G. & Vuilleumier, P. Parametric modulation of error-related ERP components by the magnitude of visuo-motor mismatch. *Neuropsychologia* **49**, 360–367 (2011).

[17] Luu, P., Flaisch, T. & Tucker, D. Medial Frontal Cortex in Action Monitoring. *Journal of Neuroscience* **20**, 464–469 (2000).

[18] Valt, C. & Stürmer, B. On the correct side of performance: Processing of internal and external signals in response speed evaluation. *International Journal of Psychophysiology* **117**, 26–36 (2017).

[19] Valt, C. & Stürmer, B. On the processing of optimal performances: Studying arousal evoked by being correct and fast. *Brain and Behavior* **11**, e02162 (2021). URL https://onlinelibrary.wiley.com/doi/10.1002/brb3.2162.

[20] De Bruijn, E. R. A., Hulstijn, W., Meulenbroek, R. G. J. & Van Galen, G. P. Action monitoring in motor control: ERPs following selection and execution errors in a force production task. *Psychophysiology* **40**, 786–795 (2003). URL https://onlinelibrary.wiley.com/doi/abs/10.1111/1469-8986.00079.

[21] Meckler, C., Carbonnell, L., Ramdani, C., Hasbroucq, T. & Vidal, F. On-line action monitoring of response execution: An electrophysiological study. *Biological Psychology* **129**, 178–185 (2017).

[22] Ramdani, C., Hasbroucq, T. & Vidal, F. Why is there an error negativity on correct trials? A reappraisal. *Neuroscience Letters* **828**, 137731 (2024). URL https://www.sciencedirect.com/science/article/pii/S0304394024001083.

[23] Files, B. T., Pollard, K. A., Oiknine, A. H., Khooshabeh, P. & Passaro, A. D. Correct response negativity may reflect subjective value of reaction time under regulatory fit in a speed-rewarded task. *Psychophysiology* **58**, e13856 (2021). URL https://onlinelibrary.wiley.com/doi/abs/10.1111/psyp.13856. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/psyp.13856.

[24] Grützmann, R., Endrass, T., Klawohn, J. & Kathmann, N. Response accuracy rating modulates ERN and Pe amplitudes. *Biological Psychology* **96**, 1–7 (2014).

[25] Nunez, P. *et al.* A theoretical and experimental study of high resolution EEG based on surface Laplacians and cortical imaging. *Electroencephalography and Clinical Neurophysiology* **90**, 40–57 (1994). URL https://www.sciencedirect.com/science/article/pii/0013469494901120.

[26] Babiloni, F. *et al.* Performances of surface Laplacian estimators: a study of simulated and real scalp potential distributions. *Brain Topography* **8**, 35–45 (1995).

[27] Law, S. K., Rohrbaugh, J. W., Adams, C. M. & Eckardt, M. J. Improving spatial and temporal resolution in evoked EEG responses using surface Laplacians. *Electroencephalography and Clinical Neurophysiology* **88**, 309–322 (1993).

[28] Burle, B. *et al.* Spatial and temporal resolutions of EEG: Is it really black and white? A scalp current density view. *International Journal of Psychophysiology* **97**, 210–220 (2015). URL http://dx.doi.org/10.1016/j.ijpsycho.2015.05.004.

[29] Pavailler, N., Gevers, W. & Burle, B. Temporal metacognition: direct readout or mental construct? the case of introspective reaction time. *Journal of Experimental Psychology: General* **in press** (2024).

[30] Burle, B., Roger, C., Vidal, F. & Hasbroucq, T. Spatio-temporal dynamics of information processing in the Brain: Recent advances, current limitations and future challenges. *International Journal of Bioelectromagnetism* **10**, 17–21 (2008).

[31] Burle, B., Roger, C., Allain, S., Vidal, F. & Hasbroucq, T. Error negativity does not reflect conflict: a reappraisal of conflict monitoring and anterior cingulate cortex activity. *Journal of Cognitive Neuroscience* **20**, 1637–1655 (2008).

[32] Jung, T. *et al.* Analysis and visualization of single-trial event-related potentials. *Human Brain Mapping* **14**, 166–185 (2001).

[33] Rabbitt, P. M. Error correction time without external error signals. *Nature* **212**, 438 (1966).

[34] Coles, M. G., Gratton, G., Bashore, T. R., Eriksen, C. W. & Donchin, E. A psychophysiological investigation of the continuous flow model of human information processing. *Journal of Experimental Psychology: Human Perception and Performance* **11**, 529–553 (1985).

[35] Burle, B., Possamai, C.-A., Vidal, F., Bonnet, M. & Hasbroucq, T. Executive control in the Simon effect: an electromyographic and distributional analysis. *Psychological Research* **66**, 324–336 (2002). URL http://dx.doi.org/10.1007/s00426-002-0105-6.

[36] Scheffers, M. K., Coles, M. G., Bernstein, P., Gehring, W. J. & Donchin, E. Event-related brain potentials and error-related processing: an analysis of incorrect responses to go and no-go stimuli. *Psychophysiology* **33**, 42–53 (1996).

[37] Friston, K. A theory of cortical responses. *Philosophical Transaction of the Royal Society London, B. Biological Sciences* **360**, 815–836 (2005).

[38] Holroyd, C. B. & Coles, M. G. H. The neural basis of human error processing: reinforcement learning, dopamine, and the error-related negativity. *Psychological Review* **109**, 679–709 (2002).

[39] Alexander, W. H. & Brown, J. W. Medial prefrontal cortex as an action-outcome predictor. *Nature Neuroscience* **14**, 1338–1344 (2011).

[40] Silvetti, M., Seurinck, R. & Verguts, T. Value and prediction error in medial frontal cortex: Integrating the single-unit and systems levels of analysis. *Frontiers in Human Neuroscience* **5** (2011).

[41] Brown, J. W. Beyond conflict monitoring: Cognitive control and the neural basis of thinking before you act. *Current Directions in Psychological Science* **22**, 179–185 (2013).

[42] LoTemplio, S., Silcox, J., Murdock, R., Strayer, D. L. & Payne, B. R. To err is human- to understand error-processing is divine: Contributions of working memory and anxiety to error-related brain and pupil responses. *Psychophysiology* **60** (2023).

[43] Boudewyn, M. A., Luck, S. J., Farrens, J. L. & Kappenman, E. S. How many trials does it take to get a significant ERP effect? It depends. *Psychophysiology* **55** (2017).

[44] Jensen, K. M. & MacDonald, J. A. Towards thoughtful planning of ERP studies: How participants, trials, and effect magnitude interact to influence statistical power across seven ERP components. *Psychophysiology* **60** (2022).

[45] Peirce, J. *et al.* PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods* **51**, 195–203 (2019). URL https://doi.org/10.3758/s13428-018-01193-y.

[46] Liu, J. & Liu, Q. 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 (2016).

[47] Gramfort, A. *et al.* MEG and EEG data analysis with MNE-Python. *Frontiers in Neuroscience* **7** (2013). URL https://www.frontiersin.org/articles/10.3389/fnins.2013.00267/full.

[48] Perrin, F., Pernier, J., Bertrand, O. & Echallier, J. F. Spherical splines for scalp potential and current density mapping. *Electroencephalography and Clinical Neurophysiology* **72**, 184–187 (1989).

[49] Vallat, R. Pingouin: statistics in Python. *Journal of Open Source Software* **3**, 1026 (2018). URL https://joss.theoj.org/papers/10.21105/joss.01026.

[50] Seabold, S. & Perktold, J. Statsmodels: Econometric and Statistical Modeling with Python 92–96 (2010). URL https://conference.scipy.org/proceedings/scipy2010/seabold.html.

[51] Bates, D., Mächler, M., Bolker, B. & Walker, S. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software* **67**, 1–48 (2015). URL https://doi.org/10.18637/jss.v067.i01.

[52] Bakdash, J. Z. & Marusich, L. R. Repeated Measures Correlation. *Frontiers in Psychology* **8**, 456 (2017). URL https://www.frontiersin.org/article/10.3389/fpsyg.2017.00456/full.

Metacognitive context tunes medial prefrontal cortex performance monitoring activity. Supplementary Information

Nathalie Pavailler[1], Franck Vidal[1], Boris Burle[1*]

[1]CRPN, Aix-Marseille Univ., CNRS, Marseille, France.

*Corresponding author(s). E-mail(s): boris.burle@univ-amu.fr;

Supplementary Notes

Force-RT relationship

To assess a potential specificity, one needs two continuous and comparable, yet non-correlated, dimensions. Reaction times (RT) and force were chosen as they both are produced internally and accessible at the moment of response. Examination of their respective distributions within each group revealed notable similarities, with both displaying slightly right-skewed distributions (Fig. 1). In the RT-group, repeated measure correlation revealed a small positive correlation between RT and force ($r_{rm}(8314) = .09, p < .001$). Individual Pearson r values ranged from -.21 to .3 with a mean of .06 (and a standard deviation of .13). These low r values coupled with a high variability between participants suggest that the correlation is most likely non-existent. Conversely, in the Force-group, the same analysis revealed a negative correlation between RT and force ($r_{rm}(8671) = -.18, p < .001$). Individual Pearson r values range from -.45 to .13 with a mean of -.16 and a standard deviation of .16. While these values suggest the existence of a negative correlation, it remains notably heterogenous and weak (see Fig. 2).

Although the cause of the discrepancy between groups (or individuals) remains unknown, it most certainly cannot explain the observed EEG results. The absence of correlation between RT and force in the RT-group makes us confident about the observed dissociation between effects of RT and force on Ne amplitude. Concerning the Force-group, the weak negative correlation between RT and force implies that short RT are more often associated with strong forces. In this case, even though there is a small relationship between Ne amplitude and RT, it is nonlinear and weak, it is hence unlikely that it is responsible of the observe stronger linear effect with force. Conversely, it is also unlikely that the

relationship with force would have erased a relationship with RT. Higher Ne amplitude for stronger forces would have instead induced higher amplitude for shorter RTs, but this linear effect is largely reduced in the Force-group.

Effect of difficulty on Ne amplitude

As both RT and force were highly affected by task difficulty/stimulus angle in both groups, we looked at a potential effect of difficulty on Ne amplitude (Fig. 3). Epochs were averaged by difficulty level and Ne amplitude was extracted. No effect of difficulty was observed in the RT-group ($t(19) = 1.59, p = .13$) nor in the Force-group ($t(18) = -1.16, p = .26$).

Compare RT and force between groups

RTs and forces were compared between groups to check that no differences can explain the differences observed in the EEG activity. RTs were slower in the Force-group ($t(32) = -4.28, p < .001$), but there was no difference in variability ($t(36) = -1.78, p = .08$). Forces were also stronger in the Force-group ($t(24) = -2.64, p = .01$), with a higher variability ($t(26) = -5.12, p < .001$). Difference in force could not be due to normalization by maximum force as no difference in maximum voluntary force was observed between groups ($t(29) = .35, p = .73$).

To address potential confounding variables between groups, we examined several factors. First, in our bin analysis (as depicted in Fig. 2 in main text), we observed overlapping mean reaction times (RTs) and force levels between groups. Despite this overlap, the Ne amplitude patterns still exhibited notable distinctions. This indicates that the observed Ne pattern in one group cannot simply be a subset of the pattern observed in the other group. Furthermore, our raster plot analysis, sorted by RTs, also revealed a large overlap of RT values between groups, further suggesting that the observed difference in Ne patterns is not due to group-specific variations in RTs. When trials were sorted by force, and as illustrated in the regression analyses (Fig. 3 in main text), there was indeed a discernible difference in variability between groups. To ensure that the linear relationship with Ne amplitude observed in the Force-group was not merely a consequence of this higher variability, we conducted a secondary analysis. Specifically, we excluded force values exceeding the maximum value observed in the RT-group (45% of maximum force). Even with this more restricted dataset, a significant linear prediction of Ne amplitude by force persisted ($\beta = -.04, t = -6.11, p < .001$). This suggests that the relationship between force and Ne amplitude in the Force-group is robust and not solely driven by a higher variability.

Conditional Accuracy Functions

To assess if the observed Ne pattern could be due to differences in accuracy for different RT and force values, we computed Conditional Accuracy Functions (CAF): for each participant, trials were sorted by RT/force and divided into 5 bins, and mean RT/force and accuracy were computed within each bin and plotted one against the other (Fig. 4). As task difficulty affected RT and Accuracy, this analysis

was performed by level of difficulty for each group. First, mixed ANOVAs with RT bin, difficulty, and group as factors were performed. Both RT bins ($F(2, 75) = 8.8, p < .001$) and difficulty ($F(1, 37) = 177.27, p < .001$) influenced Accuracy, but not group ($F(1, 37) = 2.67, p = .11$). However, group and RT bins interacted ($F(2, 75) = 4.5, p = .014$). Separate ANOVAs for both groups revealed that Accuracy was lower in the difficult condition in both groups (RT-group: $F(1, 19) = 113.89, p < .001$, Force-group: $F(1, 18) = 70.15, p < .001$), and varied as a function of RT bins (RT-group: $F(4, 76) = 8.69, p = .001$, Force-group: $F(4, 72) = 3.64, p = .03$). Factors did not interacted (RT-group: $F(4, 76) = .82, p = .47$, Force-group: $F(4, 72) = 1.70, p = .18$).

Same analysis was performed with force bins. Both Force bins ($F(4, 148) = 36.34, p < .001$) and difficulty ($F(1, 37) = 178.53, p < .001$) influenced Accuracy, but not group ($F(1, 37) = 2.6, p = .12$). However, group and Force bins interacted ($F(4, 148) = 13.05, p < .001$). Force bins and difficulty also interacted ($F(4, 148) = 6.23, p < .001$). Separate ANOVAs for both groups revealed that Accuracy was lower in the difficult condition in both groups (RT-group: $F(1, 19) = 114.4, p < .001$, Force-group: $F(1, 18) = 70.83, p < .001$), and varied as a function of Force bins (RT-group: $F(4, 76) = 42.25, p < .001$, Force-group: $F(4, 72) = 3.55, p = .014$). Factors interacted in the RT-group ($F(4, 76) = 4.93, p = .006$) but not in the Force-group ($F(4, 72) = 2.02, p = .12$).

In both groups, accuracy was hence lower for short RTs and weak forces. This effect was weaker in the Force-group, probably because of overall higher forces and longer RTs. Higher error risk for shorter RTs could explain the higher Ne amplitude for these trials. Additionally, flatter conditional accuracy functions with RT in the Force-group would account for the reduced observed pattern between RT and Ne amplitude. However, it most likely cannot explain the nonlinear relationship with increased Ne amplitude for the longest RTs. Moreover, higher amplitude in conditions of high error risk would predict higher amplitudes for weakest forces. The opposite pattern is observed, and only in the Force-group, which has the weakest effect of force on accuracy. We can thus be confident that the difference in the EEG patterns between groups is not a by-product of these differences in CAFs.

Supplementary Figures

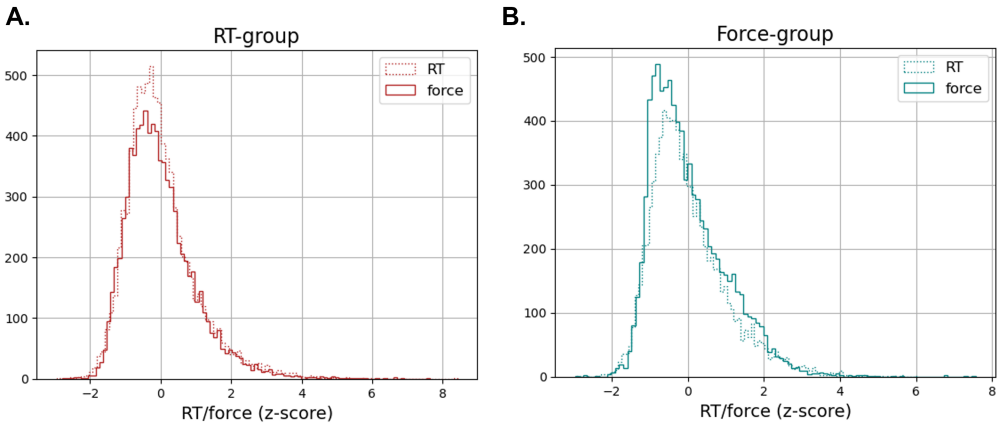


Fig. 1 Distributions of z-score-transformed RT (dashed) and force (solid) of RT-(A) and Force-(B) groups.

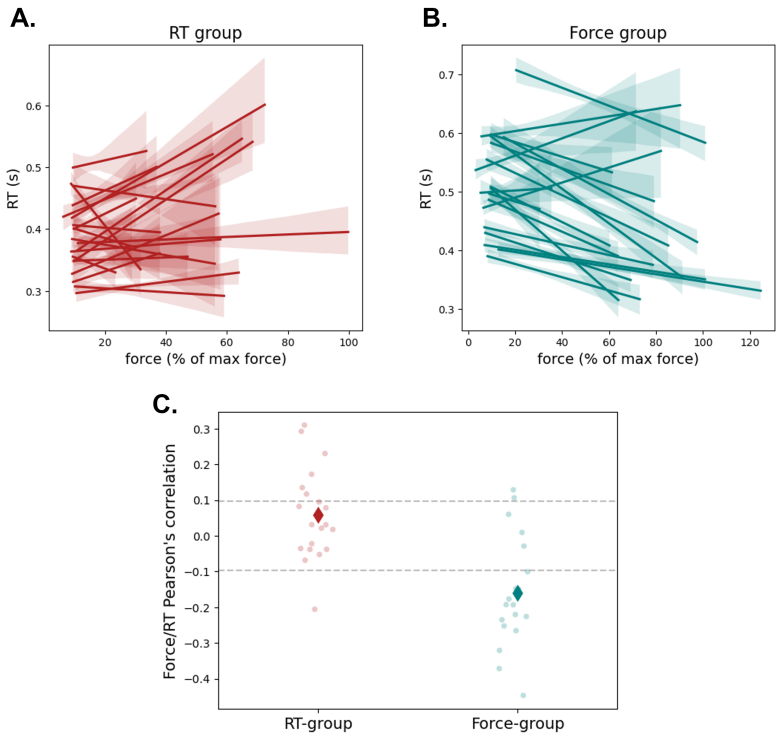


Fig. 2 Individual trial-by-trial correlations between RT and force in RT-group (A) and Force-group (B). Each regression line represents a participant. (C) individual Pearson's r values of force-group (blue) and RT-group (red). Dashed grey horizontal lines represent critical r values for alpha=0.05 and n=400.

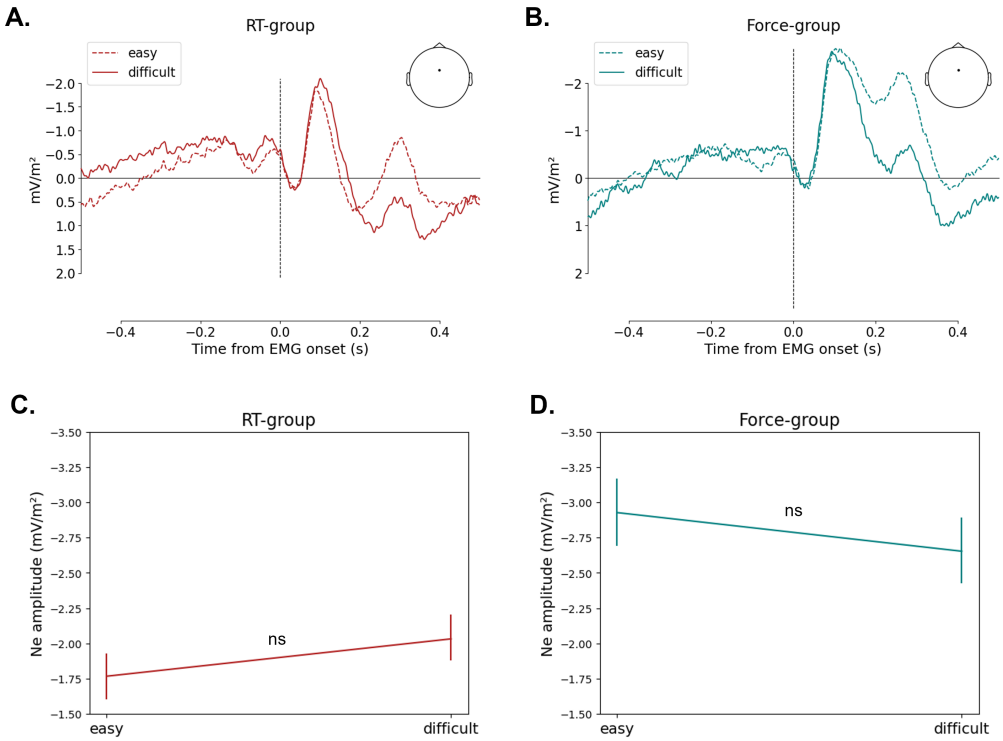


Fig. 3 Grand averages per level of difficulty at electrode FCz for (A) RT-group and (B) Force-group. Peak-to-peak amplitudes for each level of difficulty are presented in bottom panels for (C) RT-group and (D) Force-group.

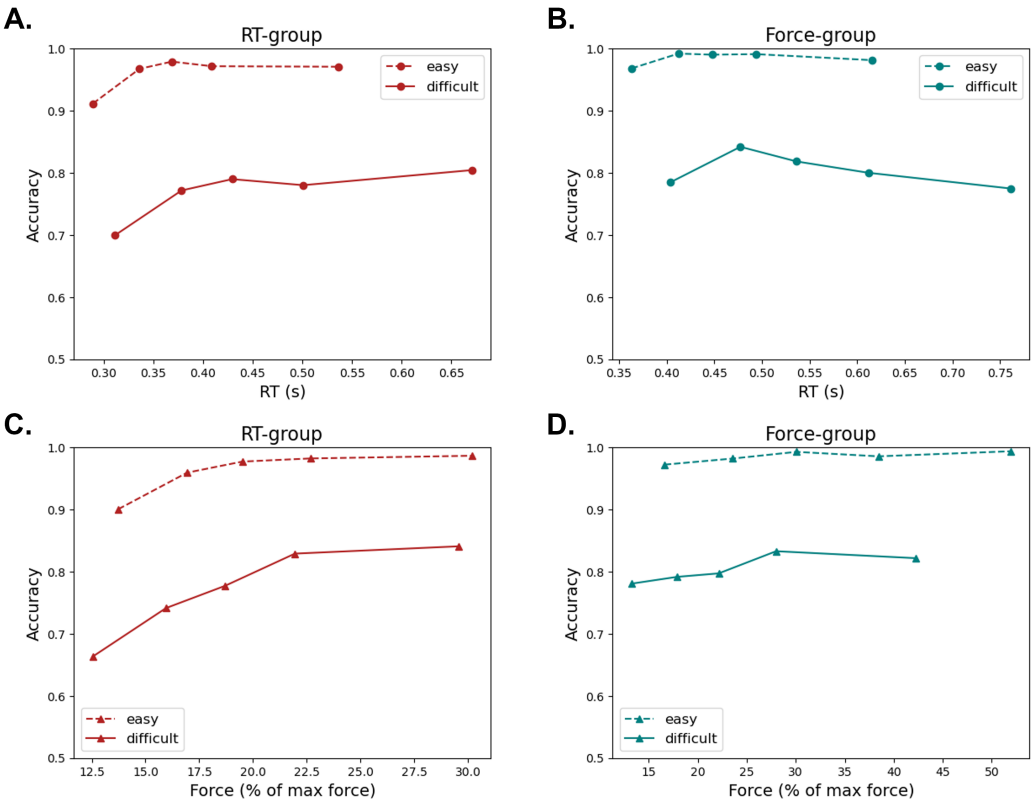


Fig. 4 Conditional Accuracy Functions for easy (dashed line) and difficult (solid line) conditions. Top row presents accuracy as a function of mean RT per quantile for (A) RT-group and (B) Force-group. Bottom row presents accuracy as a function of mean produced force per quantile for (C) RT-group and (D) Force-group.

3.3. From objective to subjective performance monitoring: MPFC activity and beyond

3.3.1. Introduction

One of the central questions of this thesis concerns the potential contribution of MPFC activity, as a proxy of a general performance monitoring system, to the subjective evaluation of RT, iRT. As stated in section 3.1, if it has such a role, it should necessarily be sensitive to the objective RT, and this sensitivity should be specific of this metacognitive context. This first hypothesis has been verified in section 3.2, showing that MPFC activity is modulated by RT when it has to be subjectively estimated. However, while revealing new properties of MPFC performance monitoring activity, this previous section does not give much evidence for a support of the subjective judgment.

The fact that the Ne is increased in case of subjective judgment (Grützmann et al., 2014) and that it is sensitive to RT only in condition of RT judgment (section 3.2), makes it an obvious candidate to support iRT. While it would appear intriguing that a system able to monitor performance would not contribute to an explicit subjective evaluation of this same performance, there is no direct evidence supporting this.

In fact, the nature of the relationship between Ne and iRT might not be as straightforward as it could seem: although the above arguments provide good reasons to expect that Ne contributes to iRT, the results obtained in chapter 2 already suggest that the Ne will not be the sole contributor. Indeed, iRT appears to be the sum of an iPMT and an iMT, and can be biased by different contextual sources of information. But the Ne can very likely not directly access iMT, as it starts only tens of milliseconds after EMG onset (during MT, and before its end); it could however have access to iPMT, the latter being finished as the Ne starts.

As no measures of iPMT and iMT are accessible, we first looked at the relationship between Ne and iRT, and predicted that it would not be strong, given the arguments presented above. To go further and assess its relationship with subparts of the judgment, we investigated its link with PMT and MT, and hypothesized that it would only be related to PMT, and hence, support iPMT.

To explore which activity could support iMT, we searched for activities which could be related to MT. A natural candidate lies in the primary motor cortex (M1) contralateral to the response. Indeed, motor execution of a response is associated with an activation of the contralateral M1. This activation is associated with the strength of the motor command, a larger activation being related to a more 'efficient' motor command (Arezzo & Vaughan, 1980). In EEG (after CSD computation), a negative wave developing just before the EMG onset of the response over M1 contralateral to the responsive hand can be observed (Vidal, Grapperon, et al., 2003, see Fig. 3.4.A). Based on data using stimulation techniques, such as the Hoffman-reflex (Hasbroucq et al., 2000) or TMS (Burle, Bonnet, et al., 2002; Leocani et al., 2000), this negativity can safely be interpreted as reflecting the descending motor command (see Burle et al.,

2004; Vidal, Grapperon, et al., 2003 for discussions).

M1 activation is more intense and prolonged in conditions where participants have a high response force threshold compared to a weak one (Meckler, 2010). Without any experimental manipulation, fluctuations in motor execution might as well be related to fluctuations in the strength of the motor command. Increased M1 activation should hence lead to a more efficient corticospinal activity and shorter MT. In addition, M1 activity has previously been hypothesized to contribute to partial errors detection, and could be used by the metacognitive system (Ficarella et al., 2019). We thus also looked at M1 activation as a function of PMT and MT, with the hypothesis that it would specifically correlate with MT.

Finally, as most metacognitive judgments, iRT is given retrospectively, several seconds after the end of RT (the response). Consequently, more information can be integrated before the metacognitive evaluation is provided (Fleming & Daw, 2017; Yeung & Summerfield, 2012), likely making this evaluation diverge from the original performance. One can expected, however, that activities close to the response would mainly follow first-order measures and later activities would be more closely related to the subjective representation measured at the time of the retrospective judgment. The Ne appears to be followed by a slow positive wave[1]. It develops in the same latency range as the Pe, which has been associated to an accumulation process leading to error awareness (Steinhauser & Yeung, 2010). Following this line, it is possible that when having to evaluate RT, information related to time also accumulate after the response, modulating the initial representation, and leading to the finally produced iRT. In a more exploratory way, we lastly studied this activity in relation to RT and iRT.

3.3.2. Methods

All the data reported in this chapter are from the RT-group of the experiment described in section 3.2. Similarly, only correct trials with only one EMG burst were analysed. All RT were fractioned as in chapter 2: premotor time (PMT) was defined as the time between stimulus onset and EMG onset, and motor time (MT) was defined as the time between the EMG onset and the mechanical response (time of the force threshold crossing). To look at different EEG activities according to either PMT, MT, RT or iRT duration, we followed the same methodology as in section 3.2. Briefly, EEG data was segmented in 1s-epochs around EMG onset. Trials were binned in five categories as a function of either variables. For this purpose, PMT, MT, RT and iRT distributions were splitted into five bins of equal sample sizes, from short (bin 0) to long (bin 4). Epochs were then averaged within each bin.

Ne amplitude was extracted in the same way as in section 3.2. Peak latency was also measured by fitting a polynomial function on EEG data in a window of 50 points around peak index determined individually. Time of the minimum of this fitted function was used as peak latency.

[1] During this positivity, another negative component develops (N300, Bartholow et al., 2005). The nature and status of this activity is still unclear.

To extract M1 activation, left and right responses activity were combined by switching activities of left and right electrodes for left responses. As for the Ne analysis, epochs were averaged time-locked to the EMG onsets leading to the responses, with a window from 0.5s before to 0.5s after EMG onset. M1 activation was extracted at C3 electrode. Onset and offset were determined on averages across all trials for each participant. Then, epochs were averaged within each MT, or PMT bin, and we measured both amplitude and slope of M1 activation (Fig. 3.4.A). Amplitude was calculated by taking the difference between values at offset and onset (peak-to-peak amplitude). To measure slope, we fitted a linear regression to data between onset and offset.

Finally, to study the activity following the Ne, we measured the slope in a window from 400 to 500ms after the EMG onset by fitting a linear regression on the corresponding data. This analysis was done at electrode Cz, at which the ramping activity was clearer.

Analyses were performed with one-way ANOVAs using the Pingouin python package (Vallat, 2018) and mixed models using the statsmodel python package (Seabold & Perktold, 2010) and the lme4 R-package (Bates et al., 2015).

3.3.3. Results

3.3.3.1. MPFC activity as a function of iRT

To assess whether the MPFC activity could support the subjective evaluation of RT, we first looked at the relationship between Ne amplitude and iRT. One-way ANOVA did not reveal a significant effect of iRT bins on Ne amplitude ($F(4,76) = 0.56, p = 0.69$, Fig.3.1.A and .B). In a second step, following previous chapter's methodology, we also did a raster plot analysis by averaging all individual raster plots sorted by iRT, as well as individual sorted iRT. Ne amplitude extracted for each of the obtained grand averaged 'pseudo' trials was regressed with grand averaged iRT and iRT squared in a mixed model (for detailed method see section 3.2). iRT tended to weakly predict Ne amplitude ($\beta = 7.59, t = 1.76, p = 0.08$), but iRT squared did not ($\beta = -6.71, t = -1.58, p = 0.11$, Fig.3.1.C and .D).

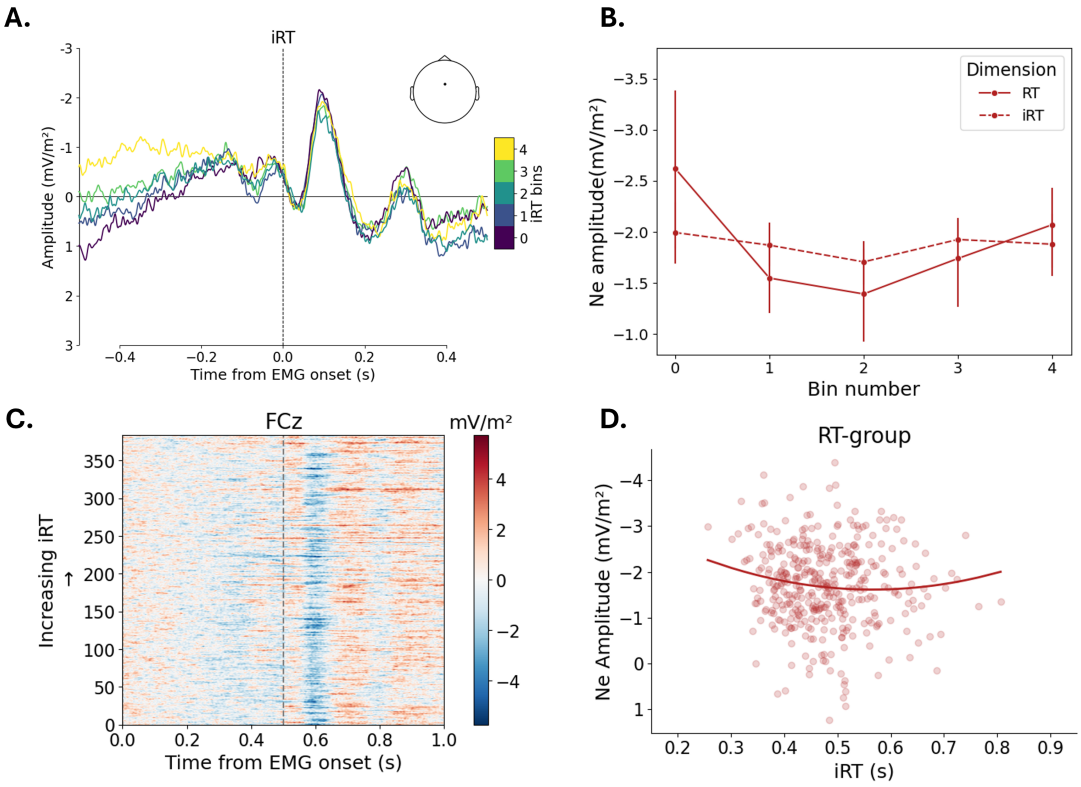


Figure 3.1. *(A) Grand averages per iRT bins at electrode FCz. (B) Ne amplitude as a function of RT (solid line) and iRT (dashed line) bins. (C) Grand averaged ERP image sorted by iRT. Epoch 0 corresponds to smallest iRT. (D) Ne amplitude of each grand averaged 'pseudo trial' is plotted as a function of corresponding iRT with fitted quadratic function.*

3.3.3.2. MPFC activity as a function of PMT and MT

Since no relationship with iRT was obtained, we sought further by looking at the link between Ne and the RT subcomponents.

One-way ANOVAs revealed an effect of PMT bins on Ne amplitude ($F(4,76) = 3.58$, $p = 0.038$, Fig.3.2.A). As for the relation with RT, the relation with PMT follows a U-shape, with higher amplitude for shortest and longest PMT. This relation with PMT appeared to be sharper than with RT. A two-way ANOVA with dimension (RT or PMT) and bin number as factors showed an interaction, although marginally significant, between both ($F(4,16) = 2.26$, $p = 0.07$, Fig.3.2.C). Bin number affected Ne amplitude ($F(4,76) = 3.24$, $p = 0.05$), but there was no difference between dimension ($F(4,76) = 0.01$, $p = 0.92$). Mixed model analysis with mean PMT and PMT-squared as predictors confirmed these observations (PMT: $\beta = 27.48$, $t = 2.87$, $p = 0.004$; PMT-squared: $\beta = -40.96$, $t = -3.02$, $p = 0.003$, Fig.3.2.D).

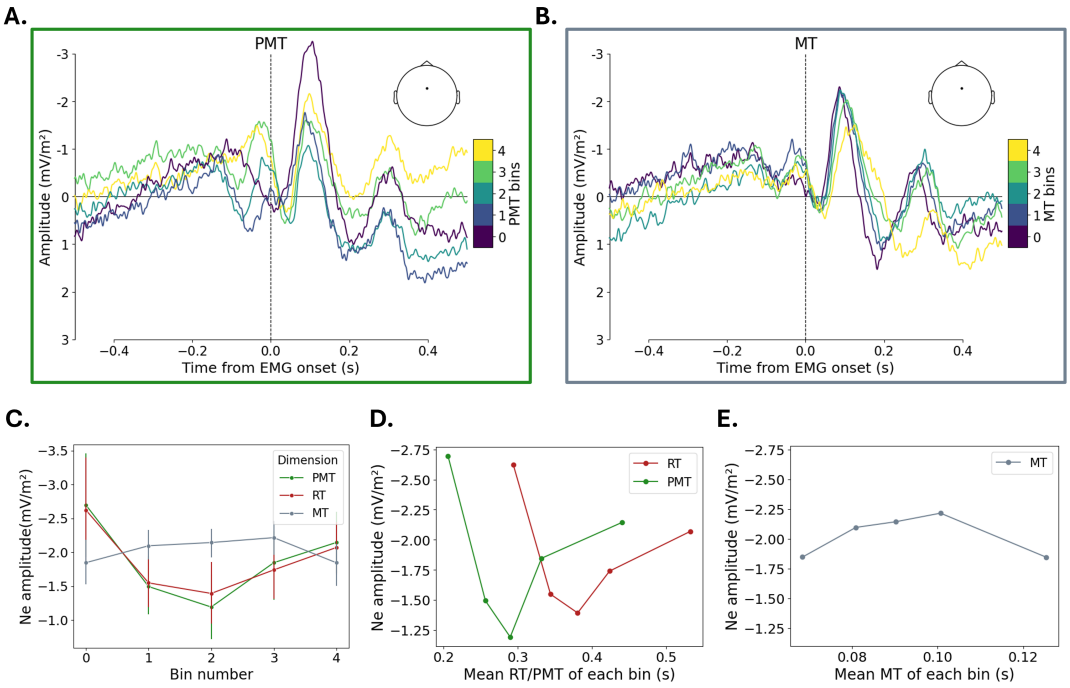


Figure 3.2. *(A) Grand averages per PMT bins at electrode FCz. (B) Grand averages per MT bins at electrode FCz. (C) Ne amplitude as a function of RT (red), PMT (green) and MT (grey) bins. (D) Ne amplitude plotted as a function of mean RT (red) and mean PMT (green) of each bin. (E) Ne amplitude plotted as a function of mean MT of each bin.*

Assessing a relationship between Ne and MT is problematic, as inspection of Fig.3.2.B suggests that Ne peak latency increases with MT. This was confirmed by an ANOVA showing that the Ne peaked later with increasing MT ($F(4,76) = 8.04$, $p < 0.001$, Fig.3.3.B). While this difference in latency is interesting in itself, we also had to take it into account to measure the amplitude as a function of MT. To do so, we used these peak latencies to calculate the peak-to-peak amplitude for each MT bin. No effect of MT bins on Ne amplitude was observed ($F(4,76) = 1.29$, $p = 0.28$, Fig.3.2.B) and mean MT did not predict Ne amplitude in the mixed model analysis ($\beta = 0.57$, $t = 0.17$, $p = 0.87$, Fig.3.2.E).

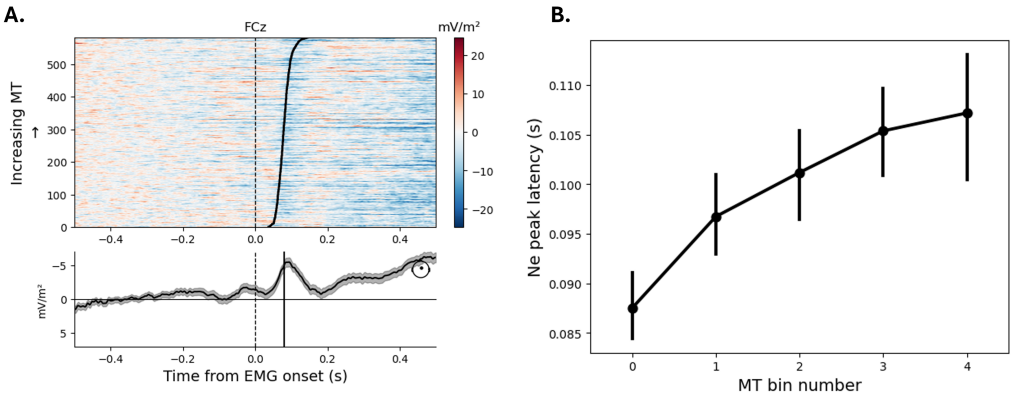


Figure 3.3. *(A) Example of an individual ERP-image sorted by MT (top). Solid black line represents the response (end of MT). ERP obtained by averaging all trials is plotted (bottom). Vertical black line is plotted at averaged MT. (B) Ne peak latency as a function of MT bins.*

3.3.3.3. Primary motor cortex activation

Amplitude of M1 activation was higher for shorter MTs ($F(4,76) = 2.96$, $p = 0.025$, Fig.3.4.B). Slope of the activity also tended to be steeper for shorter MTs ($F(4,76) = 2.34$, $p = 0.06$). Mixed models with mean MTs per bin as predictor showed similar results. MT predicted both the amplitude of M1 activation ($\beta = 0.006$, $t = 3.22$, $p = 0.001$), and its slope ($\beta = 0.0002$, $t = 2.63$, $p = 0.009$).

Same analyses were made with PMT bins (Fig.3.4.C). ANOVAs did not show any effect of PMT bins on amplitude of M1 activation ($F(4,76) = 0.99$, $p = 0.42$, Fig.3.4.D), nor on the slope ($F(4,76) = 1.35$, $p = 0.26$). Mixed models with mean PMT per bin as predictor showed comparable results for amplitude ($\beta < 0.001$, $t = -0.34$, $p = 0.73$) and slope ($\beta < 0.001$, $t = -1.08$, $p = 0.28$).

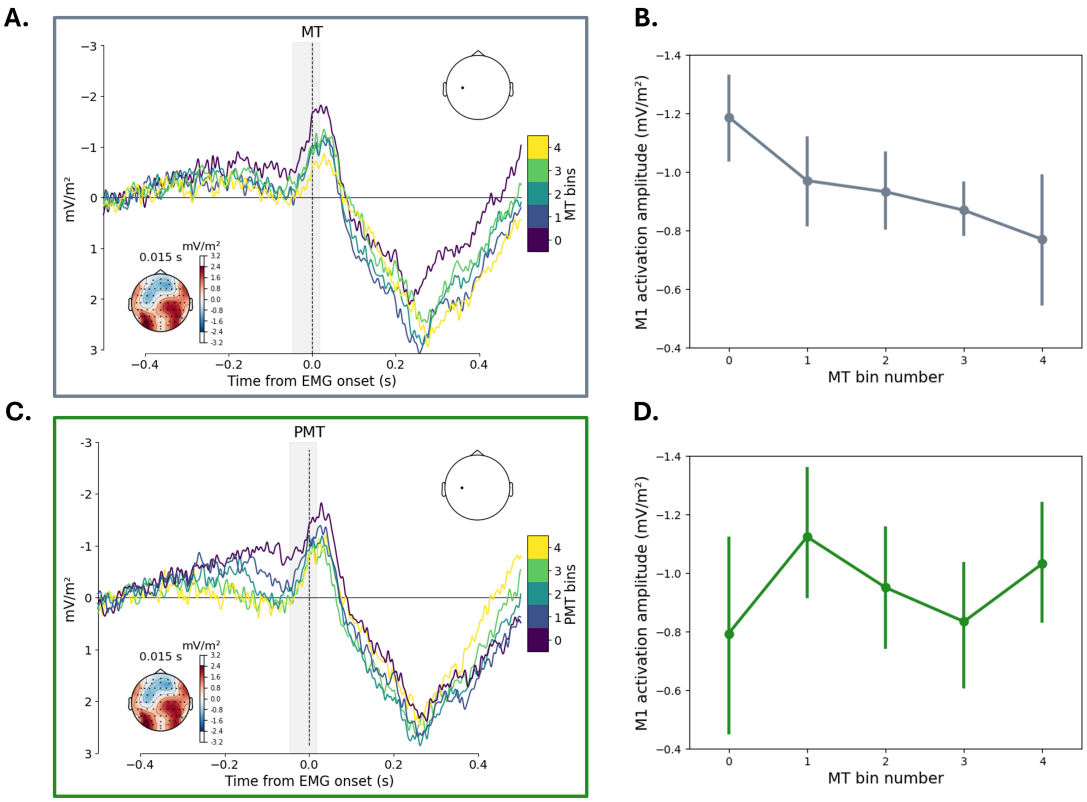


Figure 3.4. *(A) Grand averages per MT bins at electrode C3. Grey zone represents the mean window of analysis. Topography at mean peak latency of M1 activation is plotted at the bottom left. (B) M1 activation amplitude as a function of MT bins. (C) Grand averages per PMT bins at electrode C3. Grey zone represents the mean window of analysis. Topography at mean peak latency of M1 activation is plotted at the bottom left. (D) M1 activation amplitude as a function of PMT bins.*

3.3.3.4. Activity following the response

We finally looked at the slope of the slow positive activity following Ne, as it seemed highly related to RT (see Fig.3.5.A). One-way ANOVAs revealed an effect of RT bins on this slope ($F(4,76) = 2.5, p = 0.049$, Fig.3.5.B), with steeper slope for longer RT, but failed to bring out the same effect with iRT bins ($F(4,76) = 1.68, p = 0.16$). Mixed models using the slope of this activity as dependant variable and either mean RT or iRT per bin revealed a linear relationship with both RT ($\beta = -0.018, t = -2.08, p = 0.037$) and iRT ($\beta = -0.014, t = -2.03, p = 0.043$, Fig.3.5.C).

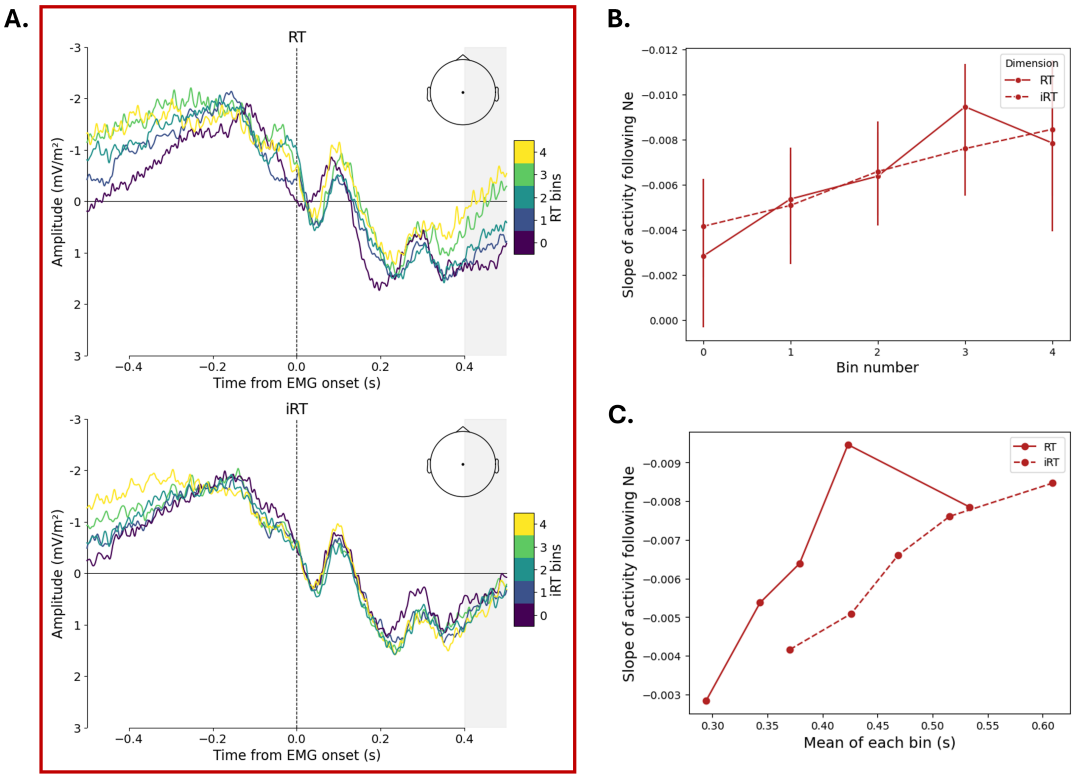


Figure 3.5. *(A) Grand averages per RT bins (top) and iRT bins (bottom) at electrode Cz. Grey zone represents the window of analysis. (B) Slope of the late activity as a function of RT bins (solid line) and iRT bins (dashed line). (C) Slope of the late activity plotted as a function of mean RT (solid) and mean iRT (dashed) of each bin.*

3.3.4. Discussion

3.3.4.1. Contribution of MPFC activity to iRT

Based on the arguments presented in the general introduction and in the previous sub-chapter, we could predict a contribution of MPFC activity (through the Ne) to the subjective judgment of RT.

For an activity to support a metacognitive judgment, it both needs to be related to the first-order variable that must be evaluated, but also to the second-order judgment. Section 3.2 highlighted that MPFC activity is indeed related to RT, confirming the first assessment. We also showed that this is the case only when participants are asked to make a metacognitive judgment about RT. This could intuitively mean that in such a condition, the RT tracking by MPFC is directly related to the subsequent judgment, implying a pure metacognitive effect. A few previous studies reported a relationship between RT and MPFC activity outside a context of metacognitive judgment. However, they used either a monetary punishment for slow RT (Valt & Stürmer, 2021) or a reward for fast RT (Files et al., 2021). We could argue that in these situations, to maximize reward, one needs to know when one is being fast or slow, and thus still requires an

implicit monitoring, even without having to make an explicit judgment.

However, the present results failed to demonstrate a relationship between MPFC activity and iRT. While this could be taken as evidence that MPFC activity does not contribute to the subsequent judgment, this is not the only possible explanation. Chapter 2 demonstrated that iRT represents the sum of multiple information. First, it is most likely the sum of an iPMT and iMT, which has been reproduced with the present data (see Appendix A.3). Second, iRT also contains several biases induced by task context. Even if MPFC is involved in subjective evaluation, this does not mean that it tracks all this different information. In this case, the relation between its activity and the subjective measure is necessarily weaker. Indeed, when looking at the data of the Force-group, a linear relationship with iForce was still observed, although with a less steep slope, i.e. a weaker sensitivity (see Appendix B.1).

The present results showed that Ne amplitude varied as a function of PMT, but not as a function of MT. Ne amplitude appears to code PMT, and consequently could constitute a basis for iPMT. However, we saw that this relation was quadratic, with higher amplitudes for both shorter and longer PMT. We argued in section 3.2 that MPFC computes a prediction error. The present data suggests that it does so on PMT rather than RT. One could wonder whether this implies that Ne amplitude codes for iPMT. In such case, one should however expect a linear relationship. Given the quadratic relationship, a large Ne can code either for a short or a long RT, and hence cannot provide an unambiguous information about RT length. In chapter 6, we will propose a potential mechanism reconciling the present data with an implication of Ne in iPMT.

While Ne amplitude did not vary as a function of MT, its latency did, and could be useable for iMT. This would mean that two ways of coding (amplitude and latency, Gerstner et al., 1997) would be present in the same signal. It is possible that such latency effect would be mechanical. Previous studies have shown that the duration of the Ne was related to the duration of partial EMG activation (Burle et al., 2008). It is possible that the Ne is stopped once the mechanical response is given, and hence peaks at different times for differents MTs. In any case, this information could be used to estimate MT.

Finally, in light of its timing, MPFC activity most certainly does not include biases induced by task context, at least not all of them. While it would be interesting to further investigate the impacts of these biases, we can only speculate (see General Discussion), as the present study was not designed to address this question.

Overall, MPFC activity appears to be related, at least to some extent, to both PMT and MT, and thus most probably contributes to iRT. However, these observations, despite the fact that they give essential evidence, are not a demonstration. To really back up an hypothesis about a causal relationship between MPFC activity and iRT, interventional research would be needed. Being able to inhibit this activity at the time of the EMG onset would allow to investigate the effects on iRT. If the causal relationship exists, disrupting MPFC activity should also disrupt the RT/iRT relationship (see section 5.1). In addition, we highlight that MPFC activity cannot lead to iRT by itself, and this metacognitive representation of RT most probably emerges from a combination of

different cerebral activities.

3.3.4.2. iRT likely arises from multiple cerebral activities

Just as RT is the sum of the duration of different processes, iRT is likely the sum of the estimations of these subdurations. If one attempts to search for activities supporting iRT, one first has to look for activities correlating with PMT and MT. Ne amplitude was shown to correlate with PMT, and its latency with MT. However, its relationship with PMT presented a U-shape, which we hypothesized in section 3.2 could represent the deviation from an expected value. On one hand, if MPFC computes this prediction error, it needs to receive the actual PMT value as an input. On another hand, if this activity reflects by how much the actual PMT deviates from the expected value, it cannot inform about whether it is shorter or longer. In both cases, it means that another activity must be tracking PMT duration. Such activity should likely be a ramping activity developping during the interval of interest, that is RT. While this was not investigated in the present thesis, potential candidates will be discussed in the General Discussion.

Concerning MT, we found that shorter MT were linked to stronger M1 activation. It is not surprising that if M1 activation reflects the strength of the motor command, a stronger one would lead to shorter MT. Hence, reading the strength of the motor command would directly inform about MT duration. Indeed, during partial errors, M1 activity of the incorrect response is larger for detected partial errors than for undetected ones, and even larger for overt errors (Ficarella et al., 2019), suggesting that it could contribute to partial errors detection.

If these different activities support iPMT and iMT, then iRT should be better predicted by considering all of them. Applying the logic used in chapter 2 according to which iRT reflects a sum of iPMT and iMT, it would be explained by the sum of cerebral activities supporting PMT and those supporting MT.

Bringing out activities correlated with PMT and MT does not imply that they are involved in metacognition, and this question cannot be addressed without measures of iPMT and iMT. Considering that they arise early, close to the response, the previously described activities are more likely related to first order than second order variables. As they are closely linked to first order variables, i.e. the ones to estimate, they could serve as sources of information integrated later in time to give rise to an iRT representation. Indeed, many metacognitive models stipulates that more information can be integrated after the first-order response and participate in the second-order judgment (post decisional locus models, see Yeung and Summerfield, 2012). Following this idea, we looked at a later activity, around 300 ms after the response, whose slope increases linearly with both RT and iRT. This could be compatible with an integration and accumulation of information giving rise to the subjective judgment, paralleling what is hypothesized for the Pe (J. Wessel et al., 2011). However, the slope of the relationship remains steeper with RT than with iRT. As RT and iRT are strongly correlated, it is possible that any activity correlated with RT would also be with iRT. Another possibility is that this later activity could have integrated temporal information about PMT and

MT, but does not include the potential biases observed in chapter 2. Indeed, we do not know when and how these biases come into play. If a priori knowledge could affect early activities, biases due to metacognitive experiences should necessarily affect later processing.

Based on all the results presented so far, it appear clear that, contrary to our initial (simple) working hypothesis, it is very unlikely that a single brain activity could predict iRT. Instead, our results suggests that it is the weighted sum of different activities (or different aspects of one activity) that gives rise to the subjective representation of RT, of which we take iRT as a proxy. On chapter 2, we argued that iRT = iPMT + iMT + biases. We have identified potential candidates for iPMT (Ne amplitude, although not alone) and for iMT (M1 activity and Ne latency), paving the way to probe more directly those hypothesis. We did not (yet) identified the activities that could implement the biases. However, the current experiment was not design to do so, and rather different designs would be necessary to reach this goal.

3.3.4.3. A need to dissociate first and second order measures

The different activities described above are related to RT, or parts of RT, and are consequently good candidates to support iRT. But simply looking at correlations between variables and cerebral activities is not enough. Indeed, cerebral correlates of metacognition can vary in nature. For instance, neural activities tied to first-order processes serve as prerequisites for metacognition, ensuring accurate representations, yet they do not necessarily indicate a metacognitive process itself (de Graaf et al., 2012). Conversely, other neural activities may correlate with subjective judgments as a consequence of those judgments, such as the emotional response triggered by perceiving oneself as slow to respond. Investigating the temporal dynamics of these activities could provide insights into their actual roles: an actual cerebral substrate of iRT should manifest closer to the moment when this metacognitive representation emerges. However, this remains challenging due to the difficulty in pinpointing when a metacognitive representation arises.

As an example, we can notice a ramping activity on the grand average of Fig. 3.1.A, between 500 and 250ms before EMG onset, with a slope clearly sorting on iRT. While at first sight, this can look like a perfect correlate for iRT, it has to be interpreted with caution. Indeed, this activity appears early, and happens even before the appearance of the stimulus, as shortest RTs are around 300ms. It could rather reflect some preparatory activity, leading to short or long RTs. In addition, it seems impossible that such an early activity can index MT.

It highlights an important problematic concerning potential cerebral basis of iRT. We set that such an activity should track RT, and potentially include biases revealed in chapter 2. First, it is not guaranteed that RT representation arises from a unique activity. Second, and most importantly, RT and iRT are highly correlated, it is hence likely that an activity correlating with RT would also do with iRT, and vice versa. However, we can hypothesize that activities linked to metacognition would only be observed in case of subjective judgment, as it was observed with MPFC activity in

section 3.2. To further explore activities specific to the subjective judgment, one needs to find dissociation with the objective measures (Fleming & Dolan, 2012), and assess causal implications. Indeed, these activities should differ depending on the 'quality' of the judgment. One way of dissociating first and second order variables would be to look at activities related to different iRTs associated to the same RT. However, too few trials would remain in each category.

One prediction is that sensitivity to RT should be higher in condition of high metacognitive performance. At the trial level, it is challenging to dissociate well-estimated from poorly-estimated RT. Another option is to separate participants based on their metacognitive abilities, indexed by the individual RT/iRT correlations (see Appendix B.2). One prediction is that participants with the highest correlations should present activities with the highest sensitivities to RT.

Overall, we showed that several activities correlated with RT, some of them specifically when it had to be estimated. While our findings and explorations offer essential insights, they also underscore the complexity of the neural mechanisms underlying metacognition, highlighting the need for further investigation to clarify the roles of various cerebral activities in supporting iRT.

4. Second order judgments influence first order behavior

Summary

4.1. Note on the present chapter 122
4.2. Introduction 122
4.3. Methods 126
4.3.1. Comparison of first order measures 126
4.3.2. Mixed models for sequential effects 126
4.4. Results 127
4.4.1. Effect of global metacognitive context on first-order behavior . . 127
4.4.2. Effect of local metacognitive judgments on subsequent behavior 129
4.5. Discussion 131

4.1. Note on the present chapter

The two first chapters of this thesis investigated metacognitive monitoring of RT, highlighting which processes can be monitored (chapter 2) and which cerebral functions could underlie this ability (chapter 3). The function of metacognitive monitoring has usually been interpreted in terms of control and regulation of behaviour and mental processes (Nelson, 1990). This functional perspective led to the question regarding when and how humans use their knowledge about their own cognition to guide their behaviour. These questions have been largely addressed in the fields of memory and decision-making (see general introduction, sections 1.1.3 and 1.1.4). The same question can be asked about the functional role of estimating RT or produced force, as participants presented a high ability in doing so. In the present chapter, we used behavioral data of chapters 2 and 3 to explore whether and how subjective RT (and force) estimations could be used to regulate first-order behavior.

4.2. Introduction

Most of the time, the study of metacognitive abilities in humans relies on a second-order instruction requiring a subjective judgment. In this framework, researchers classically dissociate first-order cognitive processing with second-order metacognitive processing, both being seen as globally independant. More or less implicitly, it is

considered that asking for a metacognitive judgment does not affect the way the first-order task is performed. This is, however, not guaranteed. In a study in which participants had to judge their accuracy, Grützmann et al. (2014) showed that their RT were slower in the rating condition, implying that they probably shifted their response strategy to a more cautious type. This observation parallels effects of experimental emphasis on accuracy (Wickelgren, 1977). In a similar way, in the case of iRT or iForce judgments, the mere fact that participants have to evaluate one particular aspect of their performance could impact the way they performed the task. We could hypothesize that RT judgment would have an opposite effect to accuracy judgement, with a global decrease in response time, paralleling an effect of speed emphasis (Wickelgren, 1977). Indeed, effectively following the instruction to be as fast as possible while avoiding overly risky, hasty responses requires the ability to accurately detect and regulate RT (P. Rabbitt, 1979). Being required to explicitly judge these durations might enhance this control, and thus globally speed-up RTs.

As a matter of fact, according to one of the first model of metacognition (Nelson, 1990), object and meta-levels are related by both top-down and bottom-up processes: metacognition encompasses both the ability to be aware of one's cognitive processes (metacognitive monitoring) and to regulate them (metacognitive control). Metacognitive control allows individuals to adapt their cognitive processes in response to their own monitoring and evaluation of those processes. This means that not only metacognitive judgments depend on first-order decision-making process, but the other way around too: decision-making depends on metacognitive judgments. There are evidence that confidence indeed conditions perceptual accumulation of evidence (Balsdon et al., 2020). This bidirectional influence indicates that metacognitive judgments can shape how individuals process information, challenging the notion of a strict separation between metacognitive and first-order processes.

This metacognitive control effect can manifest in two distinct ways. First, as previously mentioned, it can lead to a strategic adaptation in how the first-order task is globally processed, influenced by the mere fact of being required to make a judgment. Second, metacognitive judgments can induce trial-by-trial adaptations, where the way one assesses one's performance in one trial affects how one approaches the subsequent trial. Such sequential effects are more likely driven by the specific judgment made.

Indeed, metacognitive control has mostly been addressed by looking at how subjective judgments influence subsequent behaviors. For examples, low confidence in the ability to remember an information will more likely lead to setting up a reminder (Boldt & Gilbert, 2019), and lack of confidence for decision-making will also trigger the search for additional information (Desender et al., 2018). Sequential, adaptation effects represent a large part of research on cognitive control in general. In choice RT task, one of the most studied adaptation effect is post error slowing (PES, P. M. Rabbitt, 1966). PES is traditionally attributed to adaptive adjustments, favoring a slower, more cautious response to improve performance in the subsequent trial (Dutilh et al., 2012; Purcell & Kiani, 2016). Some authors observed a PES only after errors consciously detected by the participant (Endrass et al., 2007; Nieuwenhuis et al., 2001; J. Wessel

et al., 2011; see however Harsay et al., 2018; Hester et al., 2005), suggesting a metacognitive control mechanism. In a DDM framework, decision bounds have been shown to depend on the reported confidence in previous trial (Desender et al., 2019), suggesting that participants use confidence to continuously adjust the speed-accuracy tradeoff of their decisions.

Other effects interpreted in terms of behavioral adaptations have also been associated with metacognitive control. The Gratton effect is observed in conflict tasks and corresponds to a smaller congruency effect following an incongruent trial (Gratton et al., 1992). It is often interpreted as an increase in cognitive control after detection of conflict (M. M. Botvinick et al., 2001), but it seems that bottom-up factors also contribute to the effect (Duthoo et al., 2014; Mayr et al., 2003). In such tasks, participants are able to report different metacognitive experiences such as urge-to-err (Morsella et al., 2009), and experience of congruency (Desender et al., 2014), control (Chambon & Haggard, 2012), and difficulty (Desender et al., 2016). These experiences can be observed in subliminal priming tasks, i.e without conscious access to the stimulus. It has been shown that the Gratton effect could be observed in situation with no visual conflict, but when participants did experience one (Desender et al., 2014; Questienne, Van Opstal, et al., 2018), suggesting a role of the subjective experience in triggering cognitive control and adaptation.

Beyond explicit metacognitive effects, sequences of events significantly impact performance. In two choices RT tasks, stimulus repetitions are notably faster than alternation (M. C. Smith, 1968), at least at short response-stimulus intervals (RSI) (Gao et al., 2009). These sequential effects have been linked to stimulus expectations (Kirby, 1976). Overall, it means that the RT on any given trial depends on the history the previous trial(s). As the instruction is to respond as fast as possible, there is a high probability that participants will speed-up after a slow trial. Note that such an explanation implies that participants have a fair representation of their RT duration. This ability was highlighted in previous studies (Corallo et al., 2008) and in previous chapters of this thesis. This would also suggest that this decrease in RT would happen after perceived slow iRT rather than after objectively slow RT.

In the present chapter, we used data presented in chapters 2 and 3, in which different groups of participants performed an orientation detection task, with either introspective reports of RT (iRT) or force (iForce) after each trial. We first explored if, and how, metacognitive judments globally and strategically modified performance, by comparing first-order performance of both groups. Then, we investigated whether participants would modify their RT as a function of their previous metacognitive judgment.

An operationalization of this second question would consist in knowing if previous iRT would predict present RT, beyond a prediction only based on previous RT. This ties in with Granger's notion of causality (Granger, 1969), where a time serie x_1 is said to cause, in the sense of Granger, another time serie x_2 when the past of x_1 can predict x_2, beyond information contained in the past of x_2 alone (see Fig. 4.1). Formally, testing granger causality consists in comparing an autoregressive model of x_2, with an autoregressive model of x_2 that includes lagged values of x_1.

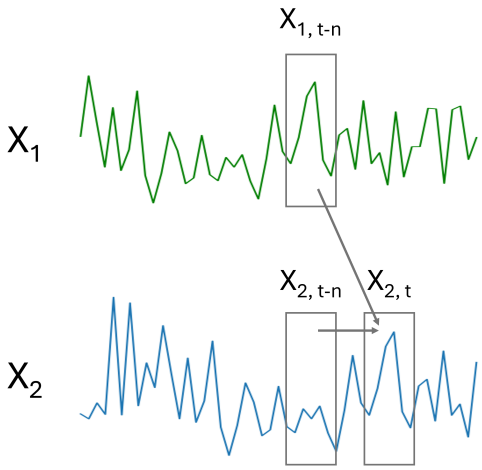


Figure 4.1. *Visual representation of Granger Causality. A time serie x_1 is said to have a causal influence on another time serie x_2, if the past of x_1 $(x_{1,t-n})$ improves the prediction of x_2 $(x_{2,t})$ by its own past $(x_{2,t-n})$*

Classically, RTs show a positive autocorrelation in 2AFC tasks (Laming, 1979a), revealing interesting properties of RT fluctuations which are not due to the experimental conditions. This means that RTs are not independant across trials, and that the cognitive state or performance from one trial influences subsequent ones. Given that iRT is positively correlated with RT, without accounting for RT autocorrelation, past iRT would very likely be positively correlated with present RT. Following the idea of Granger causality, to investigate whether iRT has a causal influence on RT, one needs to test if the addition of previous iRT on the RT autoregressive model improves the prediction of RT. While Granger has no hypothesis on the lag of the influence of x_1 on x_2, in our case, we could suppose that lag 1 (i.e. the effect of the first preceding trial) would be sufficient, especially considering high RSI due the the presence of the metacognitive judgments in-between trials. In addition, Granger assumes that both time series are homogenous. However, we also manipulated perceptual difficulty, in a mixed design, affecting both RT and iRT. Hence, both RT and iRT times series are mixtures of easy and difficult conditions. Difficulty of the previous stimulus can also lead to sequential effects ('mixing costs', see Los, 1996). We hence had to account for difficulty while testing for sequential effects with iRT. Considering this, while we followed the theory of Granger, we implemented it with another method, namely mixed models (see Methods).

4.3. Methods

4.3.1. Comparison of first order measures

Data from the experiments of chapter 3 were used to investigate whether metacognitive instructions could globally influence first-order performance. More precisely, we compared first order measures (RT and force) between groups who received different metacognitive instructions (RT-group and Force-group). RT and force means and intra-individual variabilities were compared between groups with unpaired ttests, using pingouin Python-package (Vallat, 2018). The sample sizes were unequal between groups, so the Welch–Satterthwaite equation was used to approximate the adjusted degrees of freedom. To compare variabilities independently of means, we computed the coefficients of variation (CV), i.e divided standard deviations by means, for each participant and both RT and force, and compared them between groups. Correlations between RT and force were also investigated.

4.3.2. Mixed models for sequential effects

To avoid sequential effects related to error trials, we analyzed sequences of correct trials following correct trials. To see whether and how iRT of the preceding trial influenced RT of the current trial, we performed a mixed model analysis with iRT_{n-1} as a predictor of RT. As RT and iRT are highly correlated, we had to consider RT autocorrelation and hence added RT_{n-1} as a predictor. Finally, we also entered $difficulty_{n-1}$ and *difficulty* to explain part of variance related to difference in difficulties between trials. We used random intercepts for participants.

The final model was the following:

$$RT \sim RT_{n-1} + iRT_{n-1} + difficulty + difficulty_{n-1} \tag{4.1}$$

with RT_{n-1}, iRT_{n-1} and $difficulty_{n-1}$ representing the RT, iRT and *difficulty* level of the preceding trial, respectively.

We applied this model to the data from the RT-group of the EEG experiment described in chapter 3 and from Experiment 2 described in chapter 2. However, we chose not to use the data from Experiment 1 described in chapter 2 (the online experiment) because participants only completed 100 trials, resulting in too few instances of correct responses following correct responses. Experiment 2 of chapter 2 also used a manipulation of response force, with a weak and a strong level. As this manipulation was blocked, we ran two different mixed models for each force level, using the same predictors as above.

We used the same logic to explore sequential effects related to estimation of produced force, using data from the Force-group of chapter 3's experiment. To this aim, we ran the following model:

$$force \sim force_{n-1} + iForce_{n-1} + difficulty + difficulty_{n-1} \tag{4.2}$$

with $force_{n-1}$, $iForce_{n-1}$ and $difficulty_{n-1}$ representing the $force$, $iForce$ and *difficulty* level of the preceding trial, respectively.

Each model was first compared to a similar model without the introspective variable of preceding trial, with likelihood ratio test, to see if its addition improved the prediction of RT (or $Force$). All analyses were done with R package 'lmerTest' (Kuznetsova et al., 2017) and Python package 'statsmodels' (Seabold & Perktold, 2010).

4.4. Results

4.4.1. Effect of global metacognitive context on first-order behavior

RTs were shorter in the RT-group ($t(32) = -4.28, p < 0.001$, Fig.4.2.B) and variabilities tended to be higher in the Force-group ($t(37) = -1.78, p = 0.08$, Fig.4.2.C). Accuracy was also slightly lower in the RT-group (mean=0.87) compared to the Force-group (mean=0.90) but this difference was not significant ($t(36) = -1.6, p = 0.12$). Produced forces were higher in the Force-group ($t(24) = -2.64, p = 0.01$, Fig.4.2.B). This difference in mean force was not due to difference in maximum voluntary force ($t(29) = 0.35, p = 0.73$, see methods of chapter 3 for details on force measurement). Force variability was also higher in the Force-group ($t(26) = -5.12, p < 0.001$, Fig.4.2.C).

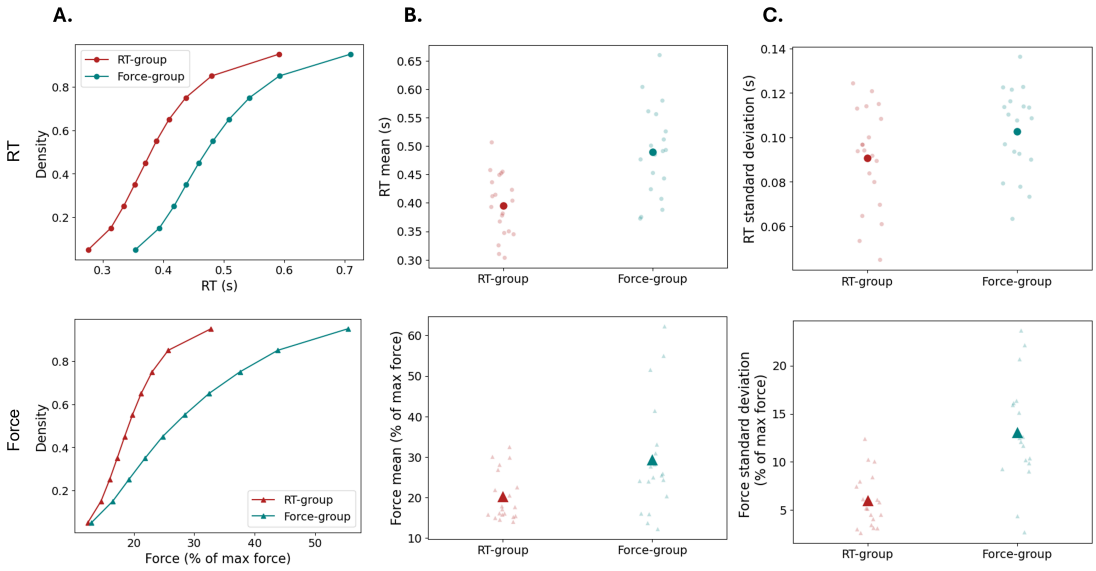


Figure 4.2. *(A) Cumulative density functions of RTs and forces for both groups, obtained after averaging of vincentized distributions of each participant. (B) Comparison of mean RTs and forces between groups. (C) Comparison of RT and force standard deviations between groups.*

Force CVs were higher in the Force-group ($t(27) = -5.4, p < 0.001$, Fig.4.3.A) but RT

CVs did not differ between groups ($t(32) = -1.52, p < 0.14$, Fig.4.3.A).

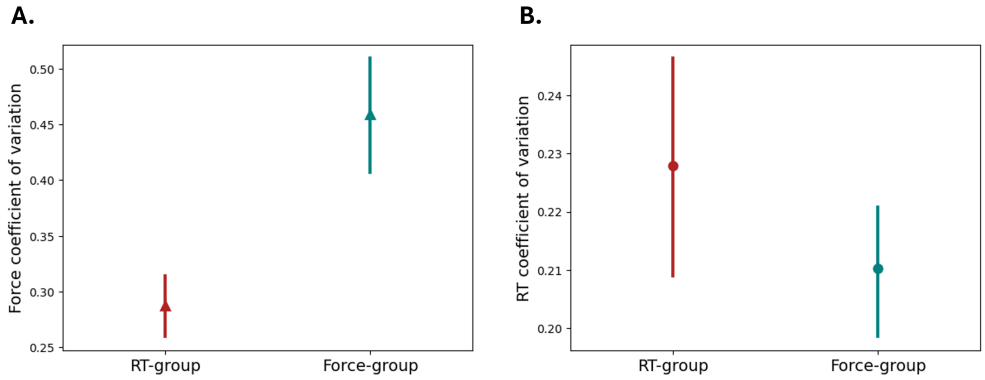


Figure 4.3. *(A) Force and (B) RT coefficients of variation of both groups.*

Finally, we looked at correlations between both variables (also reported in the Supplementary Material of the article in section 3.2). In the RT-group, repeated measure correlation revealed a small positive correlation between RT and force ($r_{rm}(8314) =$ à.09, $p < 0.001$). Individual Pearson r values ranged from -0.21 to 0.3 with a mean of 0.06 and a standard deviation of 0.13 (Fig.4.4.A). These low r values coupled with a high variability between participants suggest that the correlation is most likely non-existent. Conversely, in the Force-group, the same analysis revealed a negative correlation between RT and force ($r_{rm}(8671) = -0.18, p < 0.001$). Individual Pearson r values range from -0.45 to 0.13 with a mean of -0.16 and a standard deviation of 0.16 (Fig.4.4.B). To compare these correlations between groups, we applied fischer transformation to individual pearson r values. They were lower in the Force-group ($t(34) = 4.71, p < 0.001$, Fig.4.4.C).

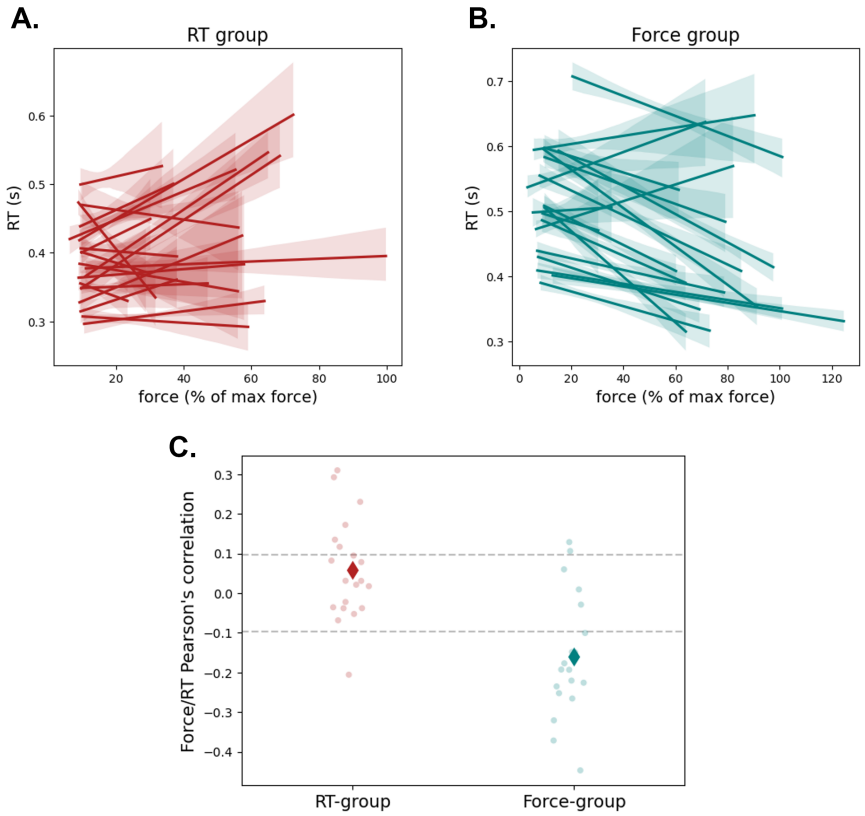


Figure 4.4. *Individual trial-by-trial correlations between RT and force in RT-group (A) and Force-group (B). Each regression line represents a participant. (C) individual Pearson's r values of force-group (blue) and RT-group (red). Dashed grey horizontal lines represent critical r values for alpha=0.05 and n=400.*

Note that in another group of participants who performed a similar RT estimation task (data of Experiment 2 of chapter 2), RT/force correlations were also weakly positive. As two conditions of force thresholds (weak and strong) were used in this experiment, we computed the correlations independently for both (weak force:$r_{rm}(9098) = 0.06, p < 0.001$; strong force: $r_{rm}(9288) = 0.17, p < 0.001$). This suggest that the difference observed between RT-group and Force-group reported above is probably not a group effect, but effectively due to the different metacognitive instructions.

4.4.2. Effect of local metacognitive judgments on subsequent behavior

With the data of the RT-group of chapter 3, the model including iRT_{n-1} outperformed the autoregressive model ($\chi^2(1) = 23.29, p < 0.001$). It revealed an effect of difficulty ($\beta = 0.075, t = 30, p < 0.001$), simply reflecting the fact that RT are slower in difficult conditions. RT_{n-1} also significantly predicted RT ($\beta = 0.16, t = 11.82, p < 0.001$), revealing the autocorrelation. More interestingly, effects of $difficulty_{n-1}$ ($\beta = -0.008, t = -2.9, p = 0.004$) and iRT_{n-1} ($\beta = -0.067, t = -4.83, p < 0.001$, see Fig.4.5.B)

were also observed. Negative coefficients mean that RT decreases with increasing iRT_{n-1} and $difficulty_{n-1}$.

The same model applied to the data of the Force-group with force and iForce values showed similar results. Model including $iForce_{n-1}$ outperformed the autoregressive model ($\chi^2(1) = 50.16, p < 0.001$). An effect of difficulty ($\beta = -7.09, t = -25.55, p < 0.004$) was observed, reflecting the fact that produced force are weaker in the difficult condition. As for RT, $force_{n-1}$ predicted force ($\beta = 0.41, t = 31.45, p < 0.001$), with the positive coefficient indicating a positive autocorrelation for force. Effects of $difficulty_{n-1}$ ($\beta = 2.8, t = 9.67, p < 0.001$) and $iForce_{n-1}$ ($\beta = -0.012, t = -7.09, p < 0.001$, see Fig.4.5.B) were also observed.

However, with data of chapter 2 (see Fig.4.5.A), models including iRT_{n-1} did not outperform the autoregressive models (weak force:$\chi^2(1) = 0.61, p = 0.43$, strong force:$\chi^2(1) = 0.61, p = 0.44$). They revealed similar effects of difficulty (weak force: $\beta = 0.019, t = 22.69, p < 0.001$; strong force: $\beta = 0.019, t = 24.03, p < 0.001$) and RT_{n-1} (weak force: $\beta = 0.2, t = 16.7, p < 0.001$; strong force: $\beta = 0.1, t = 9.43, p < 0.001$) for both force levels. $Difficulty_{n-1}$ predicted RT in the weak force condition ($\beta = -0.005, t = -5.54, p < 0.001$) but not in the strong one ($\beta = -0.001, t = -1.13, p = 0.26$). But, for both force conditions, iRT_{n-1} failed to predict RT (weak force: $\beta = -0.008, t = -0.78, p = 0.43$; strong force: $\beta = -0.008, t = -0.78, p = 0.44$).

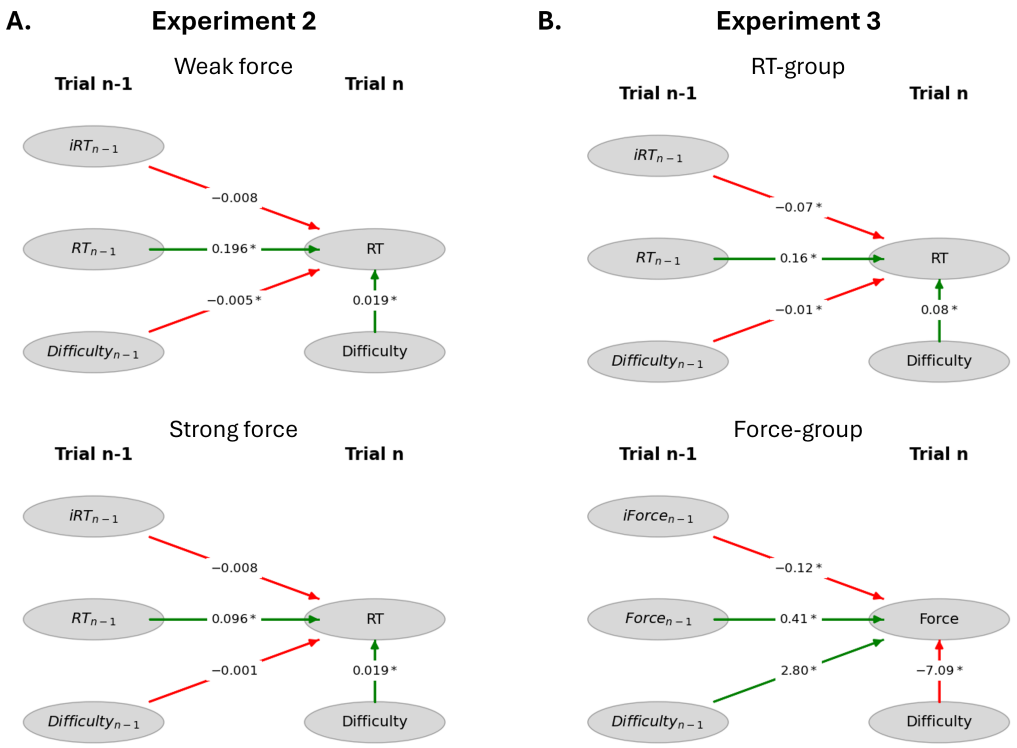


Figure 4.5. *Results of mixed models 4.1 and 4.2 run on data of Experiment 2 (A) and Experiment 3 (B). Grey nodes represent predictors of the models. Left columns represent* $trial_{n-1}$ *and right columns* $trial_n$*. Arrows indicate predictors' influence on the dependant variables and are associated with the corresponding coefficient from the model. Green arrows represent positive link and red arrows negative ones. Asterisks are associated with significant predictors.*

4.5. Discussion

The present results show that second-order judgments influence first-order performance at different scales. First, comparing the two groups of the EEG experiment described in chapter 3 revealed global differences in first-order performance. Indeed, RTs were faster in the RT-group, but they had higher error rates, reflecting the typical speed-accuracy tradeoff observed when speed is emphasized over accuracy (Heitz, 2014). Metacognitive instruction regarding RT likely heightened focus on RT, with its explicit awareness helping to optimize behavior toward faster responses. This is in accordance with the observation of slower RT in situation of accuracy judgment (Grützmann et al., 2014). We might hypothesize that the RTs in both groups would be higher than in a task without metacognitive judgments, as these judgments likely increase attentional demand and create a dual-task-like situation (S. W. Brown, 1997; Pashler, 1994). Although this hypothesis cannot be directly tested since none of the experiments included a condition without judgment, the difference between the groups

suggests it may not be the case. It is also possible that any slowing due to the dual-task situation in the RT-group was offset by the enhanced focus on RT. Similarly, the increased produced force observed in the Force-group could indicate a heightened attention towards this dimension, with participants pressing harder to ensure they crossed the threshold.

While a group effect cannot be ruled out, some evidence points to strategic influences. In addition to mean differences, the Force-group also showed a higher variability in produced force, but a RT variability comparable to that of the RT-group. Wagenmakers-Brown law predicts a linear increase of RT variability with its mean (Wagenmakers & Brown, 2007). Similarly, isometric produced force variability increases as a function of force level (Jones et al., 2002; Slifkin & Newell, 1999). Hence, differences in means should be associated with differences in variability. To account for this, we also compared coefficients of variations, which were higher for force in the Force-group. This indicates that the increased variability in produced force in the Force-group is not solely due to participants having higher produced forces. Regarding RT, the coefficients of variations did not differ between groups but there was a tendency for them to be higher in the RT group. While such an increase in variability would hardly optimize first-order performance, it could make the second-order judgment easier. Whether this reflects a strategy by participants to intentionally vary their performance, allowing them to make more informed judgments, remains to be explored.

In addition to these global differences in mean and variability, we observed a difference in the RT/force correlation of both groups. RT-groups exhibited a weak positive correlation, while Force-group showed a negative one. While this effect remains unexplained and would need further investigation, it supports the idea that the metacognitive instruction modifies the way the first-order task is processed.

On the trial-to-trial scale, the present results also showed that subjective judgments influenced subsequent performance, for both force and RT. However, this was not the case with data of chapter 2, where iRT of preceding trial did not predict RT. One major difference was the number of experimental conditions. First, there was two different levels of force (2 different thresholds), but this manipulation being blocked, it most likely did not influence sequences of events. Second, and contrary to experiment 3, 5 stimulus orientations could be possible. This higher number of difficulty levels necessarily captured RT variability and complexified sequential effects. In such case, it could become much harder to control subsequent RT, because it was already influenced by an important number of factors.

With data of chapter 3, these effects of preceding judgment were observed beyond effects of preceding first-order performance. The autocorrelation observed in both RT measure is consistent with previous findings (Laming, 1979a), indicating that performance in a given trial is influenced by performance in preceding trials. Interestingly, a similar positive autocorrelation was observed with produced force. These autocorrelations reflect a form of temporal dependency, where the current response is not entirely independent but rather shaped by the ongoing sequence of cognitive and motor demands (Gilden, 2001). Manipulation of perceptual difficulty also affected subsequent RT and force. Results showed a tendency for faster (stronger) responses

after difficult trials and slower (weaker) responses after easier trials. As both RT and difficulty can be confounded with iRT, they had to be considered in the assessement of these metacognitive sequential effects.

The modification of RT or force following a subjective judgment can be interpreted as a form of cognitive control, wherein the awareness of one's performance is used to optimize subsequent behavior. Participants tended to speed up after a high iRT, and slow down after a low one, independently of the preceding RT. Such a behavior would allow to increase speed so as to be closer to the instruction while avoiding too fast responses with higher error risk. Similarly, the sequential effects observed on force could reflect a use of the subjective judgment to optimize produced force toward a value that is not too strong, to minimize effort, but is strong enough to exceed the threshold and avoid omissions. This dynamic adjustments suggest that metacognitive awareness plays a crucial role in the real-time regulation of behavior, allowing individuals to optimize their actions in line with the task's demands.

These trial-to-trial variations are also in line with a strategic increase in variability. Participants may intentionally adjust their RT or force after assessing themselves as too fast/slow or too weak/strong, not necessarily to optimize their first-order task performance but rather to enhance their second-order performance. By doing so, they adapt their behavior to better calibrate their self-assessment on the visual scale, fine-tuning their responses to align more accurately with their subjective judgments. Such adjustments could be crucial for second-order perception, much like how perceptual systems are built to detect changes in the environment. This strategic variability ensures that participants are not only aware of their performance but are also actively modifying it to improve the accuracy of their metacognitive evaluations, since giving an accurate judgment is part of the task goal.

Overall, the findings underscore the interdependence between first-order cognitive processes and second-order metacognitive judgments. Modifying one inevitably influences the other, which has important implications for both experimental design and cognitive training. For example, training aimed at enhancing metacognitive skills could, in turn, improve first-order task performance by promoting more accurate self-monitoring and strategic adjustments (Schaeffner et al., 2021), and thus be used for cognitive remediation (Cella et al., 2015).

However, this interdependence also presents a challenge at the experimental level. When metacognitive judgments are explicitly required, they may alter the very cognitive processes being studied, thereby complicating the interpretation of results. This highlights the importance of considering metacognitive factors in the design and interpretation of cognitive experiments.

5. A new method for the online detection of EMG activity

Summary

5.1. Note on the present chapter . . . 134
5.2. Towards an online detection of EMG onsets . . . 135
5.3. Introduction . . . 136
5.3.1. EMG as a marker of interest for cognitive architecture and cognitive control . . . 136
5.4. Existing methods of EMG onset detection . . . 138
5.5. General constraints of online processing . . . 139
5.6. Core algorithm and specific constraints . . . 140
5.6.1. Specific constraints of our acquisition system . . . 140
5.6.2. Core detection algorithm . . . 140
5.7. Technical solutions tested for online EMG onset detection . . . 142
5.7.1. OpenVibe . . . 142
5.7.1.1. Difficulties encountered . . . 142
5.7.1.2. Developing a Custom Python Box . . . 142
5.7.2. Full custom python program through TCP/IP protocol . . . 144
5.7.2.1. Using the offline pseudo-online code to optimize detection parameters . . . 147
5.7.3. Other potential solutions to solve the remaining issues . . . 147
5.8. Conclusion . . . 149
5.9. Appendix . . . 151
5.9.1. OpenVibe python box . . . 151
5.9.2. Python code for online processing . . . 153

5.1. Note on the present chapter

One of the main objectives of the thesis was to assess the contribution of the MPFC, whose monitoring activity was reflected in the Ne, to subjective evaluation of RT. Chapter 3 allowed to establish that the Ne covaries with RT/PMT only when a metacognitive judgement is requested on RT (chapter 3.2), providing strong arguments for its involvement in such a representation.

Although we initially hypothesized that the Ne should also covary with the subjective counterpart, iRT, no such relationship was found (chapter 3.3). Although, based on

results of the experiment reported on chapter 2, this abscence of correlation is not really surprising, we could not firmly establish a link between metacognitive timing judgments and MPFC activity.

More generally, we cannot bring strong evidence with correlations only. Indeed, to dig further on the question, we need causal evidence: if the Ne is (at least part of) the signal that allows our RT to be evaluated, blocking its development should wipe out its informational content, and hence impair RT evaluation. To this aim, we initially planned to run a Transcranial Magnetic Stimulation (TMS) experiment. TMS can induce so-called "virtual lesions" by transiently disrupting the normal functioning of the focally stimulated structure. One challenge is to determine the precise timing of the TMS pulse. Triggering the TMS at a fixed time after the stimulus would not be sufficiently accurate to block the Ne because of the intrinsic variability of RT. The Ne starts developing about 30ms after EMG onset, so, time-locking the TMS to the EMG onset would be the most precise and optimal solution. This implies detecting EMG onsets online in (almost) real time (*i.e.* in less than 30ms). Hence, before being able to run such an experiment, we had to work on the development of a new method that would allow us to reach this temporal requirement.

While we could not have this criteria met on time to run the TMS experiment, we worked on the development of a new pipeline, adapted from the ones that can be run offline, to use for an online detection of the EMG onset. The current chapter presents our efforts towards reaching this goal, along with difficulties encountered, the current state of this development and future directions to overcome those difficulties. Once effective, this method, will offer broad perspectives for its application in cognitive psychology/neuroscience.

5.2. Towards an online detection of EMG onsets

Towards an online detection of EMG onsets

Nathalie Pavailler & Boris Burle

November 27, 2024

Abstract

Behavior results from cognitive processing, which is not directly accessible. Obtaining intermediate neurophysiological indices between inputs and outputs helps to constrain models of these cognitive processes. However, these neurophysiological indices are typically processed offline, leading to, at best, correlational analyses. Detecting relevant neurophysiological indices online, in (almost) real time, allows for the investigation of more causal relationships. Here, we focused on a specific neurophysiological marker that has proven very useful in the field of cognitive control: the electromyographic (EMG) activity of muscles involved in responses during Reaction Time tasks.

1 Introduction

To interact with the world, one needs to transform sensory inputs into behavioral outputs, through a series of information processing steps. One generally infers the nature of the underlying processes by analysing the relationship between inputs and outputs (How long did it take? Was the output correct? etc.). Such behavioral analysis can be augmented by neurophysiological indices allowing to infer the timing and/or the properties of intermediate processing (see, for examples Kutas et al., 1977; Renault et al., 1982; Schall, 2004). Many indices can be used, derived from different measures, such as EEG, single unit recording etc., that differ in their temporal and spatial resolutions (Churchland and Sejnowski, 1988). Although very usefull, their signal-to-noise ratio is often too low to be able to extract reliable information at the single measurement level (typically an experimental trial). It is therefore necessary to extract the information by combining different repetitions (averaging, peri-stimulus time histogram etc.). One neurophysiological measure, electromyographic activity, in contrast, has a excellent signal-to-noise ratio.

1.1 EMG as a marker of interest for cognitive architecture and cognitive control

In manual RT tasks, responses are typically given by key presses on dedicated response button (Fig. 1). EMG is the last physiological measure before overt behavior, that is button press. Although it is a peripheral measure (that can be considered as rather far from the brain), it has proven very fruitfull to decipher some of the properties of the cognitive architecture (Coles et al., 1985; Gratton et al., 1992; Hasbroucq et al., 2001; Possamaï et al., 2002), and to investigate

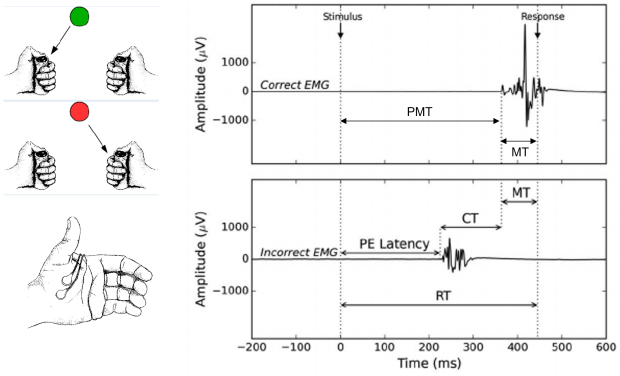


Figure 1: EMG use in RT task, with examples of RT fractioning and partial error.

cognitive control (Burle et al., 2002; Bonini et al., 2014). Furthermore, while the link between some brain activities and cognitive processes might sometimes be complex (Teller, 1984; Schall, 2004), the functional interpretation of EMG is (rather) straightforward, and one generally agrees that it is tigthly linked to the activation of its corresponding response.

Furthermore, EMG enables to reveal hidden processes. The high signal-to-noise ratio allows to decompose the RT of every single trial into two sub-components: the time from stimulus onset to EMG onset (pre-motor time, PMT), and the time from EMG onset to the behavioral response (motor time, MT, Botwinick and Thompson, 1966; Burle et al., 2002, figure 1). PMT is associated with decision time and studies using this decomposition have shown that MT was undoubtedly linked to the voluntary motor command (Hasbroucq et al., 1995; Possamaï et al., 2002) and that it contributes significantly to RT variability.

In choice RT tasks, EMG recordings also allow to detect, in approximatly 15 % of trials, small muscular activities on the side of the incorrect response, preceeding the overt correct response (figure 1). Too small to reach the response threshold, such 'partial errors' reveal the activation of the incorrect response. They have been interpreted as errors that are detected, inhibited and corrected before triggering an overt incorrect response (Coles et al., 1985; Scheffers et al., 1996; Hasbroucq et al., 1999; Burle et al., 2002) and are important indices of online cognitive control.

RT fractioning and partial errors detection are usually performed offline, limiting the analyses to correlations. Detecting these EMG bursts online would allow to interfere with the activation of the motor response, be it correct or erroneous, and hence to have some more causal approaches (close-loop).

However, "online" often refers to "shortly after the actual movement", "shortly" being undefined. Depending on the needs, this delay can be crucial and determines the method to be used (see below). Our present goal was to detect EMG

onsets in real-time to be able to block the development of a given cerebral activity with Transcranial Magnetic Stimulation (TMS). The targeted activity is the the Error Negativity (Ne) that occurs time-locked to EMG onset, with a latency of about 30 ms.

2 Existing methods of EMG onset detection

Recently, Carvalho et al. (2023) performed a review of the EMG onset detection methods that could be applied online. Note, however, that among the 247 studies reported, only 25 really used **online** detection for robotics purposes; the other ones were tested offline. Note also that their temporal efficiency is not reported.

Carvalho et al. (2023) identified three main families of methods for EMG onset detection:

- Visual inspection: detection of EMG onset relies on the expertise of the experimenter. While this is sometimes criticized for objectivity concerns, it remains the "gold standard" to which all methods are compared. However, this approach is, obviously, not applicable online.

- Threshold based: this simple approach considers that an EMG burst is present if the amplitude of the recorded activity crosses a threshold (see below for more details). This family of methods is simple, both conceptually and computationally, and rather easy to implement. It remains the most used tool for onset detection. Within this broad family, three types of threshold based detection are identified: single (ST), double (DT) and adaptive (AT) thresholds. ST compares the amplitude of the EMG signal with a previously selected threshold. The onset is detected when the EMG amplitude is higher than the threshold. DT adds a second threshold to determine the muscle activation onset time, with the final goal of avoiding false positives and enhance EMG onset detection precision: the second threshold is often the minimal number of samples that crosses the threshold for the EMG burst to be considered as present. AT segments the signal using the signal-to-noise ratio to adapt the threshold of muscle activation by windows, which might enable a more precise EMG onset detection over time.

- Statistical (should be better termed "probabilistic"): this family of approaches estimates the likelihood of an EMG burst. Based on the likelihood ratio between the distribution of noise ($H0$) and of signal ($H1$). If the likelihood of $H1$ crosses a given threshold, EMG activity is detected. Note that this method requires a reliable estimation of both $H0$ and $H1$. When applied online, this implies that the estimations are bad at the beginning of the segment of interest (few points) and increase as samples are being accumulated. Hence, the reliability of the method, when applied online, likely changes over time.

Note that other detection methods (not documented in Carvalho et al., 2023) exist, but their computations prevent their use online (for example, "Integrated Profile" Liu and Liu, 2016, that proved to be very resistant to noise or the "Second order moment" approach, Journée et al., 2008).

Besides detection algorithms, pre-processing is important to improve the quality of EMG before applying the onset detection method itself. Several pre-processing steps are evaluated by Carvalho et al. (2023):

- Envelope: either low-pass filter or moving average
- Teager-Kayser operator (TKEO, Li et al., 2007; Solnik et al., 2010): estimates the energy of the signal. The TKEO operator Ψ of a signal x is defined as $\Psi[x(n)] = x(n)^2 - x(n-1) \times x(n+1)$

Another pre-processing step might have to be taken into account, depending on the acquisition system. EMG signals are high-frequency components, without any activity below 10Hz. Most acquisition system devoted to EMG implement a high pass filter at acquisition time. But if EMG is recorded at the same as other physiological measures (with different frequency contents), the actual frequency span of the recorded signal might be much larger, and the recorded signal might be contaminated by low frequency components, that have to be suppressed.

3 General constraints of online processing

Online detection of EMG onset is associated with several constraints.

- Online detection implies working on stream of data, of which only the past is known.
- one needs to detect *as fast/early as possible*, or at the very least whithin a given time window.
  - onset detection is necessarily after the real onset (as one needs to have identified it).
  - how much *after* depends on different factors:
    - the sampling frequency (the higher, the smaller the delay)
    - the way data are delivered by the acquisition system: some deliver sample–by–sample, some by chunks of samples, whose size depend on the system (or settings)
    - the delay introduced by the algorithms

All those delays can be brought into one single *time unit* (TU): the intrinsic, hardware-based, TU_{intr} can be defined as:

$$TU_{intr} = n_{chunk} \times \frac{1}{sf} \tag{1}$$

where n_{chunk} corresponds to the number of samples within a chunk of data; the minimal being 1 (in this case the acquisition systems delivers individual samples), and $\frac{1}{sf}$ corresponds to the sampling interval; for example, if the sampling frequency (sf) is set to 2048 Hz, the duration of a sample will be 0.488 ms. If $n_{chunk} = 1$ (one sample at a time), and $sf = 2048$ Hz, TU_{intr} will hence be 0.488 ms, but, for the same sampling frequency, if the acquisition system delivers chunk of, for example, 8 samples, the TU_{intr} will be $8 \times 0.488 = 3.906$ ms.

Besides this TU_{intr}, which is determined only by the acquisition system, the detection algorithm may requires more than one TU_{intr}, hence introducing additional delays. For example, if, after having suspected an onset, the algorithm needs another (or some other) data to decide, the actual TU_{act} detection time will be delayed by $v \times TU_{intr}$, where v determines how many validation data are needed. To generalize, if the algorithm needs v chunks, TU_{act} will be:

$$TU_{act} = v \times TU_{intr} \tag{2}$$

Obviously, the choice of the algorithm will depend on TU_{act}, as it must be inferior to the acceptable delay in the onset detection.

So far, the duration of processing operations implemented in the algorithms are not considered as inducing any further delay. Obviously, those operations take time, and it will be important to estimate them. Note, however, that as long as the processing time for one chunk of data is inferior to TU_{act}, this assumption seems plausible, and since TU_{act} is pretty large (for a computer clock...), we will see later that this assumption generally holds.

4 Core algorithm and specific constraints

4.1 Specific constraints of our acquisition system

Our acquisition system is a Biosemi active 2 system, which imposes several important preprocessings:

1. data are acquired broad band, between DC and nyquist frequency (since the sampling frequency was set at 2048 Hz, the nyquist value is 1024 Hz, ; actually Biosemi starts filtering large before this value to avoid filtering artefacts). A high-pass filter has to be implemented to remove slow frequency bands (note that this is not trivial to apply a 10 Hz high-pass filter on data of about 4 ms long... see code for details).

2. EMG is acquired in a monopolar way by two independent channels (EXG1 and EXG2); a bipolar montage has to be implemented as EMG = EXG1 - EXG2.

3. since a TKEO operator would consume one chunk of data (see above), a less efficient, but faster, signal booster can be used, by simply squaring the signal.

4.2 Core detection algorithm

As we chose to use an adaptive threshold to detect EMG onsets, two periods have to be defined in each trial. A baseline period, without any expected activity, needs to be processed to calculate the threshold and update its value at each trial. Then, once the stimulus is presented, one needs to check whether each received chunk of data exceeds the threshold (see figure 2).

On the pre-processed data stream, the following processing operations have to be implemented:

1. when a stimulation value coding for the onset of the baseline period $Code_{bl}$ is received:

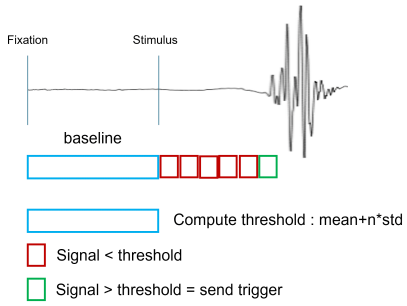


Figure 2: Timeline of an individual trial with associated computations needed for EMG onset detection.

- segment 1 second of the stream after $Code_{bl}$
- compute the mean (m) and the standard deviation (sd) of the segment

2. when a stimulation value coding for the stimulus $Code_{stim}$ is received:
   - for each chunk of data x, compute if $|x| > m + n \times sd$ (n needing to be determined per participant)
   - check that this condition remains true for m value (m needing to be determined *a piori*)
   - if so, send a trigger that marks the onset of the EMG burst

A more formal description is given by algorithm 1.

Algorithm 1 Algorithm for online EMG detection in a RT task

```
initialize Onset to False
high-pass filter (10 Hz)
set bipolar montage
square signal (x → x^2)                                   ▷ optionnal
if Received_Stimulation == Code_bl then
    Segment 1 second of signal
    compute mean (m) and standard deviation (sd) of the segment
else if Received_Stimulation == Code_stim then
    for each new chunk c do
        if c > (m + n × sd) then
            if Onset == 0 then
                Send Stimulation EMG_onset
                Onset = 1                    ▷ so that trigger is sent only once
        else
            continue
```

We tried to implement this online detection with two technical solutions presented below.

5 Technical solutions tested for online EMG onset detection

5.1 OpenVibe

OpenVibe (Renard et al., 2010) is a general purpose Brain-Computer-Interface (BCI) and Online Neuroscience system. It relies on "scenarii", that are made of elementary processing boxes implementing more or less complex functions that can be combined into more or less complex processing pipelines (see figure 4). A great interest of OpenVibe, despite its rather mature state, is that it allows to replay, in realtime, as if it was really online, data and scenarii, allowing to easily tune algorithms on the very same data.

In our case, the pipeline to implement was, a priori, pretty simple: one first needs to compute the mean m and standard deviation sd of a baseline period (before the stimulus), and then, after the stimulus, compare if the incoming signal was larger than $m + n \times sd$. Several pre-processing steps and transformations were needed before detection (see above). We implemented these different processing steps by organizing corresponding processing boxes in an Openvibe scenario (figure 4).

5.1.1 Difficulties encountered

However, we faced several problems while developping the pipeline, some due to our poor knowledge of online processing, and some due to unforeseeable temporal limitations of OpenVibe.

First, we faced issues related to asynchronuous processing. Indeed, m and sd are extracted and used at different moments. While this is not a problem on offline systems (the values of the variables are just "reserved" for later use), it turned to be a non trivial problem to deal with in OpenVibe since everything happens "in time". Since debugging (even very simple things such as checking the content of the variables, for example) is very complex in an online system (or at least we did not manage to do it easily...), it has been very difficult to set-up a pipeline with asynchronous processing.

Second, while an OpenVibe box allowing to compare streaming values to a threshold does exist, this threshold is hard coded in the box, and cannot be changed based on operations performed on the data. This is, however, precisely what we want to do, that is to compute m and sd, and hence the threshold ($m + n \times sd$) on every trial to adjust the threshold to the background noise. We tried different "tricks" to overcome this limitation, implemented in different scenarii, but did not succeed.

After discussing with different OpenVibe experts, it turned out that the only way to do what we wanted was to build our own OpenVibe box. This is clearly at another level than simply connecting already programmed boxes... We, however, decided to take this endeavour. Although OpenVibe is written in C++, fortunately, it is also possible to write boxes in Python.

5.1.2 Developing a Custom Python Box

Based on existing examples, we slowly understood the different parts constituting an Openvibe box, the different internal OpenVibe objects, and the under-

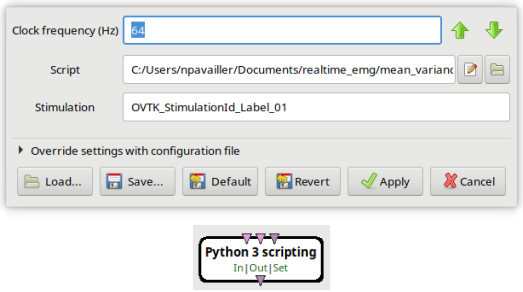


Figure 3: Graphical interface for OV boxes.

lying stream model.

An OpenVibe box is composed of two parts:

- a graphical interface, where are defined:
  - the inputs and outputs of the box. They can be of type:
    - ∇ signal
    - ∇ stimulation
    - ∇ streamed matrix
  - the clock of the box (*i.e.*, its processing resolution)
  - the location of the python script
  - the code of the stimulation (if any) to be sent as output
- the python script implementing the transformations

Our python box has three inputs, represented as the small colored arrows on top of the box (see figure 3):

∇ two signal imputs (rose arrows): one for the segment (baseline), one for the continuous following (detection period)

∇ a stimulation (purple arrow) to react to triggers sent by the computer running the task

and one output:

∇ the triggers to be sent once the EMG onset is detected (purple arrow at the bottom).

The full scenario, including the python box is represented on figure 4, and the python code is reproduced on appendix 6.

A critical parameter is the clock frequency of the box. The default value is 64Hz, which corresponds to about 16 ms. This is slightly too slow to detect EMG onset within 30 ms, as it allows for less than two cycles of the box. Therefore, we increased the frequency to 128Hz, which is just under 8 ms, providing more time for accurate detection.

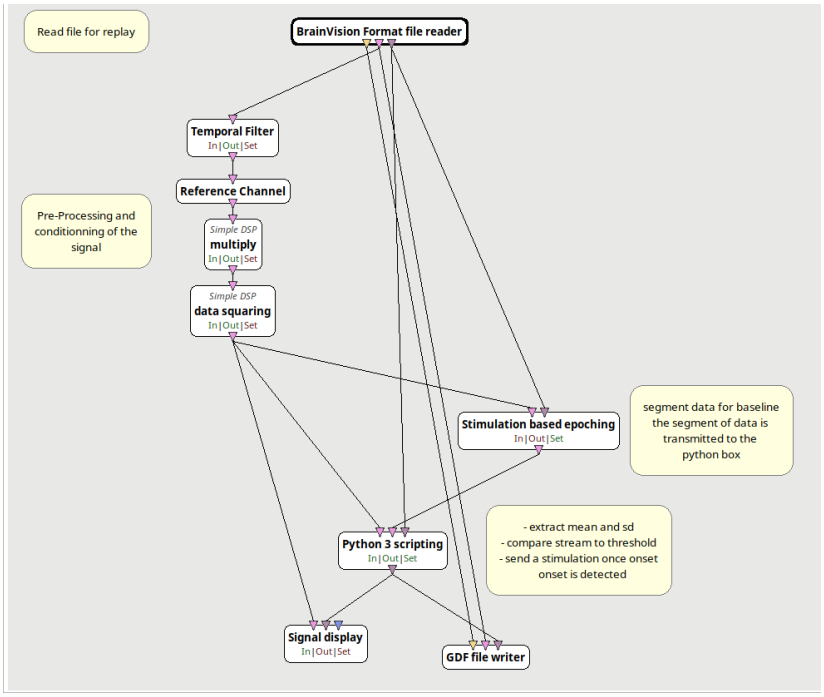


Figure 4: OpenVibe Scenario. First (top) box allows to replay in real time data from an existing file. Stream of data is then passed to preprocessing boxes. Preprocessed signal is sent to two different boxes: one to extract the baseline, and the custom-made Python box to compare the signal to the calculated threshold. Finally, a trigger is sent by this final Python box if an EMG onset is detected.

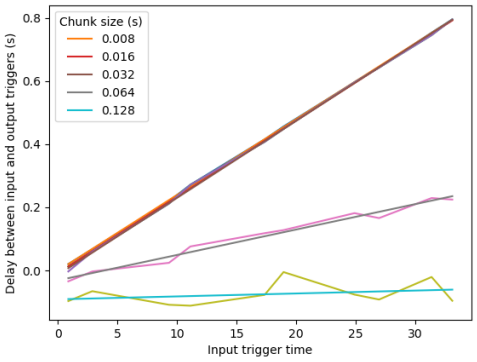


Figure 5: Delays between times of input and output triggers for different chunk durations (from 0.008s (128Hz) to 0.128s (8Hz)).

The initial tests demonstrated that the box was functioning quite well, but we noticed that while the detection was accurate at the start of the (replayed) experiment, the error increased over time, with the detected onset gradually drifting away from the actual onset. It seemed as though a cumulative error was building up; although the onset was detected every time, it appeared with an increasing delay compared to the real onset. After extensive investigation, we identified that this growing delay was related to the clock frequency: the faster the clock, the greater the drift (see Fig.5).

To achieve a relatively stable delay, we had to set the clock very slow, at 8 Hz (128 ms), which is far too long. We attempted to identify the source of this cumulative delay by varying the clock speed across scenarii of different complexity levels and by implementing or omitting certain critical functions. Eventually, we suspected that the issue might be related to an improper emptying of buffers. However, possibly due to our limited experience with OpenVibe, we were unable to eliminate the delay at acceptable clock speeds. Consequently, we decided to explore another technical solution.

5.2 Full custom python program through TCP/IP protocol

The Biosemi ActiveView acquisition software allows to transfer data through the TCP/IP protocol, to get the data on another computer in (almost) real time. On the receiving computer side, we used the pyActiveView library (in which we corrected some few bugs) to get and process the data.

To avoid resources limitations, we decided to have a dedicated computer for EMG onset detection. Since this dedicated computer would only process the EMG online flow, we used a raspberryPi device. The computers network and the flow of information is represented on figure 6.

Contrary to OpenVibe, in which many tools for online processing are implemented, we had to start from scratch, and develop everything. Furthermore, as indicated above, OpenVibe allows to replay the acquired file, in the very same

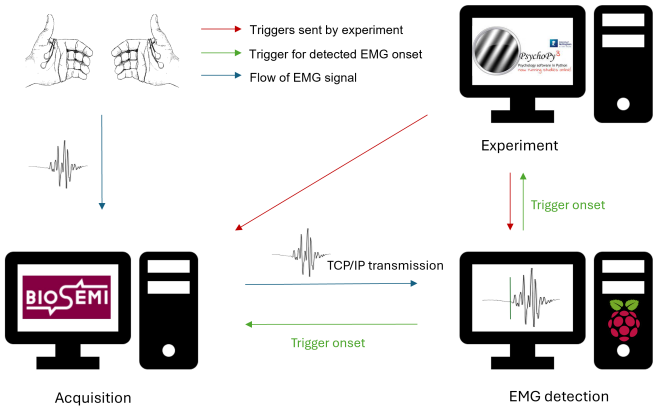


Figure 6: Transfer of information within our computers network.

conditions as if they were acquired online. To be able to optimize the detection algorithm, we developped both a version based on real online acquisition, and an adaptation of the same code for replay, simulating the stream of data. The code for online processing is presented on appendix 6, and was ran on the RaspberryPi.

While OpenVibe was tested using replayed data, the current implementation was tested online in a simple RT task. We tested only one participant performing 30 trials, due to the reasons outlined below. The detection method performed well, as all EMG bursts were detected. However, the onset position was noticeably delayed compared to the real onset marked visually offline (figure 7), with a mean delay of about 70 ms (mean: 70.5 ms, std: 12.34 ms).

Although unlikely, we investigated whether the delay could be caused by the computations performed. Using Python `time` module, we measured the time required for different operations. It was found that processing a whole chunk (corresponding to 8 samples, or just under 4 ms) took about 0.1 ms (filtering was the longest, with 0.08 ms), indicating that the computational load could not account for such a significant delay. To determine if the delay was due to the algorithm itself, we ran the same (slightly adapted) code offline to measure onset detection. These analyses, described in more detail below (section 5.2.1), revealed that the delay was between 6 and 20 ms, much shorter than the delay observed online.

In order to identify the source of this delay, we simplified the situation, and ended up with the simplest possible case: a trigger impulse is sent by the acquition system, and as soon as the rasperryPi detects it, a new trigger is sent in response. The observed delay between the two triggers turns out to be in the range comparable to the EMG (mean: 62.7 ms, std: 12.8 ms, figure 7). It hence appears that data sent through the TCP/IP protocol by the Biosemi computer is only accessible after a delay of about 60 ms, with a rather large variability.

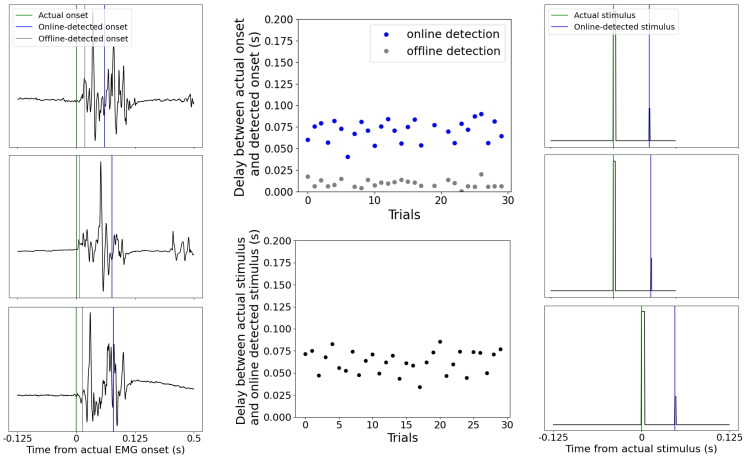


Figure 7: Delays between online and visually detected EMG onsets (top center) with individual trials examples (left), and between real and online detected triggers (bottom center) with individual trials examples (right).

Since this is an hardware delay, there is no way to reduce it. Interestingly, however, the standard deviations of the errors were extremely similar for EMG and simple triggers (12.34 vs 12.8 ms), suggesting that the detection algorithm did not add any variance to the one induced by the TCP/IP transmission. It means that, if one could get rid of the hardware delay, the detection is probably rather good. We will discuss below different possibilities to (try to) eliminate this delay.

5.2.1 Using the offline pseudo-online code to optimize detection parameters

As indicated above, we ran the very same processing script, with the same parameters (n value = 5, and at least 4 samples – out of 8 – being above the threshold) to assess which error one can expect. Note that the (pseudo) online detection necessary lags behind the real onset, since the onset has to already be present to be detected; accordingly, a negative difference indicates a false alarm. The mean error obtained was 8.14 ms, with a standard deviation of 7.61 ms (min= -23.7 ms, max= 19.96 ms). Note that only one detection was a false alarm (among only 28 analysed trials). Such a performance, which requires to be verified with more data, would clearly be sufficient for online use with, *e.g.*, TMS to block the Ne genesis.

We then explored the parameters space to assess some possible optimizations. We varied the n value (multiplier of the std) from 2 to 5, and the number of samples crossing the threshold within a chunk from 2 to 4. Results are summarized on figure 8. Remember that negative values reflect false alarms. First, decreasing n dramatically increased the number of false alarms, without improving much the timing of correct detection. Increasing the number of samples

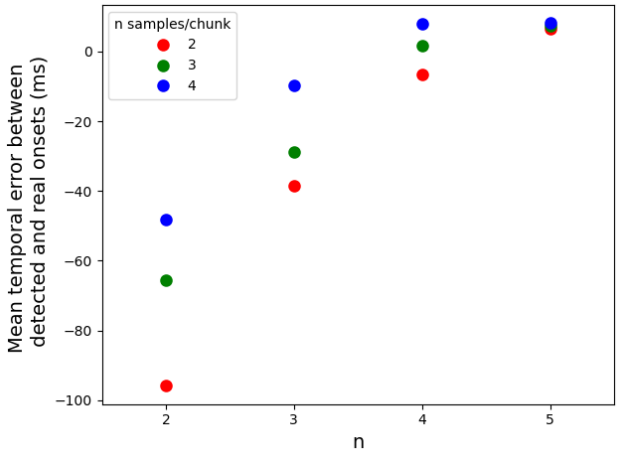


Figure 8: Mean error for different values of n and different numbers of samples crossing the threshold within a chunk (n samples/chunk). Negative values represent false alarms, with detections occuring before EMG onset.

above the threshold had a similar effect, but only for low n; for higher n, especially when $n = 5$, there is virtually no effect of this parameter. Although this parameters exploration should not be taken to draw strong conclusions, it is a proof-of-concept that our offline code can be efficiently used to explore parameters space on the very same data which is, obviously, not possible online (except if several computers are run in parallel to detect EMG with different parameters).

5.3 Other potential solutions to solve the remaining issues

None of the two technical solutions we tried ended-up with fully satisfying results. The OpenVibe solution, despite a great investment (such as designing a new box), led to systematic drift in onset detection that seems to be related to the box clock speed: it only "worked" (that is, with a small drift) for "very slow" clock (8 Hz), incompatible with our purpose. Although we could not firmly identify the origin of this drift, the numerous tests we made made us suspect that there was an issue in the way we handle, and empty, the signal buffer. For lack of time, we decided not to pursue our investigations and move forwards with another solution. However, all the "expertise" (or at least the better understanding of stream processing) acquired during the last years could maybe allow us to see this issue differently, and better address it. So, maybe we should give OpenVibe another try, with a fresh look.

Besides this drifting issue, one difficulty with OpenVibe was that even very simple debugging was complex. We thus decided to move to a solution where we master (almost) everything. Since Biosemi ActiveView software allowed to transmit the data stream through TCP/IP, we tried to understand how to collect it, and once we found out, we worked on writing code allowing to work on this

stream of data. The produced code seems to be working, both online and offline, allowing to test for optimal parameters. However, we realized (rather late) that the TCP/IP protocol delivers the data with a delay. While this delay (about 60 ms) is likely acceptable for most "online" application, it is incompatible with our needs. We tried to reduced this delay by optimizing the network connection, but the remaining one seemed irreducible. However, the code developped is easily adaptable to other incoming stream and, provided that other acquisition system has no (or much lower) delay, it should be re-useable.

One candidate to investigate is the Lab Streaming Layer ecosystem (`https://labstreaminglayer.org/`) that allows to acquire, and synchronize, different physiological and behavioral flows of data. Although some tests are necessary on its timing, to avoid last minute (bad) surprises, its description makes us confident that it would run fast enough for our needs. Furthermore, when looking at the provided example, it seems that our code would be very easy to adapt (just a few lines of code should be changed). We hope to soon be able to test this possibility to trigger TMS pulse right at EMG onset, and hence, for the current purpose, before the onset of the Ne. Besides this current project, this would open a lot of possibilities to investigate the role of this activity in performance monitoring and cognitive control in general.

6 Conclusion

In conclusion, the development and implementation of an online EMG onset detection system for RT tasks is constrained by different parameters, from the acquisition caracteritics to the algorithm of detection. With our specific constraints, notably being able to detect the onset in less than 30ms, we faced significant technical challenges, particularly in achieving real-time processing with minimal delay. Despite these obstacles, we managed to develop a python script with a working detection algorithm, which showed promising results in detecting EMG bursts, although some temporal delays remained an issue, with onset being detected too late. This script should be easily adaptable to be used with a different hardware, more precise in time. This work underscores the importance of carefully balancing algorithmic efficiency and system constraints in real-time neurophysiological applications, bringing us closer to a precise real-time detection of EMG onsets.

References

Bonini, F., Burle, B., Liégeois-Chauvel, C., Régis, J., Chauvel, P., and Vidal, F. (2014). Action monitoring and medial frontal cortex: leading role of supplementary motor area. *Science*, 343(6173):888–891.

Botwinick, J. and Thompson, L. (1966). Premotor and motor components of reaction time. *Journal of Experimental Psychology*, 71 1:9–15.

Burle, B., Possamaï, C.-A., Vidal, F., Bonnet, M., and Hasbroucq, T. (2002). Executive control in the Simon effect: an electromyographic and distributional analysis. *Psychological Research*, 66(4):324–336.

Carvalho, C. R., Fernndez, J. M., del Ama, A. J., Oliveira Barroso, F., and Moreno, J. C. (2023). Review of electromyography onset detection methods for real-time control of robotic exoskeletons. *Journal of NeuroEngineering and Rehabilitation*, 20(1).

Churchland, P. S. and Sejnowski, T. J. (1988). Perspectives on cognitive neuroscience. *Science*, 242(4879):741–745.

Coles, M. G., Gratton, G., Bashore, T. R., Eriksen, C. W., and 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.

Gratton, G., Coles, M. G., and Donchin, E. (1992). Optimizing the use of information: strategic control of activation of responses. *Journal of Experimental Psychology: General*, 121(4):480–506.

Hasbroucq, T., Burle, B., Akamatsu, M., Vidal, F., and Possamaï, C. A. (2001). An electromyographic investigation of the effect of stimulus-response mapping on choice reaction time. *Psychophysiology*, 38(1):157–162.

Hasbroucq, T., Mouret, I., Seal, J., and Akamatsu, M. (1995). Finger Pairings in Two-Choice Reaction Time Taskscolon: Does the Between-Hands Advantage Reflect Response Preparation? *Journal of motor behavior*, 27:251–262.

Hasbroucq, T., Possamaï, C. A., Bonnet, M., and Vidal, F. (1999). Effect of the irrelevant location of the response signal on choice reaction time: an electromyographic study in humans. *Psychophysiology*, 36:522–526.

Journée, H. L., Postma, A. A., Sun, M., and Staal, M. J. (2008). Detection of tremor bursts by a running second order moment function and analysis using interburst histograms. *Medical Engineering & Physics*, 30(1):75–83.

Kutas, M., McCarthy, G., and Donchin, E. (1977). Augmenting mental chronometry: The p300 as a measure of stimulus evaluation time. *Science*, 197(4305):792–795.

Li, X., Zhou, P., and Aruin, A. S. (2007). Teager-Kaiser energy operation of surface EMG improves muscle activity onset detection. *Annals of Biomedical Engineering*, 35(9):1532–1538.

Liu, J. and 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.

Possamaï, C.-A., Burle, B., Osman, A., and Hasbroucq, T. (2002). Partial advance information, number of alternatives, and motor processes: an electromyographic study. *Acta Psychol (Amst)*, 111(1):125–139.

Renard, Y., Lotte, F., Gibert, G., Congedo, M., Maby, E., Delannoy, V., Bertrand, O., and Lcuyer, A. (2010). OpenViBE: An Open-Source Software Platform to Design, Test, and Use BrainComputer Interfaces in Real and Virtual Environments. *Presence: Teleoperators and Virtual Environments*, 19(1):35–53.

Renault, B., Ragot, R., Lesevre, N., and Remond, A. (1982). Onset and offset of brain events as indices of mental chronometry. *Science*, 215(4538):1413–1415.

Schall, J. D. (2004). On building a bridge between brain and behavior. *Annual Review of Psychology*, 55:23–50.

Scheffers, M. K., Coles, M. G., Bernstein, P., Gehring, W. J., and Donchin, E. (1996). Event-related brain potentials and error-related processing: an analysis of incorrect responses to go and no-go stimuli. *Psychophysiology*, 33(1):42–53.

Solnik, S., Rider, P., Steinweg, K., DeVita, P., and Hortobgyi, T. (2010). TeagerKaiser energy operator signal conditioning improves EMG onset detection. *European Journal of Applied Physiology*, 110(3):489–498.

Teller, D. Y. (1984). Linking propositions. *Vision Research*, 24(10):1233–1246.

Appendix

OpenVibe python box

```
import numpy as np


class MyOVBox(OVBox):
    def __init__(self):
        OVBox.__init__(self)
        self.signalHeader = None
        self.stimLabel = None
        self.stimCode = None


    def initialize(self):
        # the stim label is taken from the box setting
       self.stimLabel = self.setting['Stimulation']
       # we get the corresponding code using the
                                        OpenViBE_stimulation
                                        dictionnary
       self.stimCode = OpenViBE_stimulation[self.stimLabel]
       # we append to the box output a stimulation header. This is
                                        just a header, dates are 0
                                        .
       self.output[0].append(OVStimulationHeader(0., 0.))


    def process(self):

        if (len(self.input[0])>0): # if input on raw stream,
                                         otherwise data = 0

            #input starts with a header
            if (type(self.input[0][0]) == OVSignalHeader):
                self.signalHeader = self.input[0].pop()
                #input length=+1 at each new chunk; .pop() returns
                                                the last element
                                                and suppress the
                                                others


            # check if next data block is signal (i.e. signal
                                                buffer)
            if (type(self.input[0][0]) == OVSignalBuffer):
                chunk=self.input[0].pop() # to remove the part of
                                                the header
                                                telling it is
                                                signal
                numpyBuffer=np.array(chunk) # the remaining data in
                                                 input are signal
                                                , we copy them in
                                                 a numpy array

                # different possible comparisons
                if all(numpyBuffer>threshold):  # if all elements
                                                in the array are
                                                above the
                                                threshold
                # if np.mean(numpyBuffer)>threshold: # if the mean
                                                of the array is
                                                above threshold
```

```
                # if (np.where(numpyBuffer > threshold)[0].shape[0]
                                                > m):   # if at
                                               least m point in
                                               chunk are above
                                               threshold
                    print(self.getCurrentTime())
#                           stimSet = OVStimulationSet(self.
                                getCurrentTime(), self.
                                getCurrentTime()+1./self.getClock
                                ())
                    # the date of the stimulation is simply the
                                                    current
                                                    openvibe time
                                                     when calling
                                                     the box
                                                    process
                    stimSet.append(OVStimulation(self.stimCode,
                                                    self.
                                                    getCurrentTime
                                                    (), 0.))
                    self.output[0].append(stimSet)

        if (len(self.input[1])>0):  # if in put on baseline epoch
            if (type(self.input[1][0]) == OVSignalHeader):
                self.signalHeader = self.input[1].pop()


            elif (type(self.input[1][0]) == OVSignalBuffer):  #
                                               starting to compare
                                               when there is a first
                                                epoch
                epoch=np.array(self.input[1][0])
                mean=np.mean(epoch)
                std=np.std(epoch)
                threshold=mean+3*std
                # n (=3) is hard coded. Should be better be
                                               adaptable
                stimSet = OVStimulationSet(self.getCurrentTime(),
                                               self.
                                               getCurrentTime()+
                                               1./self.getClock
                                               ())

        if (len(self.input[2])>0):
            if (type(self.input[2][0]) == OVStimulationSet):
                stim = self.input[2][0]

                if (len(stim)>0):
                    if (stim[0].identifier==33026):
                    # stim value is hard coded. Should be adaptable
                        self.output[0].append(stim)#identifier for
                                                        epoch end
                if (len(self.input[1])>1):
                    self.input[1].pop()

            self.input[2].pop()

        if len(self.input[1])>1:
            self.input[1].pop()

        if len(self.input[2])>1:
            self.input[2].pop()
```

```
box = MyOVBox()
```

Python code for Online processing

```
import sys
import RPi.GPIO as GPIO
import numpy as np
import scipy.signal as sg
import matplotlib.pyplot as plt
#from psychopy import parallel
import time

from pyactivetwo import ActiveTwo

GPIO.setmode(GPIO.BOARD)
GPIO.setup([7,13,29],GPIO.IN,pull_up_down=GPIO.PUD_UP)
#GPIO.setup(entree_7,GPIO.IN,pull_up_down=GPIO.PUD_UP)
GPIO.setup([16,18,22,36],GPIO.OUT)
GPIO.output([16,18,22,36],GPIO.LOW)


def send_trigger(code):
    GPIO.output(code,GPIO.HIGH)
    time.sleep(0.002)
    GPIO.output(code,GPIO.LOW)



#FIXME Maybe a parameter file?
sf = 2048 #FIXME: need to make it variable
nsamples = 8
nchannels = 8
acq_duration = 1
nchunks = (sf/nsamples)*acq_duration
#prepare array for online data
largeBaseline = 1.5 # in s, will contain the  filtering artefact at
                                    the beginning
endLargeBaseline = int(largeBaseline*sf) #FIXME check whether it is
                                    an int
usefullBaseline = 1 # largeBaseline on which the start has been
                                    removed to get rid
# of the artefact.
onsetUsefullBaseline = int((largeBaseline - usefullBaseline) * sf)
endUsefullBaseline = endLargeBaseline # just to get the script more
                                    readable

raw_data_array=np.zeros((int(largeBaseline*sf + nchunks*nsamples),
                                    nchannels))
filtered_data_array=np.zeros(int(largeBaseline*sf + nchunks*
                                    nsamples))

#device = ActiveTwo(host='147.94.106.83', sfreq=sf, port=8888,
                                    nchannels=nchannels,
#                         tcpsamples=nsamples)
device = ActiveTwo(host='192.168.1.1', sfreq=sf, port=8888,
                                    nchannels=nchannels,
                        tcpsamples=nsamples)
```

```
#define filter
cutoff = 10                 # high-pass cut-off: 10 Hz
forder=3                    #The order of the filter (Default 3).
Wn = 2. * cutoff / sf
b, a = sg.butter(forder, Wn, "highpass")

emg_threshold = 5           # corresponds to ''n'' in the algorithm

# value of the trigger once onset is detected
emg_onset = 36

onset = []                  # empty list to collect the onset times

outputdata=np.zeros((filtered_data_array.shape[0], 30)) #FIXME:
                                    last parameter should be number
                                    of trials
trial=0


GPIO.add_event_detect(7, GPIO.FALLING)
GPIO.add_event_detect(29, GPIO.FALLING)
GPIO.add_event_detect(13, GPIO.FALLING)

# number of points whithin a chunk that must be above threshold to
                                    consider EMG onset
consecutive_points  = 3 #FIXME: incorrect name, not necessarily
                                    consecutive

chan0 = 0
chan1 = 1

acquire=True

while acquire:
    device.read(duration=nsamples/sf)


    if GPIO.event_detected(7):
        thisOnset = 0
        # read the signal corresponding to the baseline in one
                                        block
        raw_data_array[0:int(endLargeBaseline),:]=device.read(
                                        duration=largeBaseline)

        filtered_data_array[0:endLargeBaseline] = raw_data_array[0:
                                        endLargeBaseline,chan0] -
                                         raw_data_array[0:
                                        endLargeBaseline,chan1] #
                                        FIXME define correct
                                        channel numbers
        filtered_data_array[0:endLargeBaseline] = sg.lfilter(b, a,
                                        filtered_data_array[0:
                                        endLargeBaseline])
        mean_bsl=filtered_data_array[onsetUsefullBaseline:
                                        endUsefullBaseline].mean
                                        ()
        std_bsl=filtered_data_array[onsetUsefullBaseline:
                                        endUsefullBaseline].std()

        # now read each chunck of data one after the other
```

```
        for chunk in range(int(nchunks)):

            start_full= endLargeBaseline + (chunk*nsamples)
            end_full=start_full+nsamples

            # bipolar derivation
            filtered_data_array[0:end_full] = raw_data_array[0:
                                              end_full,chan0] -
                                              raw_data_array[0:
                                              end_full,chan1] #
                                              FIXME define correct
                                              channel numbers
            # Filtering the whole signal
            filtered_data_array[0:end_full] = sg.lfilter(b, a,
                                              filtered_data_array[0
                                              :end_full])
            # Taking back only the filtered last chunck
            filteredChunck = filtered_data_array[start_full:
                                              end_full]

            sup = np.where(np.abs((filteredChunck-mean_bsl))>
                                              emg_threshold*std_bsl
                                              )
            if sup[0].shape[0] > consecutive_points :
                if thisOnset == 0:
                    send_trigger(emg_onset)
                    onset.append(end_full)
                thisOnset+=1
        outputdata[:, trial]=filtered_data_array
        trial+=1

    if GPIO.event_detected(29): # end of experiment
        acquire=False
```

6. General Discussion

"What we observe is not nature in itself but nature exposed to our method of questioning."
– Werner Heisenberg

Throughout this thesis, we delved into the processes contributing to metacognitive representations of RT. On one hand, we investigated the nature of these representations, by dissociating the different sources of information on which iRT are built. On the other hand, we searched for the potential second-order processes allowing participants to access their RT. This was organized around three main questions. First, we aimed to determine whether iRT represents a direct temporal readout of RT or whether it is inferred from external cues. Second, we explored which first-order processes underlying RT are accessible for iRT judgment. Lastly, we examined whether RT estimation relies on a generic performance monitoring system, as indexed by the Ne. This was achieved by asking participants to estimate each of their RT during a choice reaction time task, in several experiments, grouped in 3 empirical chapters. We will first recap the main findings and then provide a integrative perspective on what they can reveal about humans ability to estimate the duration of their own actions.

6.1. A general overview of the main findings

6.1.1. iRT as an accurate representation of RT

Assessing the accuracy of RT estimation requires measuring the extent to which iRT effectively tracks RT duration. As stated in the General Introduction, when asking a participant to estimate a duration, he/she can make two types of errors in the estimate. First, from one trial to the other, estimations can be shorter or longer than the duration to estimate, leading to some variability. Second, global under or overestimations can occur, when one systematically reports shorter/longer estimates.

Previous studies have found that participants were able to estimate their RT trial-by-trial variability, but they systematically underestimated their actual duration (Corallo et al., 2008). In the present studies, such underestimation was not observed. However, with the VAS methodology used to measure RT estimates, absolute mean values have to be interpreted with caution as they necessitate to transform RT representations to cursor positions on the scale, and we cannot know exactly how participants use this scale (see Bryce and Bratzke, 2015).

Consequently, one more relevant way to assess RT estimation abilities is to investigate whether RT and iRT vary in a similar manner, independently of their absolute

values. This was verified, as participants showed positive correlations between individual RTs and iRTs. Although there was interindividual variability, this means that humans globally have a good temporal representation of their RTs. However, RT can have several sources of variability (Perquin et al., 2024). First, experimental manipulations are most of the time employed to induce systematic changes in RT. Through all the experiments presented in the thesis, the changes in RT due to experimental conditions were also observed in iRT. Second, RTs exhibit rapid trial-to-trial fluctuations, that were rather well captured by iRT as evidenced by the positive correlations. Finally, RTs also display slower temporal trends over experimental blocks and sessions. In our data, RTs indeed had a tendency to decrease through the duration of the experiment, but these slow trends were not found in iRT (see Appendix C.1).

These findings suggest that while participants can accurately monitor their RTs on a trial-by-trial basis, they lack a precise global representation of their average RT. This implies that iRT judgments are likely relative rather than absolute. Supporting this idea, Rouault et al. (2019) investigated global confidence in perceptual discrimination tasks and found that participants generally underestimated their overall performance. However, when provided with external feedback, their ability to monitor global performance improved. Similar findings have been observed in the temporal domain. For instance, Riemer et al. (2019) reported that participants made more accurate judgments of their mean temporal errors in reproduction tasks when they received signed feedback after each trial. Conversely, Öztel and Balci (2023) found that although participants could accurately monitor their errors on a trial-by-trial basis, they struggled to report their average temporal bias over time. Taken together, these results suggest that while individuals can make precise judgments about their performance on a trial-to-trial basis, they often lack an accurate global representation of their performance, unless they receive consistent feedback. Although we did not investigate comparable global estimates of RT, it would certainly enable to learn more about the processes underlying RT estimation.

Overall, understanding which sources of variability in RT and iRT are common, and which are different, would provide significant insights about the underlying estimation process.

6.1.2. iRT as a composite introspective measure

6.1.2.1. iRT as both temporal read-out and inferences from other cues

Chapter 2 evidenced that RT estimation relies on a direct access to the temporal interval, but is also biased by different, non-temporal, cues. With the use of mixed models, we were able to dissociate the two.

First, even when taking into account different factors, iRT and RT were still strongly correlated, suggesting a direct access to RT duration. In addition, participants ability to estimate their RT was not completely independent of their ability to estimate the duration of an external stimulus.

On another side, non-temporal factors, such as perceptual difficulty, force thresh-

old, response correctness, and multiple subthreshold EMG activities, influenced RT estimation beyond their effects on actual RT (for other examples see Bratzke and Bryce, 2019; Bryce and Bratzke, 2014). These factors, that we classified into metacognitive knowledge (*e.g.*, prior understanding of task conditions) and metacognitive experiences (*e.g.*, difficulty, effort, or hesitation), suggest that participants rely on heuristic cues for RT judgments. This aligns with Koriat's cue utilization model (Koriat, 1997), proposing that people rely on various internal and external cues when making metacognitive judgments, and that their effectiveness depends on how well these cues actually correlate with the correct outcome or performance. In our results, effects of non-temporal factors were most of the time similar on RT and iRT, meaning that they reflected valid cues, but had an effect on iRT beyond their effect on RT, meaning that they biased the estimates. In terms of resources allocation, it appears efficient to base RT judgments on other cues than temporal one, more salient and accessible (M. D. Klein & Stolz, 2018).

The variations in biases in different task contexts suggest that the use of these different cues depends on their relevance and accessibility during the ongoing task. In addition, the number and nature of biasing factors also varied across individuals, indicating that the use of non-temporal cues may be linked to individual differences in task performance and timing abilities (see Appendix C.2). This implies that patients with disruption of timing performances could still be able to estimate their own RTs based on external evidence.

Several other metacognitive experiences, such as the feeling of difficulty (Desender et al., 2017) or the experience of urge-to-err (Questienne, Van Opstal, et al., 2018) have been shown to emerge from a combination of multiple cues, which can include RT, response conflict and motor hesitation. Following this, it is possible that processing speed is part of a global and multidimensional metacognitive experience that arises during the performance of the task, and that can be used differently depending on the required metacognitive judgment. Assessing the contribution of these different features to different judgments gives a further insight on how metacognitive representations are built. Furthermore, deciphering the conditions and context in which these cues are used would be really informative.

6.1.2.2. An introspective mental chronometry

Chapter 2 also showed that different processes underlying RT could be accessible for its metacognitive judgment. Indeed, RT is the measure of the duration of a sum of processes involved in giving a response to a stimulus, including the identification of the stimulus, the decision and the motor execution of the response. By definition, RT variability hence depends on the variabilities of each of these processes.

The manipulations of perceptual diffculty and response force of chapter 2, allowed to experimentally induce variability in decision and motor execution times, respectively. We showed that iRT was sensitive to both factors, although such externally-driven variability could lead to some biases in the RT estimates (see previous section 6.1.2.1). In addition, the EMG measures of PMT and MT allowed to show that the combination

of both subdurations best predicted iRT, and that their weights in this prediction differed. It means that, in the same way RT is a sum of processes, iRT is most likely the sum of the representations of these processes.

While we can be rather confident that motor execution time is a subcomponent well captured by MT, it is clear that PMT remains a mix of different subprocesses, including (at least) perception and decision times. Decision processes and factors affecting it constitute a large body of litterature, but the processes happening before have been mostly neglected in the field of decision-making. We can be certain that participants do not take into account the time of detection of the stimulus in their RT estimate, because they need to know that the stimulus has been presented to start their estimation. However, whether they can have access to some identification process remains to be investigated (Weindel, Gajdos, & Burle, 2021). One main issue is that there is no objective measure of this perceptual processing time. One solution is to use experimental manipulations, for example with variations of chroma, which have been shown to affect non-decision time in a DDM (Servant et al., 2016). This demonstrates an important limitation: without objective measures of the mental processes, it is difficult to know whether they can be directly accessed for the subjective judgment. While experimental manipulation can reveal that a process is not considered, if it has no effect on iRT, the opposite is not guaranteed as manipulations affecting iRT might be doing so through biasing inference rather than temporal read-out (see section 6.1.2.1). Further investigations with different protocols manipulating different factors should be developped to go further on the different processes that can be accessed by the participants.

In the present thesis, we evidenced that at least decision and motor execution processes can be introspectively accessed. Metacognition about decision processes have mostly been studied with confidence judgments, and some authors suggested that decision time could inform certainty (Kiani et al., 2014). Additionnaly, many studies have highlighted an important role of motor activities for metacognition. Metacognition performance is indeed better for decisions comitted with a keypress than simply observed (Pereira et al., 2020). Confidence also increases in the presence of subthreshold motor activity before responses (Gajdos et al., 2019), and is affected by transcranial magnetic stimulation of the premotor cortex involved in the response (Fleming et al., 2015). This implies that perceptual metacognition is not only based on the quality of sensory processing, but integrates information across the perception-action cycle.

While both decision-making and motor processes undoubtedly contribute to certain forms of metacognition, we here demonstrate that their duration variability can be assessed on a trial-by-trial basis. Overall, chapter 2 highlighted the core finding that iRT = iPMT + iMT + biases (see Fig.6.1). This raises the question of how these subcomponents can be accessed, which will be discussed in following sections.

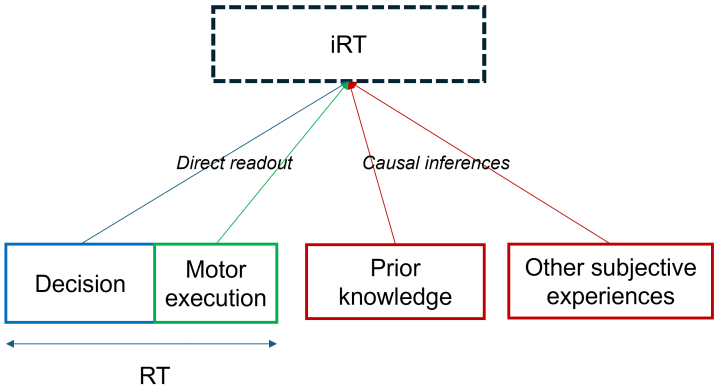


Figure 6.1. *Different cues contributing to iRT judgment.*

6.1.3. RT estimation is embedded within a broader performance monitoring process

Throughout this thesis, one central question was whether RT estimation could be rooted in a broader performance monitoring system. In chapter 3, we demonstrated that mPFC activity, measured through the Ne, reflects a performance monitoring system much more generic than previously thought. More specifically, we highlighted that the Ne could be sensitive to different action dimensions, including RT. With this first result, we expanded the area of competence of a system that has mostly been associated with error monitoring (Ullsperger et al., 2014).

We showed that the relationship between Ne amplitude and RT had a U-shape, with higher amplitudes for both fast and slow RTs, suggesting that it does not track absolute RT value but rather computes its deviation from an internal reference (see below, section 6.2.1). Crucially, this sensitivity to RT was specific to situations in which participants were required to make an iRT judgment, which can be taken as an evidence that is has a role in such metacognitive judgment.

In a second set of results, we failed to find a link between the Ne and iRT, and showed that its amplitude was sensitive only to PMT, with a similar U-shape than with RT. Given the complexity of the iRT judgment, including PMT, MT and other non-temporal cues, we proposed that the Ne is one activity contributing to iRT among others, probably supporting an introspective PMT. On another hand, we also found that the latency of this component increased with increasing MT. Hence, Ne amplitude could support iPMT, although partially due to the quadratic relationship, and its latency could inform iMT. MT duration was also related to differences of activity in the motor cortex contralateral to the response, possibly informing the metacognitive system.

6.2. An integrative mechanism for RT estimation

Building on concepts and mechanisms put forward by existing models of performance monitoring on one hand, and temporal metacognition on the other, we will now try to integrate elements from both into a common framework to propose a potential mechanism for RT estimation.

6.2.1. Performance monitoring as a prediction error computation

The great majority of performance monitoring models is in fact accounting for error monitoring only (see section 1.3.1 of General Introduction). However, in the present thesis, we exclusively focused on the evaluation of correct trials. One of the only models accounting for the presence of a Ne in correct trials is the Predicted Response Outcome (PRO) model.

Following preceding ones (Holroyd & Coles, 2002), this model relies on reinforcement learning principles, but contrary to the previous ones, it does not assume that the Ne is larger on errors because the outcome is worst than expected (that is, a negative reward), but simply because the outcome is unexpected (irrespecive of reward, Alexander and Brown, 2011). In this model, mPFC activity is assumed to code the probabilities of the possible action outcomes, greater probabilities being associated with greater prediction activity. When an expected outcome occurs, the prediction activity is inhibited. Hence, larger activities are observed when expected outcomes fail to occur (negative surprise). While such a "surprise" certainly also applies to reward, it is much more general according to the PRO model, and, although the original model still focuses on reward prediction and outcomes concerning choice accuracy, we propose that its idea could be extended to other dimensions. Indeed, extensions of the model predict a larger activity when predicted outcome occurs at an unexpected time (J. W. Brown, 2013).

From an empirical point of view, results of chapter 3, suggest that a form of prediction error can be computed on different dimensions depending on the context of the task. In the case of RT monitoring, we evidenced that the link between Ne amplitude and RT was due to a high sensitivity to PMT, following a U-shape. This observation can be in accordance with the PRO model, provided that the system computes the deviation from a predicted PMT. In this case, higher Ne amplitudes for PMT values further from the predicted one would be the reflection of larger prediction errors (see Fig. 6.2). The PMT value for which the Ne amplitude is minimal should hence correspond to the predicted PMT. By looking at the U-shaped pattern, this predicted value seems close to the maximal density of the distribution, which makes a lot of sense from a probabilistic point of view. It is, however, also possible that it represents an "optimal PMT", and thus probably an optimal RT: not too fast, to avoid error risk, nor too slow, to be in accordance with the instruction requiring to be as fast as possible. It is likely that such internal prediction is aquired through learning, during the course of the experiment, or during training blocks.

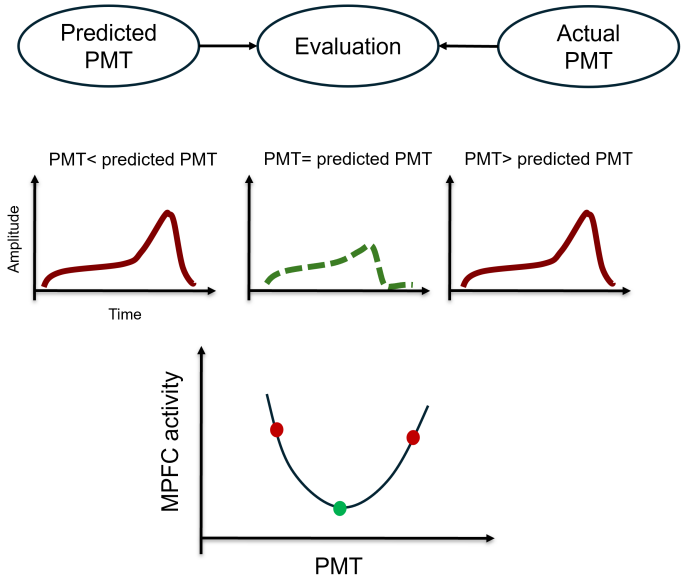


Figure 6.2. *We propose that MPFC activity reflects the comparison between a predicted PMT and the actual one. Highest prediction errors for PMTs that deviate from the predicted value lead to larger MPFC activities.*

Overall, we propose that the mPFC activity reflects a general performance monitoring function, whose role is to compute deviations from intended goal, and that this goal is depending on the context of the ongoing task. In a large number of cases, the most important aspect to monitor is the accuracy of response selection, leading to the well-known error monitoring activity of the region. However, we showed that the competence of the monitoring system was much larger and could extend to the monitoring of other dimensions such as time and response execution. To compute prediction error, the system needs to receive both predicted and actual values. We suggest that the relevant dimension that has to be monitored determines on what input the prediction error will be computed, and thus which systems will transmit information to the performance monitoring system. We will here focus on the timing aspect. In the case of RT, the predicted value most likely comes from the timing system, and the actual value to wich it is compared arises from the decision process (see below section 6.2.3)[1]. In this context, SMA appears as a pivotal structure, as it is involved in all three funtions, namely decision, timing and action monitoring. It would indeed appear efficient that a system involved in programming different aspects of motor outputs would also be engaged in monitoring them.

Whether it is directly involved in the subsequent metacognitive judgment or serve as a source of information integrated later in time to give rise to a metacognitive representation remains to be explored. While the mechanism underlying produced

[1]Note that, in a case of force evaluation, the system would instead receive information about motor execution, possibly also present in pre-motor areas and part of an efference copy.

force monitoring is beyond the scope of this thesis, we will propose a mechanism for RT monitoring, based on elements from temporal error monitoring (see below).

6.2.2. Temporal metacognition as a two-stage accumulation process

Most timing models rely on the idea of an accumulation of pulses emitted by a clock until it reaches a threshold corresponding to the interval to time. Final number of pulses accumulated at the end of the interval is then compared to a duration stored in reference memory (Gibbon et al., 1984). More recently, some authors have proposed a decision model of timing (Balcı & Simen, 2016) by redefining the processing dynamics of the pacemaker-accumulator in the framework of a drift diffusion process, well-established in decision-making. To account for time-scale invariance (*i.e.* scalar property of timing, a strong form of Weber's law), the drift-diffusion process that constitutes temporal accumulation results from the running sum of two opponent Poisson process outputs (TopDDM, Simen et al., 2011).

Interestingly for the object of the current thesis, one model, proposed by Akdoğan and Balcı (2017), used similar mechanisms to explain temporal error monitoring (TEM), that is the ability to judge the direction and magnitude of errors in temporal reproduction tasks. It builds on the idea of two clocks running in parallel, one (motor) leading to the production, and the other (sensory) to the estimation. These two clocks accumulate temporal evidence until a first threshold, conditioned by the target duration, is reached. The first clock that reaches this threshold determines the direction of the error: if the sensory clock reaches the threshold before the motor clock triggering the behavioral response, the produced duration will be judged as too long, and vice versa. The distance between the two thresholds hitting times determines the magnitude of the error, where larger distances depict larger error magnitudes. According to the model, this information is accessed through a second diffusion process, starting once the two clocks have reached their threshold (see Fig.1.5 in General Introduction). The unsigned error magnitude, higher for reproduced times off-target whether shorter or longer, predicts a V-shape pattern in confidence judgments about reproduction, with a decrease in confidence for larger error magnitude. In the following, we will propose a way to adapt those principles to RT subjective evaluation.

6.2.3. An integrative model for decision time estimation

Building on elements of these models of performance monitoring, on one hand, and temporal metacognition, on the other hand, we propose an integrative mechanism for RT estimation. Contrary to TEM, the produced duration of interest, i.e. RT, does not depend on a timing mechanism, but rather on the duration during which evidence is accumulated until a decision threshold is crossed. Thus, instead of relying on two separate clocks, we propose that RT estimation is based on a single clock (such as a TopDDM) running in parallel with a classic drift-diffusion process. Note that,

such mechanism would explain PMT estimation, as MT only starts after the decision threshold is reached.

During PMT, two accumulation processes -— one for decision and one for temporal estimation -— run simultaneously towards their respective thresholds. A key question is how the clock threshold is determined, given that RTs do not revolve around a fixed temporal target as in temporal reproduction tasks. For this process to function, a reference RT value must exist to halt temporal accumulation. This idea aligns with our earlier suggestion (see section 6.2.1) that PMT estimation involves a comparison between predicted and actual PMT values. In this context, the predicted PMT value sets the clock threshold, and participants estimate their PMT by assessing deviations from this predicted value, supporting the notion that iRT judgments are relative (see section 6.1.1).

As for the TEM model, the sign of the deviation is informed by which accumulation process (decision or temporal) reaches its threshold first. If the decision threshold is reached first, PMT will be perceived as faster than the predicted value, and vice versa. The magnitude of this deviation is reflected in the distance between the two thresholds hitting times. To inform this magnitude, we integrate a second diffusion process, starting at the moment of the first threshold hitting time and ended by the second. The sign of the drift is conditioned by which process has reached its threshold first. The level of accumulation of this second-order process directly informs about the deviation from the predicted PMT (see Fig.6.3). We propose that the performance monitoring system compares the outcomes of these two accumulation processes – one representing the predicted PMT (temporal accumulation) and the other representing the actual PMT (DDM). The Ne amplitude reflects the unsigned difference between the two, explaining its characteristic U-shape.

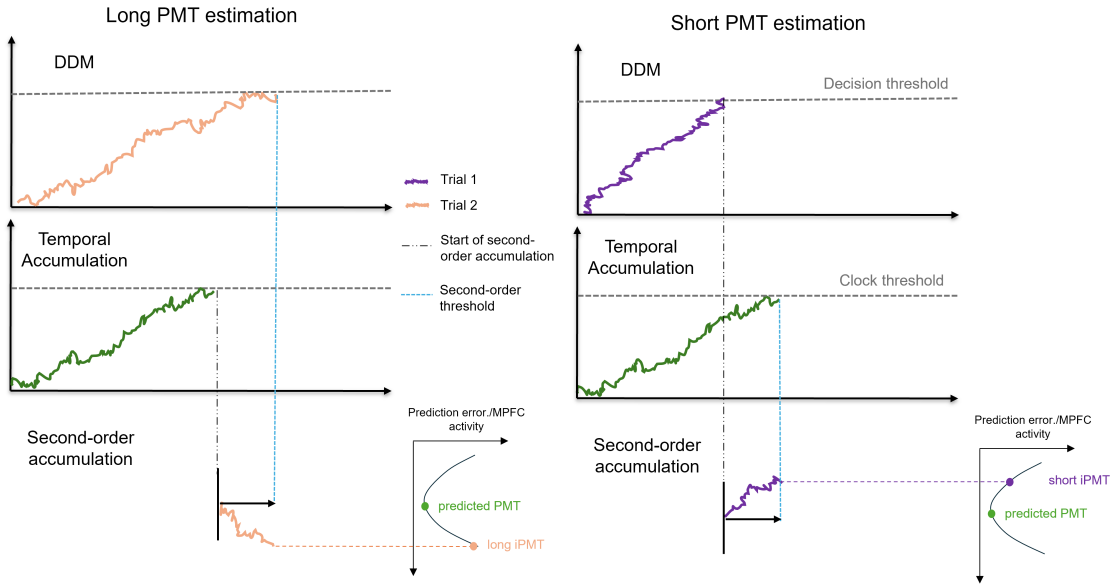


Figure 6.3. *Schematic representation of the proposed mechanism for PMT estimation for two trials. A temporal accumulation runs in parallel with the decision process to inform about PMT deviations from a predicted value. First accumulation process to reach its threshold triggers a second-stage accumulation informing about the size of the deviation. MPFC activity reflects the unsigned deviation from predicted PMT.*

However, this putative model implies that in order to access the size of the deviation, both temporal and decisional thresholds must be reached. The Ne developping shortly after the decision, this would mean that it cannot access the size of the deviation when PMT is shorter than the predicted one, as the clock would reach its threshold only after Ne onset. In such case, a longer Ne latency might be expected for short PMTs. We can also speculate that when the PMT is shorter than predicted, the performance monitoring system may only access the direction of the deviation, not its magnitude. This could be consistent with the observed asymmetry in the U-shape relation between Ne amplitude and PMT: the amplitude is higher for short PMTs than for long ones, and the decline in amplitude from short to medium PMTs is steeper than the increase from medium to long PMTs. Short PMTs, possibly linked to 'risky' responses, might be coded in an all-or-none manner, with a maximum prediction error emitted when a decision is made earlier than expected. Conversely, for longer PMTs, knowing the exact deviation from the prediction might be necessary to adjust behavior accordingly.

The quadratic shape of Ne amplitude in relation to PMT might also relate to confidence. In the TEM model, the dual-stage process informs both error direction and confidence judgments about temporal reproduction, with confidence correlating with error magnitude. Interestingly, predicted confidence patterns also follow a U-shape, with lower confidence for larger temporal errors. In iRT judgments, this U-shaped pattern could reflect two levels of confidence: deviations from an optimal predicted RT might lower confidence in the first-order response, while larger deviations might

increase confidence in the second-order judgment, as they are easier to discriminate. A protocol involving confidence judgments on both first-order performance and second-order estimation could help clarify this.

Finally, such mechanism could also explain some biases coming from non-temporal cues (see section 6.1.2.1). First, different predicted values – thus different clock threshold – could be used for different conditions: it could be that in the strong force blocks of experiment 2 of chapter 2.2, participants anticipate longer RT/PMT, and shift their predicted value. As a consequence, RT/PMT are more often perceived as longer than they should be. Note that this would lead to a clear prediction concerning the Ne: given that the clock threshold represents the predicted value, different thresholds should be reflected in differences in values for which Ne has a minimal amplitude. Second, some conditions could also modify the rate of accumulation, hence leading to over or underestimation.

To further support the proposed integrative model, it is important to consider how the different processes involved might be reflected in neural activity. For example, both first-stage accumulation processes can be indexed by different activities. Neurons representing the accumulation of sensory evidence have been found in the lateral intraparietal (LIP) cortex of monkeys asked to report the perceived direction of random dots' motion by a saccadic movement (Gold & Shadlen, 2007). During decision, these neurons show an increase in firing rate proportional to the amount of evidence, determined by motion strength, until they reach a common bound firing rate. Same neurons have also been shown to encode choice certainty (Kiani & Shadlen, 2009). Signals appearing to track the evolution of the accumulated evidence have also been observed in humans, with an EEG centro-parietal positivity emerging during choice (R. O'Connell et al., 2012). In 2AFC tasks, this signal shows a gradual ramping over time, with a slope correlated to evidence strength and RT (and thus PMT), and a constant peak amplitude putatively reflecting the end of accumulation with boundary crossing, with a peak latency following RT.

On the temporal side, the SMA may serve as a substrate for temporal accumulation, as indicated by the contingent negative variation (CNV), an EEG potential which amplitude has been shown to follow the estimated durations (Casini & Vidal, 2011; Macar et al., 1999). Such ramping activities developing during the interval to be estimated (RT) should be explored in the future, to assess their compatibility with the presented model.

While it remains hypothetical, the proposed model offers perspective to design new experiments aimed at clarifying RT estimation process.

6.3. What about motor execution time?

The previous sections highlighted a putative mechanism for a metacognitive access to PMT, with an implication of a generic performance monitoring system. However, as evidenced in section 6.1.2.2, MT is also an important contributor for iRT. This sensitivity to motor execution duration opens interesting conceptual questions. Mo-

tor execution, at least the part contained in MT, is a peripheral activity. How can it be mentally read out, especially since it does not induce any (visible) movement? Could proprioceptive or tactile reafferences be useable to estimate MT duration? It is unlikely as these signals are known to be gated during self-generated movements (Abbruzzese et al., 1981; Blakemore et al., 1999). Internal activities related to efference copy (Wolpert & Kawato, 1998), or perception of preparatory events preceding the motor command (Haggard et al., 1999) could constitute an index of this motor execution.

Following this idea, we investigated the activity of the primary motor cortex contralateral to the response, known to be related to the strength of the motor command (Vidal, Grapperon, et al., 2003). We showed that the activation of M1 preceding EMG onset was higher for shorter MTs, reflecting a more efficient motor command. In addition to leading to the effective motor command, we can hypothesized that such signal also forms an efference copy, sent to other structures to control the ongoing action (Wolpert & Kawato, 1998). In case of a metacognitive judgment, this signal could also be integrated to form a representation of MT.

A previous study investigating partial error awareness (Ficarella et al., 2019), found both larger M1 activity and larger Ne after aware partial errors compared to unaware ones. The authors proposed that the stronger motor command reflected in M1 activity led to a stronger efference copy sent to SMA, finally leading to a larger Ne. In the present case, the larger M1 activity was not reflected in a larger Ne for shorter MT, but could have affected its latency, potentially providing information about MT for the subsequent iRT judgment.

Interestingly, at the first order level, some evidence suggests that PMT and MT are not (or at least not always) that independent. Indeed, some manipulations of perceptual difficulty, such as color saturation (Servant et al., 2016) or contrast (Weindel, Anders, et al., 2021) affected MT as well as PMT. Similarly, in the present studies, we also observed effects of stimulus angle on MT. This observation, among others, have led some authors to propose an integrated diffusion model of deciding and acting (Servant et al., 2021), in which PMT and MT both rely on a similar decision variable, and two different thresholds mark the end of PMT and the end of MT/RT. If this is the case, it would appear optimal to have a unique clock for RT estimation, timing this process from its begining to the second threshold crossing. However, the results of chapter 2, not only evidence that the whole RT interval, including MT, can be estimated, but also that PMT and MT have different weights in this estimation. This suggests that PMT and MT are accessed through different processes, which seems, at first sight, incompatible with this integrated model. While this needs to be further addressed, it also points to the fact that investigating metacognitive monitoring of RT can also help in the understanding of the organization of first-order mental processes. In line with this, the very first study using a form of iRT measure with categories (Sanford, 1970), has investigated the effect of stimulus intensity in an simple auditory RT task and observed that it affected RT and its subjective judgment. The author concluded that this intensity effect on RT was not mediated by an effect on stimulus detection time. In such case, the characteristics of the second-order measure was

used to infer properties of the first-oder processing chain. Methodologically speaking, this implies that a deeper understanding of cognitive processes is essential to gain insights into metacognitive processes. However, the reverse might also hold true – improving our comprehension of metacognitive processes could, in turn, enhance our understanding of the underlying cognitive mechanisms. This interdependence between cognition and metacognition can also be observed in the way metacognitive processing can influence first-order performance, which is the object of the following section.

6.4. Knowing how long it took: what for?

So far, we focused on pure monitoring aspects of iRT, that is how second-order processes can access first-order ones. However, we showed in chapter 4 that second-order processing also has an impact on first-order performance, indicating a significant interplay between both levels.

First, we observed that metacognitive instructions, whether focusing on RT or produced force, influenced overall task performance, in terms of mean and variability. Specifically, when participants had to evaluate their RT, they exhibited faster responses on average, with a slightly enhanced variability. In parallel, we observed similar effects on the group who had to evaluate its produced force, with an increase in mean force production and in its variability. The effects on mean performance can be explained by a heightened attentional focus on these respective dimensions, that may have reinforced the first-order instruction, which is, in the case of RT, to be as fast as possible. Note that, to respect this instruction, one still needs to have some form of RT monitoring, to know when it is off-target. This implies that in RT tasks, even without explicit RT judgment, there must be a "default" RT monitoring mechanism. The raster plot analyses realized on the Force-group in chapter 3.2, seeking the relationship between Ne amplitude and RT values, indeed revealed a sensitivity to RT, although much weaker than in the RT-group. This result supports the hypothesis of a default RT monitoring, which becomes stronger and more efficient when an explicit judgment is required.

Alternatively, the (trend toward) increased variability in the dimension that is being evaluated, could reflect a process of optimization of the second-order performance. By controlling and varying their RT, participants can facilitate their discriminability, and enhance the accuracy of their subjective estimates, and the same holds for Force. In this type of metacognitive task, explicitly requesting a report about performance, participants are asked to be as accurate as possible in this report. This differs from judgments of confidence or feeling of difficulty that rely on subjective experiences that cannot be accurate or not, contrary to temporal estimations. Consequently, the second-order performance likely become a part of the task goal. This means that to optimize the global performance in the task, one has to find a balance between first and second order performances (Rosenbaum et al., 2022), that can lead to differences in the way the first order task is processed. This also highlights the importance of

developping indirect and implicit metacognitive measures, which could be added to a cognitive task, without adding a supplementary goal that can modify cognitive processing.

On a more local scale, participants also appeared to strategically adjust their behavior based on their metacognitive judgments, suggesting that they play a crucial role in real-time cognitive control. For instance, participants sped up after slower responses and tended to slow down after faster ones, indicating an active regulation of performance, probably to optimize task outcomes. Indeed, such a negative relationship between iRT judgments and subsequent RT is compatible with the idea that participants avoid both very fast and very slow response times. The instruction being to respond as fast as possible, it is clear that slow RTs are unoptimals. On the other hand, while fast RTs are in accordance with the instruction, they also constitue a higher risk of error, with responses made before processing enough information about the stimulus. Hence, an optimal RT value probably lies in between. Interestingly, this optimal value could constitute the predicted value used by the monitoring system to compute RT prediction errors. In such case, this prediction error could be used to signal the need for control, and to adapt subsequent behavior accordingly.

6.5. The measurement problem

If one seeks to understand how a subjective representation of RT arises, one needs to know what sources of information contribute to it, and to have a good measure of it. In this thesis, iRT was used to measure the subjective representation of RT, and we highlighted that it was a composite measure. First, just as RT is the sum of different processes duration, we showed that iRT most probably reflects a combination of the subjective representations of decision and motor execution times. Second, iRT also reflected inferences from non temporal cues. Hence, to better understand the processes involved in RT estimation, we need to disentangle iRT components and find good measures of each of them. While the methodology of chapter 2 allowed to dissociate and quantify the overall biases, finding a way to assess them on a trial-by-trial basis would allow more precise individual measures. On the other hand, even though we could dissociate PMT and MT with EMG, we did not have a direct measure of their subjective equivalents. To go further, it could be possible to ask for separate PMT and MT reports, although multiple introspective reports can be complex to implement (see, however, J. Miller et al., 2010). Beyond iRT complexity due to its multiple sources of information, it remains an indirect measure, objectively reflecting a spatial position on a visual scale. We do not know to what extent having to transform our representation of RT into a visuo-spatial framework could distort it. Furthermore, this report is made retrospectively, and we cannot know when the representation arises and if (and how) it changes over time.

On the other side, EEG recordings are also indirect and noisy measures of brain activity. In fact, our ERP analyses relied on an averaging of data of several trials depending on RT duration. As RT is a continuous measure, it required to make

categories of trials having different RTs. This categorization raises an issue: if we choose to make categories with similar number of trials, as we did, RT variability differs between them; but if we choose categories based on equal ranges of RT, they will be composed of a significantly different numbers of trials and thus different levels of noise after the averaging process. In addition, categorizing a continuous variable results in a significant loss of information. While we tried to overcome this by performing ERP-images analyses in chapter 3, to approximate individual trial measures, it erased the inter-individual variability. One important perspective would be to extract our component of interest on a trial-by-trial basis, allowing to have a direct mapping between individual values of RT/iRT. We attempted to extract these amplitudes for individual trials (see Appendix B.3), and the results nicely fit the one obtained on averages. However, the very low signal-to-noise ratio and the impossibility of verifying the accuracy of our measurements made us doubt about the reliability of the obtained results. More advanced methods have been developped to reach this goal, such as denoising (Benchabane et al., 2019) or direct statistical modeling of the data (Ehinger & Dimigen, 2019; Visalli et al., 2024). This would offer the perspective of using mixed models with electrophysiological data, to precise the nature of the relations between Ne, RT and iRT. We could test more precisely if Ne amplitude can predict iRT, and explore if the addition of its latency, or of M1 activation, would improve the prediction.

Ultimately, the advancements in our comprehension of subjective representation of RT depends on the reliability of our measures, and on the relationship between them. Indeed, as neuroscientists, our goal is to find brain states and neuronal computations that lead to the subjective representation of interest, herein subjective RT. However, as experimenters that are external to the studied brains and minds, we dot not have a direct access to these elements. We instead rely on indirect measures on which we make inferences. On one side, we rely on iRT, which is only a behavioral indication of the underlying subjective representation. On the other side, we only have measures of the brain activity, EEG signal in the present case. In addition, iRT is a subjective judgment supposedly reflecting the objective measure that is RT. Hence, while our theoretical goal is to assess the correlation between a subjective representation and a brain state, what we can really measure is a correlation between a subjective report and a brain activity (see Overgaard, 2006). The desired correlation is fully dependant on the relationships between the subjective representation and its report and its objective equivalent, on one side, and the brain state and its measured activity, on the other (see Fig.6.4). Beyond correlations, one aim is also to assess causal relationships between subjective representation and brain states. One way of doing that is to suppress one and observed the effect on the other. While the observation of chapter 3 that without explicit iRT judgment Ne shows no sensitivity to RT, a short term perspective is to do the opposite: inhibit the Ne to see if it disrupts iRT judgments (see section 5.1).

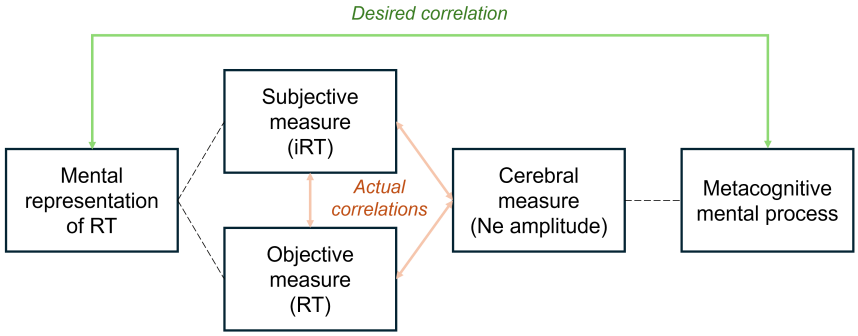


Figure 6.4. *Relations between the studied processes and the different measured used.*

The study of metacognitive representations also raises an important paradox. From the experimental point of view, instruction requires iRT to be an accurate temporal estimation of RT. Behaviorally, we hence expect an significant positive correlation between the two measures, which is what is observed. However, if we want to find brain signatures that are specific to the subjective judgments, we need to be able to dissociate them. In the present thesis, we dissociated iRT and RT to some extent, by showing that iRT is a composite measure not only reflecting the processes underlying RT but also other non-temporal cues, probably arising from the integration of multiple cerebral activities. We also observed that a percentage of participants did not have a good representation of their RT, indexed by an absence of correlation between RT and iRT. It would propably be insightful to compare brain activities of different groups of participants, depending on their ability to accurately estimate their RT (see Appendix B.2). On the same line, all participants were not biased by the same external cues, which may provide a clue as to how these biases are implemented in the brain. Ideally, it would also be interesting to dissociate different types of trials within the same participant, some trials being more accurate than others. For that, we would need to have some clear criteria of what constitutes a good estimation. Furthermore, while we often consider that the relationship between objective and subjective measures is due to the addition of a metacognitive process simply reading out first-order processes, chapter 4 suggests that the presence of subjective judgment can also modify the first-order process itself. This bidirectional relationship complicates our understanding and highlights the need for methods that can disentangle these interactions.

Overall, while we cannot consider all these points at the same time, clarifying the links between objective, subjective and cerebral measures bring us closer to the mechanism underlying subjective representation of response times.

6.6. Conclusion

The results of the present thesis allow to advance our understanding of temporal metacognition by addressing the complex nature of subjective experience in the temporal domain. We demonstrate that subjective representation of RT is both a direct read-out of RT subdurations and a reconstruction from inferences based on other cues. We could dissociated these different components, although there is potential to refine this understanding even further. We bring evidence that RT estimation relies on a generic performance monitoring system, but that this link is not as straightforward as previously thought. These findings led us to propose an integrative mechanism, based on both temporal error monitoring and predicted response outcome models, that needs to be more clearly established. This work provides key pieces to the puzzle that is the understanding of subjective representation of self-produced time, by dissecting the building blocks, at both behavioral and electrophysiological levels, of iRT and quantifying their contributions. By bridging performance monitoring and timing mechanisms, this thesis paves the way for a deeper exploration of how subjective temporal judgments are formed and how they intersect with broader cognitive mechanisms.

Bibliography

Abbruzzese, G., Ratto, S., Favale, E., & Abbruzzese, M. (1981). Proprioceptive modulation of somatosensory evoked potentials during active or passive finger movements in man. *Journal of Neurology, Neurosurgery, and Psychiatry*, *44*(10), 942–949 (cit. on p. 168).

Akdoğan, B., & Balcı, F. (2016). Stimulus probability effects on temporal bisection performance of mice (Mus musculus). *Animal Cognition*, *19*(1), 15–30. https://doi.org/10.1007/s10071-015-0909-6 (cit. on p. 28).

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 (cit. on pp. 28, 29, 164).

Alain, C., McNeely, H. E., He, Y., Christensen, B. K., & West, R. (2002). Neurophysiological Evidence of Error-monitoring Deficits in Patients with Schizophrenia. *Cerebral Cortex*, *12*(8), 840–846. https://doi.org/10.1093/cercor/12.8.840 (cit. on p. 43).

Alexander, W. H., & Brown, J. W. (2011). Medial prefrontal cortex as an action-outcome predictor. *Nature neuroscience*, *14*(10), 1338–1344. https://doi.org/10.1038/nn.2921 (cit. on pp. 40, 162).

Alexander, W. H., & Brown, J. W. (2014). A general role for medial prefrontal cortex in event prediction. *Frontiers in Computational Neuroscience*, *8*. https://doi.org/10.3389/fncom.2014.00069 (cit. on p. 40).

Allain, S., Burle, B., Hasbroucq, T., & Vidal, F. (2009). Sequential adjustments before and after partial errors. *Psychonomic Bulletin & Review*, *16*(2), 356–362. https://doi.org/10.3758/PBR.16.2.356 (cit. on p. 23).

Allain, S., Carbonnell, L., Falkenstein, M., Burle, B., & Vidal, F. (2004). The modulation of the Ne-like wave on correct responses foreshadows errors. *Neuroscience Letters*, *372*(1-2), 161–166. https://doi.org/10.1016/j.neulet.2004.09.036 (cit. on p. 43).

Allan, L. G. (1975). Temporal order psychometric functions based on confidence-rating data. *Perception & Psychophysics*, *18*(5), 369–372. https://doi.org/10.3758/BF03211214 (cit. on p. 27).

Amiez, C., Joseph, J.-P., & Procyk, E. (2005). Anterior cingulate error-related activity is modulated by predicted reward. *The European Journal of Neuroscience*, *21*(12), 3447–3452. https://doi.org/10.1111/j.1460-9568.2005.04170.x (cit. on pp. 40, 46).

Amiez, C., Neveu, R., Warrot, D., Petrides, M., Knoblauch, K., & Procyk, E. (2013). The Location of Feedback-Related Activity in the Midcingulate Cortex Is Predicted

by Local Morphology. *The Journal of Neuroscience*, *33*(5), 2217–2228. https://doi.org/10.1523/JNEUROSCI.2779-12.2013 (cit. on p. 39).

Anguera, J. A., Seidler, R. D., & Gehring, W. J. (2009). Changes in Performance Monitoring During Sensorimotor Adaptation. *Journal of Neurophysiology*, *102*(3), 1868–1879. https://doi.org/10.1152/jn.00063.2009 (cit. on p. 46).

Arezzo, J., & Vaughan, H. G. (1980, January). Intracortical Sources and Surface Topography of the Motor Potential and Somatosensory Evoked Potential in the Monkey. In H. H. Kornhubek & L. Deecke (Eds.), *Progress in Brain Research* (pp. 77–83, Vol. 54). Elsevier. https://doi.org/10.1016/S0079-6123(08)61610-6 (cit. on p. 110).

Armbrecht, A.-S., Gibbons, H., & Stahl, J. (2012). Monitoring force errors: Medial-frontal negativity in a unimanual force-production task. *Psychophysiology*, *49*(1), 56–72. https://doi.org/10.1111/j.1469-8986.2011.01282.x (cit. on p. 46).

Armbrecht, A.-S., Gibbons, H., & Stahl, J. (2013). Effects of response force parameters on medial-frontal negativity. *PloS One*, *8*(1), e54681. https://doi.org/10.1371/journal.pone.0054681 (cit. on p. 46).

Audley, R. J. (1960). A stochastic model for individual choice behavior. *Psychological Review*, *67*, 1–15. https://doi.org/10.1037/h0046438 (cit. on p. 24).

Bader, F., & Wiener, M. (2021). Awareness of errors and feedback in human time estimation. *Learning & Memory (Cold Spring Harbor, N.Y.)*, *28*(5), 171–177. https://doi.org/10.1101/lm.053108.120 (cit. on p. 30).

Bader, F., & Wiener, M. (2023). Neuroimaging Signatures of Metacognitive Improvement in Sensorimotor Timing. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, JN–RM–1789–22. https://doi.org/10.1523/JNEUROSCI.1789-22.2023 (cit. on p. 30).

Baird, B., Cieslak, M., Smallwood, J., Grafton, S. T., & Schooler, J. W. (2015). Regional white matter variation associated with domain-specific metacognitive accuracy. *Journal of Cognitive Neuroscience*, *27*(3), 440–452. https://doi.org/10.1162/jocn_a_00741 (cit. on p. 20).

Baird, B., Smallwood, J., Gorgolewski, K. J., & Margulies, D. S. (2013). Medial and lateral networks in anterior prefrontal cortex support metacognitive ability for memory and perception. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *33*(42), 16657–16665. https://doi.org/10.1523/JNEUROSCI.0786-13.2013 (cit. on p. 20).

Balci, F., Freestone, D., & Gallistel, C. R. (2009). Risk assessment in man and mouse. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(7), 2459–2463. https://doi.org/10.1073/pnas.0812709106 (cit. on p. 28).

Balcı, F., Freestone, D., Simen, P., deSouza, L., Cohen, J. D., & Holmes, P. (2011). Optimal Temporal Risk Assessment. *Frontiers in Integrative Neuroscience*, *5*. https://doi.org/10.3389/fnint.2011.00056 (cit. on p. 28).

Balcı, F., & Simen, P. (2014). Decision processes in temporal discrimination. *Acta Psychologica*, *149*, 157–168. https://doi.org/10.1016/j.actpsy.2014.03.005 (cit. on pp. 25, 26).

Balcı, F., & Simen, P. (2016). A decision model of timing. *Current Opinion in Behavioral Sciences*, *8*, 94–101. https://doi.org/10.1016/j.cobeha.2016.02.002 (cit. on pp. 25, 26, 28, 164).

Balsdon, T., Wyart, V., & Mamassian, P. (2020). Confidence controls perceptual evidence accumulation. *Nature Communications*, *11*(1), 1753. https://doi.org/10.1038/s41467-020-15561-w (cit. on p. 123).

Barch, D. M., Braver, T. S., Sabb, F. W., & Noll, D. C. (2000). Anterior Cingulate and the Monitoring of Response Conflict: Evidence from an fMRI Study of Overt Verb Generation. *Journal of Cognitive Neuroscience*, *12*(2), 298–309. https://doi.org/10.1162/089892900562110 (cit. on p. 39).

Barthelmé, S., & Mamassian, P. (2009). Evaluation of Objective Uncertainty in the Visual System. *PLoS computational biology*, *5*, e1000504. https://doi.org/10.1371/journal.pcbi.1000504 (cit. on p. 18).

Bartholow, B. D., Pearson, M. A., Dickter, C. L., Sher, K. J., Fabiani, M., & Gratton, G. (2005). Strategic control and medial frontal negativity: Beyond errors and response conflict. *Psychophysiology*, *42*(1), 33–42. https://doi.org/10.1111/j.1469-8986.2005.00258.x (cit. on p. 111).

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 (cit. on p. 112).

Benchabane, B., Benkherrat, M., Burle, B., Vidal, F., Hasbroucq, T., Djelel, S., & Belmeguenai, A. (2019). Wavelets statistical denoising (WaSDe): individual evoked potential extraction by multi-resolution wavelets decomposition and bootstrap. *IET Signal Processing*, *13*(3), 348–355. https://doi.org/10.1049/iet-spr.2018.5389 (cit. on p. 171).

Bjork, R. A., Dunlosky, J., & Kornell, N. (2013). Self-regulated learning: beliefs, techniques, and illusions. *Annual Review of Psychology*, *64*, 417–444. https://doi.org/10.1146/annurev-psych-113011-143823 (cit. on p. 18).

Blakemore, S.-J., Frith, C. D., & Wolpert, D. M. (1999). Spatio-Temporal Prediction Modulates the Perception of Self-Produced Stimuli. *Journal of Cognitive Neuroscience*, *11*(5), 551–559. https://doi.org/10.1162/089892999563607 (cit. on p. 168).

Bogacz, R., Wagenmakers, E.-J., Forstmann, B. U., & Nieuwenhuis, S. (2010). The neural basis of the speed-accuracy tradeoff. *Trends in Neurosciences*, *33*(1), 10–16. https://doi.org/10.1016/j.tins.2009.09.002 (cit. on p. 24).

Boldt, A., & Yeung, N. (2015). Shared Neural Markers of Decision Confidence and Error Detection. *Journal of Neuroscience*, *35*(8), 3478–3484. https://doi.org/10.1523/JNEUROSCI.0797-14.2015 (cit. on pp. 22, 23, 45).

Boldt, A., & Gilbert, S. J. (2019). Confidence guides spontaneous cognitive offloading. *Cognitive Research: Principles and Implications*, *4*(1), 45. https://doi.org/10.1186/s41235-019-0195-y (cit. on p. 123).

Bonini, F., Burle, B., Liégeois-Chauvel, C., Régis, J., Chauvel, P., & Vidal, F. (2014). Action Monitoring and Medial Frontal Cortex: Leading Role of Supplementary Motor

Area. *Science*, *343*(6173), 888–891. https://doi.org/10.1126/science.1247412 (cit. on pp. 22, 30, 42–44).

Botvinick, M., Nystrom, L. E., Fissell, K., Carter, C. S., & Cohen, J. D. (1999). Conflict monitoring versus selection-for-action in anterior cingulate cortex. *Nature*, *402*(6758), 179–181. https://doi.org/10.1038/46035 (cit. on p. 39).

Botvinick, M. M., Carter, C. S., Braver, T. S., Barch, D. M., & Cohen, J. D. (2001). Conflict Monitoring and Cognitive Control. *Psychological Review*, *108*(3), 624–652 (cit. on pp. 22, 39, 124).

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 (cit. on p. 33).

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 (cit. on p. 35).

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 (cit. on pp. 35, 159).

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 (cit. on p. 35).

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 (cit. on p. 34).

Brázdil, M., Roman, R., Daniel, P., & Rektor, I. (2005). Intracerebral error-related negativity in a simple go/nogo task. *Journal of Psychophysiology*, *19*(4), 244–255. https://doi.org/10.1027/0269-8803.19.4.244 (cit. on pp. 38, 39).

Brázdil, M., Roman, R., Falkenstein, M., Daniel, P., Jurák, P., & Rektor, I. (2002). Error processing – evidence from intracerebral ERP recordings. *Experimental Brain Research*, *146*(4), 460–466. https://doi.org/10.1007/s00221-002-1201-y (cit. on p. 38).

Brocas, I., Carrillo, J. D., & Tarrasó, J. (2018). Self-awareness of biases in time perception. *Journal of Economic Behavior & Organization*, *148*, 1–19. https://doi.org/10.1016/j.jebo.2018.02.001 (cit. on p. 27).

Brown, A. L. (1977, June). *Knowing When, Where, and How to Remember: A Problem of Metacognition. Technical Report No. 47* (tech. rep.). (Cit. on p. 17).

Brown, J. W. (2013). Beyond Conflict Monitoring: Cognitive Control and the Neural Basis of Thinking Before You Act. *Current Directions in Psychological Science*, *22*(3), 179–185. https://doi.org/10.1177/0963721412470685 (cit. on pp. 41, 162).

Brown, J. W., & Braver, T. S. (2005). Learned Predictions of Error Likelihood in the Anterior Cingulate Cortex. *Science*, *307*(5712), 1118–1121. https://doi.org/10.1126/science.1105783 (cit. on p. 40).

Brown, R., & McNeill, D. (1966). The "tip of the tongue" phenomenon. *Journal of Verbal Learning & Verbal Behavior*, *5*(4), 325–337. https://doi.org/10.1016/S0022-5371(66)80040-3 (cit. on p. 17).

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 (cit. on p. 131).

Brown, S. W. (2006). Timing and executive function: Bidirectional interference between concurrent temporal production and randomization tasks. *Memory & Cognition*, *34*(7), 1464–1471. https://doi.org/10.3758/BF03195911 (cit. on p. 35).

Brown, S. W., & Boltz, M. G. (2002). Attentional processes in time perception: Effects of mental workload and event structure. *Journal of Experimental Psychology: Human Perception and Performance*, *28*(3), 600–615. https://doi.org/10.1037/0096-1523.28.3.600 (cit. on p. 25).

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 (cit. on pp. 35, 50, 159).

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 (cit. on pp. 35, 157).

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 (cit. on p. 35).

Buhusi, C. V., & Meck, W. H. (2005). What makes us tick? Functional and neural mechanisms of interval timing. *Nature Reviews Neuroscience*, *6*(10), 755–765. https://doi.org/10.1038/nrn1764 (cit. on p. 25).

Burgio-Murphy, A., Klorman, R., Shaywitz, S. E., Fletcher, J. M., Marchione, K. E., Holahan, J., Stuebing, K. K., Thatcher, J. E., & Shaywitz, B. A. (2007). Error-related event-related potentials in children with attention-deficit hyperactivity disorder, oppositional defiant disorder, reading disorder, and math disorder. *Biological Psychology*, *75*(1), 75–86. https://doi.org/10.1016/j.biopsycho.2006.12.003 (cit. on p. 22).

Burle, B., Bonnet, M., Vidal, F., Possamaï, C.-A., & Hasbroucq, T. (2002). A transcranial magnetic stimulation study of information processing in the motor cortex: relationship between the silent period and the reaction time delay. *Psychophysiology*, *39*(2), 207–217. https://doi.org/10.1017/S0048577202010077 (cit. on p. 110).

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 (cit. on pp. 21, 33).

Burle, B., Roger, C., Allain, S., Vidal, F., & Hasbroucq, T. (2008). Error negativity does not reflect conflict: a reappraisal of conflict monitoring and anterior cingulate cortex activity. *Journal of Cognitive Neuroscience*, *20*(9), 1637–1655. https://doi.org/10.1162/jocn.2008.20110 (cit. on pp. 39, 118).

Burle, B., Vidal, F., Tandonnet, C., & Hasbroucq, T. (2004). Physiological evidence for response inhibition in choice reaction time tasks. *Brain and Cognition*, *56*(2), 153–164. https://doi.org/10.1016/j.bandc.2004.06.004 (cit. on p. 110).

Burns, J. T. (1971). Error-induced inhibition in a serial reaction time task. *Journal of Experimental Psychology*, *90*(1), 141–148. https://doi.org/10.1037/h0031335 (cit. on p. 23).

Call, J., & Carpenter, M. (2001). Do apes and children know what they have seen? *Animal Cognition*, *3*(4), 207–220. https://doi.org/10.1007/s100710100078 (cit. on p. 21).

Carrasco, M. C., Bernal, M. C., & Redolat, R. (2001). Time Estimation and Aging: A Comparison between Young and Elderly Adults. *The International Journal of Aging and Human Development*, *52*(2), 91–101. https://doi.org/10.2190/7NFL-CGCP-G9E1-P0H1 (cit. on p. 25).

Carter, C. S., Braver, T. S., Barch, D. M., Botvinick, M. M., Noll, D., & Cohen, J. D. (1998). Anterior cingulate cortex, error detection, and the online monitoring of performance. *Science (New York, N.Y.)*, *280*(5364), 747–749. https://doi.org/10.1126/science.280.5364.747 (cit. on p. 39).

Carter, C. S., Macdonald, A. M., Botvinick, M., Ross, L. L., Stenger, V. A., Noll, D., & Cohen, J. D. (2000). Parsing executive processes: Strategic vs. evaluative functions of the anterior cingulate cortex. *Proceedings of the National Academy of Sciences*, *97*(4), 1944–1948. https://doi.org/10.1073/pnas.97.4.1944 (cit. on p. 39).

Casini, L., & Vidal, F. (2011). The SMAs: Neural Substrate of the Temporal Accumulator? *Frontiers in Integrative Neuroscience*, *5*. https://doi.org/10.3389/fnint.2011.00035 (cit. on pp. 26, 167).

Çavdaroğlu, B., Zeki, M., & Balcı, F. (2014). Time-based reward maximization. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1637), 20120461. https://doi.org/10.1098/rstb.2012.0461 (cit. on p. 28).

Cella, M., Reeder, C., & Wykes, T. (2015). Lessons learnt? The importance of metacognition and its implications for Cognitive Remediation in schizophrenia. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.01259 (cit. on p. 133).

Chambon, V., & Haggard, P. (2012). Sense of control depends on fluency of action selection, not motor performance. *Cognition*, *125*(3), 441–451. https://doi.org/10.1016/j.cognition.2012.07.011 (cit. on p. 124).

Chomsky, N. (1959). A Review of B. F. Skinner's Verbal Behavior. *Language*, 26–58 (cit. on p. 15).

Chua, E. F., Schacter, D. L., & Sperling, R. A. (2009). Neural correlates of metamemory. *Journal of cognitive neuroscience*, *21*(9), 1751–1765. https://doi.org/10.1162/jocn.2009.21123 (cit. on p. 17).

Cohen, R. L., Sandler, S. P., & Keglevich, L. (1991). The failure of memory monitoring in a free recall task. *Canadian Journal of Psychology / Revue canadienne de psychologie*, *45*(4), 523–538. https://doi.org/10.1037/h0084303 (cit. on p. 17).

Coles, M. G., Scheffers, M. K., & Holroyd, C. B. (2001). Why is there an ERN/Ne on correct trials? Response representations, stimulus-related components, and the theory of error-processing. *Biological Psychology*, *56*(3), 173–189. https://doi.org/10.1016/s0301-0511(01)00076-x (cit. on pp. 22, 37).

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*, 529–553. https://doi.org/10.1037/0096-1523.11.5.529 (cit. on p. 21).

Comte, A. (1830). *Cours de philosophie positive*. (Cit. on p. 14).

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 (cit. on pp. 34, 35, 49, 124, 157).

Coull, J. T., Cheng, R.-K., & Meck, W. H. (2011). Neuroanatomical and Neurochemical Substrates of Timing. *Neuropsychopharmacology*, *36*(1), 3–25. https://doi.org/10.1038/npp.2010.113 (cit. on pp. 26, 47).

Coull, J. T., Hwang, H. J., Leyton, M., & Dagher, A. (2012). Dopamine precursor depletion impairs timing in healthy volunteers by attenuating activity in putamen and supplementary motor area. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *32*(47), 16704–16715. https://doi.org/10.1523/JNEUROSCI.1258-12.2012 (cit. on p. 26).

Coull, J. T., Nazarian, B., & Vidal, F. (2008). Timing, storage, and comparison of stimulus duration engage discrete anatomical components of a perceptual timing network. *Journal of Cognitive Neuroscience*, *20*(12), 2185–2197. https://doi.org/10.1162/jocn.2008.20153 (cit. on p. 26).

Coull, J. T., Vidal, F., Nazarian, B., & Macar, F. (2004). Functional Anatomy of the Attentional Modulation of Time Estimation. *Science*, *303*(5663), 1506–1508. https://doi.org/10.1126/science.1091573 (cit. on p. 25).

Dali, G., Brosnan, M., Tiego, J., Johnson, B. P., Fornito, A., Bellgrove, M. A., & Hester, R. (2022). Examining the neural correlates of error awareness in a large fMRI study. *Cerebral Cortex (New York, NY)*, *33*(2), 458–468. https://doi.org/10.1093/cercor/bhac077 (cit. on pp. 23, 45).

Davis, K. D., Taylor, K. S., Hutchison, W. D., Dostrovsky, J. O., McAndrews, M. P., Richter, E. O., & Lozano, A. M. (2005). Human anterior cingulate cortex neurons encode cognitive and emotional demands. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *25*(37), 8402–8406. https://doi.org/10.1523/JNEUROSCI.2315-05.2005 (cit. on p. 39).

de Bruijn, E. R. A., Hulstijn, W., Meulenbroek, R. G. J., & van Galen, G. P. (2003). Action monitoring in motor control: ERPs following selection and execution errors in a force production task. *Psychophysiology*, *40*(5), 786–795. https://doi.org/10.1111/1469-8986.00079 (cit. on p. 46).

de Graaf, T. A., Hsieh, P.-J., & Sack, A. T. (2012). The 'correlates' in neural correlates of consciousness. *Neuroscience & Biobehavioral Reviews*, *36*(1), 191–197. https://doi.org/10.1016/j.neubiorev.2011.05.012 (cit. on p. 120).

De Kock, R., Zhou, W., Datta, P., Mychal Joiner, 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), 20222060. https://doi.org/10.1098/rspb.2022.2060 (cit. on p. 30).

Dehaene, S., Posner, M. I., & Tucker, D. M. (1994). Localization of a Neural System for Error Detection and Compensation. *Psychological Science*, *5*(5), 303–305 (cit. on p. 22).

Del Cul, A., Dehaene, S., Reyes, P., Bravo, E., & Slachevsky, A. (2009). Causal role of prefrontal cortex in the threshold for access to consciousness. *Brain: A Journal of Neurology*, *132*(Pt 9), 2531–2540. https://doi.org/10.1093/brain/awp111 (cit. on p. 19).

Descartes, R. (1637). Discours de la méthode (cit. on p. 13).

Desender, K., Boldt, A., Verguts, T., & Donner, T. H. (2019). Confidence predicts speed-accuracy tradeoff for subsequent decisions. *eLife*, *8*, e43499. https://doi.org/10.7554/eLife.43499 (cit. on p. 124).

Desender, K., Boldt, A., & Yeung, N. (2018). Subjective Confidence Predicts Information Seeking in Decision Making. *Psychological Science*, *29*(5), 761–778. https://doi.org/10.1177/0956797617744771 (cit. on pp. 21, 123).

Desender, K., Ridderinkhof, K. R., & Murphy, P. R. (2021). Understanding neural signals of post-decisional performance monitoring: An integrative review. *eLife*, *10*, e67556. https://doi.org/10.7554/elife.67556 (cit. on p. 45).

Desender, K., Van Opstal, F., Hughes, G., & Van den Bussche, E. (2016). The temporal dynamics of metacognition: Dissociating task-related activity from later metacognitive processes. *Neuropsychologia*, *82*, 54–64. https://doi.org/10.1016/j.neuropsychologia.2016.01.003 (cit. on p. 124).

Desender, K., Van Opstal, F., & Van den Bussche, E. (2014). Feeling the Conflict The Crucial Role of Conflict Experience in Adaptation. *Psychological Science*, *25*. https://doi.org/10.1177/0956797613511468 (cit. on pp. 24, 124).

Desender, K., Van Opstal, F., & Van den Bussche, E. (2017). Subjective experience of difficulty depends on multiple cues. *Scientific Reports*, *7*, 44222. https://doi.org/10.1038/srep44222 (cit. on pp. 24, 159).

Dhar, M., Wiersema, J. R., & Pourtois, G. (2011). Cascade of Neural Events Leading from Error Commission to Subsequent Awareness Revealed Using EEG Source Imaging. *PLoS ONE*, *6*(5), e19578. https://doi.org/10.1371/journal.pone.0019578 (cit. on p. 45).

Di Gregorio, F., Steinhauser, M., & Maier, M. E. (2016). Error-related brain activity and error awareness in an error classification paradigm. *NeuroImage*, *139*, 202–210. https://doi.org/10.1016/j.neuroimage.2016.05.074 (cit. on p. 23).

Di Pellegrino, G., Ciaramelli, E., & Làdavas, E. (2007). The Regulation of Cognitive Control following Rostral Anterior Cingulate Cortex Lesion in Humans. *Journal of Cognitive Neuroscience*, *19*(2), 275–286. https://doi.org/10.1162/jocn.2007.19.2.275 (cit. on p. 39).

Doenyas, C., Mutluer, T., Genç, E., & Balcı, F. (2019). Error monitoring in decision-making and timing is disrupted in autism spectrum disorder. *Autism Research*, *12*(2), 239–248. https://doi.org/10.1002/aur.2041 (cit. on p. 29).

Donders, F. C. (1868). Over de snelheid van psychische processen [On the speed of mental processes]. *Onderzoekingen gedaan in het Physiologisch Laboratorium der Utrechtsche Hoogeschool, 1868–1869, Tweede reeks, 2*, 92–120 (Transl. by W. G. Koster (1969), In W. G. Koster, Attention and performance II. Acta Psychologica, 30, 412–431). https://doi.org/10.1016/0001-6918(69)90065-1 (cit. on p. 31).

Droit-Volet, S., Lamotte, M., & Izaute, M. (2015). The conscious awareness of time distortions regulates the effect of emotion on the perception of time. *Consciousness and Cognition*, *38*, 155–164. https://doi.org/10.1016/j.concog.2015.02.021 (cit. on p. 27).

Droit-Volet, S., & Izaute, M. (2009). Improving time discrimination in children and adults in a temporal bisection task: The effects of feedback and no forced choice on decision and memory processes. *Quarterly Journal of Experimental Psychology*, *62*(6), 1173–1188. https://doi.org/10.1080/17470210802384180 (cit. on p. 27).

Droit-Volet, S., & Meck, W. H. (2007). How emotions colour our perception of time. *Trends in Cognitive Sciences*, *11*(12), 504–513. https://doi.org/10.1016/j.tics.2007.09.008 (cit. on p. 25).

Dunlosky, J., & Metcalfe, J. (2008, September). *Metacognition*. SAGE Publications. (Cit. on p. 17).

Dunlosky, J., & Tauber, S. U. K. (2016). *The Oxford Handbook of Metamemory*. Oxford University Press. (Cit. on p. 17).

Duthoo, W., Abrahamse, E. L., Braem, S., Boehler, C. N., & Notebaert, W. (2014). The heterogeneous world of congruency sequence effects: an update. *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.01001 (cit. on p. 124).

Dutilh, G., Vandekerckhove, J., Forstmann, B. U., Keuleers, E., Brysbaert, M., & Wagenmakers, E.-J. (2012). Testing theories of post-error slowing. *Attention, Perception & Psychophysics*, *74*(2), 454–465. https://doi.org/10.3758/s13414-011-0243-2 (cit. on pp. 23, 123).

Duyan, Y. A., & Balcı, F. (2018). Numerical error monitoring. *Psychonomic Bulletin & Review*, *25*(4), 1549–1555. https://doi.org/10.3758/s13423-018-1506-x (cit. on p. 29).

Duyan, Y. A., & Balcı, F. (2019). Metric error monitoring in the numerical estimates. *Consciousness and Cognition*, *67*, 69–76. https://doi.org/10.1016/j.concog.2018.11.011 (cit. on p. 29).

Duyan, Y. A., & Balcı, F. (2020). Monitoring line length reproduction errors. *Consciousness and Cognition: An International Journal*, *77*. https://doi.org/10.1016/j.concog.2019.102831 (cit. on p. 29).

Eakin, D. K. (2005). Illusions of knowing: Metamemory and memory under conditions of retroactive interference. *Journal of Memory and Language*, *52*(4), 526–534. https://doi.org/10.1016/j.jml.2005.01.009 (cit. on p. 17).

Ehinger, B. V., & Dimigen, O. (2019). Unfold: an integrated toolbox for overlap correction, non-linear modeling, and regression-based EEG analysis. *PeerJ*, *7*, e7838. https://doi.org/10.7717/peerj.7838 (cit. on p. 171).

Emeric, E. E., Brown, J. W., Leslie, M., Pouget, P., Stuphorn, V., & Schall, J. D. (2008). Performance Monitoring Local Field Potentials in the Medial Frontal Cortex of Primates: Anterior Cingulate Cortex. *Journal of neurophysiology*, *99*(2), 759–772. https://doi.org/10.1152/jn.00896.2006 (cit. on p. 39).

Emeric, E. E., Leslie, M., Pouget, P., & Schall, J. D. (2010). Performance Monitoring Local Field Potentials in the Medial Frontal Cortex of Primates: Supplementary Eye Field. *Journal of Neurophysiology*, *104*(3), 1523–1537. https://doi.org/10.1152/jn.01001.2009 (cit. on pp. 41, 43, 44).

Endrass, T., Franke, C., & Kathmann, N. (2005). Error Awareness in a Saccade Countermanding Task. *Journal of Psychophysiology*, *19*(4), 275–280. https://doi.org/10.1027/0269-8803.19.4.275 (cit. on p. 45).

Endrass, T., Klawohn, J., Gruetzmann, R., Ischebeck, M., & Kathmann, N. (2012). Response-related negativities following correct and incorrect responses: Evidence from a temporospatial principal component analysis. *Psychophysiology*, *49*(6), 733–743. https://doi.org/10.1111/j.1469-8986.2012.01365.x (cit. on p. 23).

Endrass, T., Klawohn, J., Schuster, F., & Kathmann, N. (2008). Overactive performance monitoring in obsessive-compulsive disorder: ERP evidence from correct and erroneous reactions. *Neuropsychologia*, *46*(7), 1877–1887. https://doi.org/10.1016/j.neuropsychologia.2007.12.001 (cit. on p. 44).

Endrass, T., Reuter, B., & Kathmann, N. (2007). ERP correlates of conscious error recognition: aware and unaware errors in an antisaccade task. *European Journal of Neuroscience*, *26*(6), 1714–1720. https://doi.org/10.1111/j.1460-9568.2007.05785.x (cit. on pp. 23, 45, 123).

Endrass, T., Schuermann, B., Kaufmann, C., Spielberg, R., Kniesche, R., & Kathmann, N. (2010). Performance monitoring and error significance in patients with obsessive-compulsive disorder. *Biological Psychology*, *84*(2), 257–263. https://doi.org/10.1016/j.biopsycho.2010.02.002 (cit. on p. 44).

Falkenstein, M., Hohnsbein, J., Hoormann, J., & Blanke, L. (1991). Effects of crossmodal divided attention on late ERP components. II. Error processing in choice reaction tasks. *Electroencephalography and Clinical Neurophysiology*, *78*(6), 447–455. https://doi.org/10.1016/0013-4694(91)90062-9 (cit. on pp. 21, 37).

Falkenstein, M. (2004). Errors, Conflicts, and the Brain. *Journal of Psychophysiology*, *18*(4), 153–163. https://doi.org/10.1027/0269-8803.18.4.153 (cit. on p. 42).

Falkenstein, M., Hielscher, H., Dziobek, I., Schwarzenau, P., Hoormann, J., Sundermann, B., & Hohnsbein, J. (2001). Action monitoring, error detection, and the basal ganglia: an ERP study. *NeuroReport*, *12*(1), 157 (cit. on p. 40).

Falkenstein, M., Hoormann, J., Christ, S., & Hohnsbein, J. (2000). ERP components on reaction errors and their functional significance: a tutorial. *Biological Psychology*, *51*(2-3), 87–107. https://doi.org/10.1016/S0301-0511(99)00031-9 (cit. on pp. 22, 44).

Ferrandez, A. M., Hugueville, L., Lehéricy, S., Poline, J. B., Marsault, C., & Pouthas, V. (2003). Basal ganglia and supplementary motor area subtend duration perception: an fMRI study. *NeuroImage*, *19*(4), 1532–1544. https://doi.org/10.1016/s1053-8119(03)00159-9 (cit. on p. 26).

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; a Journal Devoted to the Study of the Nervous System and Behavior*, *120*, 443–456. https://doi.org/10.1016/j.cortex.2019.07.007 (cit. on pp. 23, 45, 111, 119, 168).

Files, B. T., Pollard, K. A., Oiknine, A. H., Khooshabeh, P., & Passaro, A. D. (2021). Correct response negativity may reflect subjective value of reaction time under regulatory fit in a speed-rewarded task. *Psychophysiology*, *58*(9), e13856. https://doi.org/10.1111/psyp.13856 (cit. on pp. 47, 117).

Filevich, E., Koß, C., & Faivre, N. (2020). Response-Related Signals Increase Confidence But Not Metacognitive Performance. *eNeuro*, *7*(3), ENEURO.0326–19.2020. https://doi.org/10.1523/ENEURO.0326-19.2020 (cit. on p. 19).

Fitzgerald, L. M., Arvaneh, M., Carton, S., O'Keeffe, F., Delargy, M., & Dockree, P. M. (2022). Impaired metacognition and reduced neural signals of decision confidence in adults with traumatic brain injury. *Neuropsychology*, *36*(8), 776–790. https://doi.org/10.1037/neu0000854 (cit. on p. 20).

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *American Psychologist*, *34*, 906–911. https://doi.org/10.1037/0003-066X.34.10.906 (cit. on pp. 16, 17).

Fleming, S. M. (2024). Metacognition and Confidence: A Review and Synthesis. *Annual Review of Psychology*, *75*(Volume 75, 2024), 241–268. https://doi.org/10.1146/annurev-psych-022423-032425 (cit. on p. 18).

Fleming, S. M., & Daw, N. D. (2017). Self-Evaluation of Decision-Making: A General Bayesian Framework for Metacognitive Computation. *Psychological Review*, *124*(1), 91–114. https://doi.org/10.1037/rev0000045 (cit. on pp. 19, 111).

Fleming, S. M., & Dolan, R. J. (2012). The neural basis of metacognitive ability. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *367*(1594), 1338–1349. https://doi.org/10.1098/rstb.2011.0417 (cit. on pp. 20, 121).

Fleming, S. M., & Lau, H. C. (2014). How to measure metacognition. *Frontiers in Human Neuroscience*, *8*. https://doi.org/10.3389/fnhum.2014.00443 (cit. on p. 18).

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 (cit. on pp. 19, 160).

Fleming, S. M., Ryu, J., Golfinos, J. G., & Blackmon, K. E. (2014). Domain-specific impairment in metacognitive accuracy following anterior prefrontal lesions. *Brain*, *137*(10), 2811–2822. https://doi.org/10.1093/brain/awu221 (cit. on pp. 19, 20).

Fleming, S. M., Weil, R. S., Nagy, Z., Dolan, R. J., & Rees, G. (2010). Relating introspective accuracy to individual differences in brain structure. *Science (New York, N.Y.)*, *329*(5998), 1541–1543. https://doi.org/10.1126/science.1191883 (cit. on p. 20).

Foote, A. L., & Crystal, J. D. (2007). Metacognition in the rat. *Current biology : CB*, *17*(6), 551–555. https://doi.org/10.1016/j.cub.2007.01.061 (cit. on p. 27).

Foote, A. L., & Crystal, J. D. (2012). "Play it Again": a new method for testing metacognition in animals. *Animal Cognition*, *15*(2), 187–199. https://doi.org/10.1007/s10071-011-0445-y (cit. on p. 27).

Ford, J. M. (1999). Schizophrenia: The broken P300 and beyond. *Psychophysiology*, *36*(6), 667–682. https://doi.org/10.1111/1469-8986.3660667 (cit. on p. 42).

Fortin, C., Schweickert, R., Gaudreault, R., & Viau-Quesnel, C. (2010). Timing is affected by demands in memory search but not by task switching. *Journal of Experimental Psychology: Human Perception and Performance*, *36*(3), 580–595. https://doi.org/10.1037/a0017639 (cit. on p. 35).

Franken, I. H. A., van Strien, J. W., Franzek, E. J., & van de Wetering, B. J. (2007). Error-processing deficits in patients with cocaine dependence. *Biological Psychology*, *75*(1), 45–51. https://doi.org/10.1016/j.biopsycho.2006.11.003 (cit. on p. 22).

Friedemann, M., Bang, D., & Yeung, N. (2023, October). Confidence-Guided Information Seeking. https://doi.org/10.2139/ssrn.4610715 (cit. on p. 21).

Frith, C., Perry, R., & Lumer, E. (1999). The neural correlates of conscious experience: an experimental framework. *Trends in Cognitive Sciences*, *3*(3), 105–114. https://doi.org/10.1016/S1364-6613(99)01281-4 (cit. on p. 78).

Fu, Z., Beam, D., Chung, J. M., Reed, C. M., Mamelak, A. N., Adolphs, R., & Rutishauser, U. (2022). The geometry of domain-general performance monitoring in the human medial frontal cortex. *Science (New York, N.Y.)*, *376*(6593), eabm9922. https://doi.org/10.1126/science.abm9922 (cit. on p. 30).

Fu, Z., Wu, D.-A. J., Ross, I., Chung, J. M., Mamelak, A. N., Adolphs, R., & Rutishauser, U. (2019). Single-Neuron Correlates of Error Monitoring and Post-Error Adjustments in Human Medial Frontal Cortex. *Neuron*, *101*(1), 165–177.e5. https://doi.org/10.1016/j.neuron.2018.11.016 (cit. on pp. 42–44).

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), niz001. https://doi.org/10.1093/nc/niz001 (cit. on pp. 19, 160).

Galvin, S. J., Podd, J. V., Drga, V., & Whitmore, J. (2003). Type 2 tasks in the theory of signal detectability: Discrimination between correct and incorrect decisions. *Psychonomic Bulletin & Review*, *10*(4), 843–876. https://doi.org/10.3758/BF03196546 (cit. on p. 19).

Gamache, P.-L., Grondin, S., & Zakay, D. (2011). The Impact of Attention on the Internal Clock in Prospective Timing: Is It Direct or Indirect? In A. Vatakis, A. Esposito, M. Giagkou, F. Cummins, & G. Papadelis (Eds.), *Multidisciplinary Aspects of Time and Time Perception: COST TD0904 International Workshop, Athens, Greece, October 7-8, 2010, Revised Selected Papers* (pp. 137–150). Springer. https://doi.org/10.1007/978-3-642-21478-3_12 (cit. on p. 25).

Gao, J., Wong-Lin, K., Holmes, P., Simen, P., & Cohen, J. D. (2009). Sequential Effects in Two-Choice Reaction Time Tasks: Decomposition and Synthesis of Mechanisms. *Neural Computation*, *21*(9), 2407–2436. https://doi.org/10.1162/neco.2009.09-08-866 (cit. on p. 124).

Garavan, H., Ross, T. J., Murphy, K., Roche, R. a. P., & Stein, E. A. (2002). Dissociable executive functions in the dynamic control of behavior: inhibition, error detection, and correction. *NeuroImage*, *17*(4), 1820–1829. https://doi.org/10.1006/nimg.2002.1326 (cit. on p. 22).

Gardelle, V. d., & Mamassian, P. (2015). Weighting Mean and Variability during Confidence Judgments. *PLOS ONE*, *10*(3), e0120870. https://doi.org/10.1371/journal.pone.0120870 (cit. on p. 18).

Gehring, W. J., & Knight, R. T. (2000). Prefrontal-cingulate interactions in action monitoring. *Nature Neuroscience*, *3*(5), 516–520. https://doi.org/10.1038/74899 (cit. on p. 43).

Gehring, W. J., & Fencsik, D. E. (2001). Functions of the medial frontal cortex in the processing of conflict and errors. *The Journal of Neuroscience*, *21*(23), 9430–9437. https://doi.org/10.1523/JNEUROSCI.21-23-09430.2001 (cit. on p. 37).

Gehring, W. J., Goss, B., Coles, M. G. H., Meyer, D. E., & Donchin, E. (1993). A Neural System for Error Detection and Compensation. *Psychological Science*, *4*(6), 385–390. https://doi.org/10.1111/j.1467-9280.1993.tb00586.x (cit. on pp. 21, 37, 44).

Gehring, W. J., & Willoughby, A. R. (2002). The Medial Frontal Cortex and the Rapid Processing of Monetary Gains and Losses. *Science*, *295*(5563), 2279–2282. https://doi.org/10.1126/science.1066893 (cit. on p. 39).

Gemba, H., Sasaki, K., & Brooks, V. B. (1986). 'Error' potentials in limbic cortex (anterior cingulate area 24) of monkeys during motor learning. *Neuroscience Letters*, *70*(2), 223–227. https://doi.org/10.1016/0304-3940(86)90467-2 (cit. on p. 39).

Gerstner, W., Kreiter, A. K., Markram, H., & Herz, A. V. M. (1997). Neural codes: Firing rates and beyond. *Proceedings of the National Academy of Sciences of the United States of America*, *94*(24), 12740–12741 (cit. on p. 118).

Gibbon, J. (1977). Scalar expectancy theory and Weber's law in animal timing. *Psychological Review*, *84*(3), 279–325. https://doi.org/10.1037/0033-295X.84.3.279 (cit. on p. 25).

Gibbon, J., Church, R. M., & Meck, W. H. (1984). Scalar Timing in Memory. *Annals of the New York Academy of Sciences*, *423*(1), 52–77. https://doi.org/10.1111/j.1749-6632.1984.tb23417.x (cit. on pp. 25, 26, 164).

Gilden, D. L. (2001). Cognitive emissions of 1/f noise. *Psychological Review*, *108*(1), 33–56. https://doi.org/10.1037/0033-295X.108.1.33 (cit. on p. 132).

Godlove, D. C., Emeric, E. E., Segovis, C. M., Young, M. S., Schall, J. D., & Woodman, G. F. (2011). Event-Related Potentials Elicited by Errors during the Stop-Signal Task. I. Macaque Monkeys. *The Journal of Neuroscience*, *31*(44), 15640–15649. https://doi.org/10.1523/JNEUROSCI.3349-11.2011 (cit. on p. 37).

Gold, J. I., & Shadlen, M. N. (2007). The Neural Basis of Decision Making. *Annual Review of Neuroscience*, *30*(1), 535–574. https://doi.org/10.1146/annurev.neuro.29.051605.113038 (cit. on pp. 32, 167).

Granger, C. W. J. (1969). Investigating Causal Relations by Econometric Models and Cross-spectral Methods. *Econometrica*, *37*(3), 424–438. https://doi.org/10.2307/1912791 (cit. on p. 124).

Gratton, G., Coles, M. G. H., & Donchin, E. (1992). Optimizing the use of information: Strategic control of activation of responses. *Journal of Experimental Psychology: General*, *121*(4), 480–506. https://doi.org/10.1037/0096-3445.121.4.480 (cit. on p. 124).

Grimaldi, P., Lau, H., & Basso, M. A. (2015). There are things that we know that we know, and there are things that we do not know we do not know: Confidence in Decision-Making. *Neuroscience and biobehavioral reviews*, *55*, 88–97. https://doi.org/10.1016/j.neubiorev.2015.04.006 (cit. on p. 20).

Grondin, S. (2010). Timing and time perception: A review of recent behavioral and neuroscience findings and theoretical directions. *Attention, Perception, & Psychophysics*, *72*(3), 561–582. https://doi.org/10.3758/APP.72.3.561 (cit. on p. 25).

Grützmann, R., Endrass, T., Klawohn, J., & Kathmann, N. (2014). Response accuracy rating modulates ERN and Pe amplitudes. *Biological Psychology*, *96*, 1–7. https://doi.org/10.1016/j.biopsycho.2013.10.007 (cit. on pp. 44, 78, 110, 123, 131).

Haggard, P., Newman, C., & Magno, E. (1999). On the perceived time of voluntary actions. *British Journal of Psychology*, *90*(2), 291–303. https://doi.org/10.1348/000712699161413 (cit. on p. 168).

Hajcak, G., Moser, J. S., Yeung, N., & Simons, R. F. (2005). On the ERN and the significance of errors. *Psychophysiology*, *42*(2), 151–160. https://doi.org/10.1111/j.1469-8986.2005.00270.x (cit. on p. 44).

Hampton, R. R. (2001). Rhesus monkeys know when they remember. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(9), 5359–5362. https://doi.org/10.1073/pnas.071600998 (cit. on p. 18).

Hampton, R. R., Zivin, A., & Murray, E. A. (2004). Rhesus monkeys (Macaca mulatta) discriminate between knowing and not knowing and collect information as needed before acting. *Animal Cognition*, *7*(4), 239–246. https://doi.org/10.1007/s10071-004-0215-1 (cit. on p. 18).

Harsay, H. A., Cohen, M. X., Spaan, M., Weeda, W. D., Nieuwenhuis, S., & Ridderinkhof, K. R. (2018). Error blindness and motivational significance: Shifts in networks centering on anterior insula co-vary with error awareness and pupil dilation. *Behavioural Brain Research*, *355*, 24–35. https://doi.org/10.1016/j.bbr.2017.10.030 (cit. on pp. 23, 124).

Hart, J. T. (1965). Memory and the feeling-of-knowing experience. *Journal of Educational Psychology*, *56*(4), 208–216. https://doi.org/10.1037/h0022263 (cit. on p. 17).

Hasbroucq, T., Akamatsu, M., Burle, B., Bonnet, M., & Possamaï, C. A. (2000). Changes in spinal excitability during choice reaction time: the H reflex as a probe of information transmission. *Psychophysiology*, *37*(3), 385–393 (cit. on p. 110).

Hasbroucq, T., Possamaï, C. A., Bonnet, M., & Vidal, F. (1999). Effect of the irrelevant location of the response signal on choice reaction time: an electromyographic study in humans. *Psychophysiology*, *36*(4), 522–526. https://doi.org/10.1017/s0048577299001602 (cit. on pp. 21, 33).

Heitz, R. P. (2014). The speed-accuracy tradeoff: history, physiology, methodology, and behavior. *Frontiers in Neuroscience*, *8*, 150. https://doi.org/10.3389/fnins.2014.00150 (cit. on p. 131).

Herrmann, M. J., Römmler, J., Ehlis, A.-C., Heidrich, A., & Fallgatter, A. J. (2004). Source localization (LORETA) of the error-related-negativity (ERN/Ne) and positivity (Pe). *Brain Research. Cognitive Brain Research*, *20*(2), 294–299. https://doi.org/10.1016/j.cogbrainres.2004.02.013 (cit. on p. 22).

Hester, R., Foxe, J. J., Molholm, S., Shpaner, M., & Garavan, H. (2005). Neural mechanisms involved in error processing: A comparison of errors made with and without awareness. *NeuroImage*, *27*(3), 602–608. https://doi.org/10.1016/j.neuroimage.2005.04.035 (cit. on pp. 23, 45, 124).

Hester, R., Nandam, L. S., O'Connell, R. G., Wagner, J., Strudwick, M., Nathan, P. J., Mattingley, J. B., & Bellgrove, M. A. (2012). Neurochemical Enhancement of Conscious Error Awareness. *Journal of Neuroscience*, *32*(8), 2619–2627. https://doi.org/10.1523/JNEUROSCI.4052-11.2012 (cit. on p. 45).

Hewig, J., Coles, M. G. H., Trippe, R. H., Hecht, H., & Miltner, W. H. R. (2011). Dissociation of Pe and ERN/Ne in the conscious recognition of an error. *Psychophysiology*, *48*(10), 1390–1396. https://doi.org/10.1111/j.1469-8986.2011.01209.x (cit. on p. 45).

Holroyd, C. B., & Coles, M. G. H. (2002). The neural basis of human error processing: Reinforcement learning, dopamine, and the error-related negativity. *Psychological Review*, *109*(4), 679–709. https://doi.org/10.1037/0033-295X.109.4.679 (cit. on pp. 22, 39, 162).

Holroyd, C. B., Nieuwenhuis, S., Mars, R. B., & Coles, M. G. H. (2004). Anterior Cingulate Cortex, Selection for Action, and Error Processing. In *Cognitive neuroscience of attention* (pp. 219–231). The Guilford Press. (Cit. on p. 39).

Hoonakker, M., Doignon-Camus, N., & Bonnefond, A. (2016). Performance monitoring mechanisms activated before and after a response: A comparison of aware and

unaware errors. *Biological Psychology*, *120*, 53–60. https://doi.org/10.1016/j.biopsycho.2016.08.009 (cit. on p. 45).

Hume, D. (1739). *A Treatise of Human Nature*. Oxford University Press. (Cit. on p. 14).

Isomura, Y., Ito, Y., Akazawa, T., Nambu, A., & Takada, M. (2003). Neural Coding of "Attention for Action" and "Response Selection" in Primate Anterior Cingulate Cortex. *The Journal of Neuroscience*, *23*(22), 8002–8012. https://doi.org/10.1523/JNEUROSCI.23-22-08002.2003 (cit. on p. 39).

Ito, S., Stuphorn, V., Brown, J. W., & Schall, J. D. (2003). Performance Monitoring by the Anterior Cingulate Cortex During Saccade Countermanding. *Science*, *302*(5642), 120–122. https://doi.org/10.1126/science.1087847 (cit. on p. 39).

Jazayeri, M., & Shadlen, M. N. (2010). Temporal context calibrates interval timing. *Nature Neuroscience*, *13*(8), 1020–1026. https://doi.org/10.1038/nn.2590 (cit. on p. 28).

Jessup, R. K., Busemeyer, J. R., & Brown, J. W. (2010). Error Effects in Anterior Cingulate Cortex Reverse when Error Likelihood Is High. *The Journal of Neuroscience*, *30*(9), 3467–3472. https://doi.org/10.1523/JNEUROSCI.4130-09.2010 (cit. on p. 41).

Jones, K. E., Hamilton, A. F., & Wolpert, D. M. (2002). Sources of signal-dependent noise during isometric force production. *Journal of Neurophysiology*, *88*(3), 1533–1544. https://doi.org/10.1152/jn.2002.88.3.1533 (cit. on p. 132).

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 (cit. on p. 31).

Kant, I. (1786). *Premiers principes métaphysiques de la science de la nature*. (Cit. on p. 14).

Keane, B., Spence, M., Yarrow, K., & Arnold, D. (2015). Perceptual confidence demonstrates trial-by-trial insight into the precision of audio–visual timing encoding. *Consciousness and Cognition*, *38*, 107–117. https://doi.org/10.1016/j.concog.2015.10.010 (cit. on p. 27).

Kepecs, A., & Mainen, Z. F. (2012). A computational framework for the study of confidence in humans and animals. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1594), 1322–1337. https://doi.org/10.1098/rstb.2012.0037 (cit. on p. 18).

Kepecs, A., Uchida, N., Zariwala, H. A., & Mainen, Z. F. (2008). Neural correlates, computation and behavioural impact of decision confidence. *Nature*, *455*(7210), 227–231. https://doi.org/10.1038/nature07200 (cit. on p. 18).

Kiani, R., Corthell, L., & Shadlen, M. N. (2014). Choice certainty is informed by both evidence and decision time. *Neuron*, *84*(6), 1329. https://doi.org/10.1016/j.neuron.2014.12.015 (cit. on pp. 19, 24, 160).

Kiani, R., & Shadlen, M. N. (2009). Representation of Confidence Associated with a Decision by Neurons in the Parietal Cortex. *Science*, *324*(5928), 759–764. https://doi.org/10.1126/science.1169405 (cit. on pp. 18, 19, 167).

Kirby, N. H. (1976). Sequential effects in two-choice reaction time: Automatic facilitation or subjective expectancy? *Journal of Experimental Psychology: Human Perception and Performance*, *2*(4), 567–577. https://doi.org/10.1037/0096-1523.2.4.567 (cit. on p. 124).

Kirschner, H., Humann, J., Derrfuss, J., Danielmeier, C., & Ullsperger, M. (2021). Neural and behavioral traces of error awareness. *Cognitive, Affective & Behavioral Neuroscience*, *21*(3), 573–591. https://doi.org/10.3758/s13415-020-00838-w (cit. on p. 45).

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 (cit. on pp. 35, 159).

Klein, T. A., Endrass, T., Kathmann, N., Neumann, J., von Cramon, D. Y., & Ullsperger, M. (2007). Neural correlates of error awareness. *NeuroImage*, *34*(4), 1774–1781. https://doi.org/10.1016/j.neuroimage.2006.11.014 (cit. on p. 45).

Klein, T. A., Ullsperger, M., & Danielmeier, C. (Eds.). (2014). *Error Awareness – Insights From Cognitive Neuroscience, Psychiatry And Neurology*. Frontiers Media SA. https://doi.org/10.3389/978-2-88919-201-4 (cit. on p. 45).

Koch, G., Costa, A., Brusa, L., Peppe, A., Gatto, I., Torriero, S., Gerfo, E. L., Salerno, S., Oliveri, M., Carlesimo, G. A., & Caltagirone, C. (2008). Impaired reproduction of second but not millisecond time intervals in Parkinson's disease. *Neuropsychologia*, *46*(5), 1305–1313. https://doi.org/10.1016/j.neuropsychologia.2007.12.005 (cit. on p. 26).

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 (cit. on pp. 29, 30).

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 (cit. on pp. 29, 30).

Kononowicz, T. W., van Wassenhove, V., & Doyère, V. (2022). Rodents monitor their error in self-generated duration on a single trial basis. *Proceedings of the National Academy of Sciences of the United States of America*, *119*(9), e2108850119. https://doi.org/10.1073/pnas.2108850119 (cit. on p. 30).

Koriat, A. (1993). How do we know that we know? The accessibility model of the feeling of knowing. *Psychological Review*, *100*(4), 609–639. https://doi.org/10.1037/0033-295X.100.4.609 (cit. on p. 17).

Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General*, *126*(4), 349–370. https://doi.org/10.1037/0096-3445.126.4.349 (cit. on pp. 17, 159).

Koriat, A., & Goldsmith, M. (1996). Monitoring and Control Processes in the Strategic Regulation of Memory Accuracy (cit. on p. 17).

Krigolson, O., Pierce, L., Holroyd, C., & Tanaka, J. (2009). Learning to Become an Expert: Reinforcement Learning and the Acquisition of Perceptual Expertise. *Journal of*

cognitive neuroscience, *21*, 1834–41. https://doi.org/10.1162/jocn.2009.21128 (cit. on p. 40).

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, *82*, 1–26. https://doi.org/10.18637/jss.v082.i13 (cit. on p. 127).

Lake, J. I., & Meck, W. H. (2013). Differential effects of amphetamine and haloperidol on temporal reproduction: dopaminergic regulation of attention and clock speed. *Neuropsychologia*, *51*(2), 284–292. https://doi.org/10.1016/j.neuropsychologia.2012.09.014 (cit. on p. 26).

Laming, D. (1979a). Autocorrelation of choice-reaction times. *Acta Psychologica*, *43*(5), 381–412. https://doi.org/10.1016/0001-6918(79)90032-5 (cit. on pp. 125, 132).

Laming, D. (1979b). Choice reaction performance following an error. *Acta Psychologica*, *43*, 199–224. https://doi.org/10.1016/0001-6918(79)90026-X (cit. on p. 23).

Lamotte, M., Droit-Volet, S., & Izaute, M. (2017). Confidence judgment in a temporal generalization task : Accuracy and sensitivity to task difficulty. *L'Année psychologique*, *117*(3), 275–298. https://doi.org/10.3917/anpsy.173.0275 (cit. on p. 27).

Le Pelley, M. E. (2012). Metacognitive monkeys or associative animals? Simple reinforcement learning explains uncertainty in nonhuman animals. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *38*(3), 686–708. https://doi.org/10.1037/a0026478 (cit. on p. 21).

Leocani, L., Cohen, L. G., Wassermann, E. M., Ikoma, K., & Hallett, M. (2000). Human corticospinal excitability evaluated with transcranial magnetic stimulation during different reaction time paradigms. *Brain: A Journal of Neurology*, *123 (Pt 6)*, 1161–1173. https://doi.org/10.1093/brain/123.6.1161 (cit. on p. 110).

Lichtenstein, S., & Fischhoff, B. (1977). Do those who know more also know more about how much they know? *Organizational Behavior and Human Performance*, *20*(2), 159–183. https://doi.org/10.1016/0030-5073(77)90001-0 (cit. on p. 27).

Ljungberg, T., Apicella, P., & Schultz, W. (1992). Responses of monkey dopamine neurons during learning of behavioral reactions. *Journal of Neurophysiology*, *67*(1), 145–163. https://doi.org/10.1152/jn.1992.67.1.145 (cit. on p. 40).

Los, S. A. (1996). On the origin of mixing costs: Exploring information processing in pure and mixed blocks of trials. *Acta Psychologica*, *94*(2), 145–188. https://doi.org/10.1016/0001-6918(95)00050-X (cit. on p. 125).

Lütcke, H., & Frahm, J. (2008). Lateralized Anterior Cingulate Function during Error Processing and Conflict Monitoring as Revealed by High-Resolution fMRI. *Cerebral Cortex*, *18*(3), 508–515. https://doi.org/10.1093/cercor/bhm090 (cit. on p. 39).

Luu, P., & Tucker, D. M. (2001). Regulating action: alternating activation of midline frontal and motor cortical networks. *Clinical Neurophysiology: Official Journal of the International Federation of Clinical Neurophysiology*, *112*(7), 1295–1306. https://doi.org/10.1016/s1388-2457(01)00559-4 (cit. on p. 42).

Luu, P., Flaisch, T., & Tucker, D. M. (2000). Medial Frontal Cortex in Action Monitoring. *The Journal of Neuroscience*, *20*(1), 464–469. https://doi.org/10.1523/JNEUROSCI.20-01-00464.2000 (cit. on p. 47).

Luu, P., Tucker, D. M., Derryberry, D., Reed, M., & Poulsen, C. (2003). Electrophysiological Responses to Errors and Feedback in the Process of Action Regulation. *Psychological Science*, *14*(1), 47–53. https://doi.org/10.1111/1467-9280.01417 (cit. on p. 39).

Macar, F., Vidal, F., & Casini, L. (1999). The supplementary motor area in motor and sensory timing: evidence from slow brain potential changes. *Experimental Brain Research*, *125*(3), 271–280. https://doi.org/10.1007/s002210050683 (cit. on p. 167).

MacLean, S. J., Hassall, C. D., Ishigami, Y., Krigolson, O. E., & Eskes, G. A. (2015). Using brain potentials to understand prism adaptation: the error-related negativity and the P300. *Frontiers in Human Neuroscience*, *9* (cit. on p. 46).

Maier, M., Steinhauser, M., & Hübner, R. (2008). Is the error-related negativity amplitude related to error detectability? Evidence from effects of different error types. *Journal of Cognitive Neuroscience*, *20*(12), 2263–2273. https://doi.org/10.1162/jocn.2008.20159 (cit. on p. 45).

Maier, M. E., & Steinhauser, M. (2016). Error significance but not error expectancy predicts error-related negativities for different error types. *Behavioural Brain Research*, *297*, 259–267. https://doi.org/10.1016/j.bbr.2015.10.031 (cit. on p. 44).

Malapani, C., & Fairhurst, S. (2002). Scalar Timing in Animals and Humans. *Learning and Motivation*, *33*(1), 156–176. https://doi.org/10.1006/lmot.2001.1105 (cit. on p. 25).

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 (cit. on p. 31).

Mamassian, P. (2016). Visual Confidence. *Annual Review of Vision Science*, *2*(1), 459–481. https://doi.org/10.1146/annurev-vision-111815-114630 (cit. on p. 18).

Maniscalco, B., & Lau, H. (2012). A signal detection theoretic approach for estimating metacognitive sensitivity from confidence ratings. *Consciousness and Cognition*, *21*(1), 422–430. https://doi.org/10.1016/j.concog.2011.09.021 (cit. on p. 19).

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 (cit. on pp. 35, 49).

Masaki, H., Tanaka, H., Takasawa, N., & Yamazaki, K. (2001). Error-related brain potentials elicited by vocal errors. *Neuroreport*, *12*(9), 1851–1855. https://doi.org/10.1097/00001756-200107030-00018 (cit. on p. 37).

Matell, M. S., & Meck, W. H. (2004). Cortico-striatal circuits and interval timing: coincidence detection of oscillatory processes. *Cognitive Brain Research*, *21*(2), 139–170. https://doi.org/10.1016/j.cogbrainres.2004.06.012 (cit. on p. 25).

Mathalon, D. H., Bennett, A., Askari, N., Gray, E. M., Rosenbloom, M. J., & Ford, J. M. (2003). Response-monitoring dysfunction in aging and Alzheimer's disease: an event-related potential study. *Neurobiology of Aging*, *24*(5), 675–685. https://doi.org/10.1016/s0197-4580(02)00154-9 (cit. on p. 22).

Mathalon, D. H., Fedor, M., Faustman, W. O., Gray, M., Askari, N., & Ford, J. M. (2002). Response-monitoring dysfunction in schizophrenia: an event-related brain potential study. *Journal of Abnormal Psychology*, *111*(1), 22–41 (cit. on p. 43).

Matsumoto, M., Matsumoto, K., Abe, H., & Tanaka, K. (2007). Medial prefrontal cell activity signaling prediction errors of action values. *Nature Neuroscience*, *10*(5), 647–656. https://doi.org/10.1038/nn1890 (cit. on p. 40).

Mayr, U., Awh, E., & Laurey, P. (2003). Conflict adaptation effects in the absence of executive control. *Nature Neuroscience*, *6*(5), 450–452. https://doi.org/10.1038/nn1051 (cit. on p. 124).

McCurdy, L. Y., Maniscalco, B., Metcalfe, J., Liu, K. Y., de Lange, F. P., & Lau, H. (2013). Anatomical coupling between distinct metacognitive systems for memory and visual perception. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *33*(5), 1897–1906. https://doi.org/10.1523/JNEUROSCI.1890-12.2013 (cit. on p. 20).

Meck, W. H. (2005). Neuropsychology of timing and time perception. *Brain and Cognition*, *58*(1), 1–8. https://doi.org/10.1016/j.bandc.2004.09.004 (cit. on p. 25).

Meckler, C. (2010). *Proactive and reactive processes of response monitoring for sensorimotor activities in humans: behavioral and electrophysiological studies* [Doctoral dissertation]. (Cit. on pp. 46, 111).

Meckler, C., Allain, S., Carbonnell, L., Hasbroucq, T., Burle, B., & Vidal, F. (2011). Executive control and response expectancy: A Laplacian ERP study. *Psychophysiology*, *48*(3), 303–311 (cit. on pp. 41, 43).

Meckler, C., Carbonnell, L., Ramdani, C., Hasbroucq, T., & Vidal, F. (2017). On-line action monitoring of response execution: An electrophysiological study. *Biological Psychology*, *129*, 178–185. https://doi.org/10.1016/j.biopsycho.2017.09.002 (cit. on p. 46).

Merchant, H., Harrington, D. L., & Meck, W. H. (2013). Neural Basis of the Perception and Estimation of Time. *Annual Review of Neuroscience*, *36*(1), 313–336. https://doi.org/10.1146/annurev-neuro-062012-170349 (cit. on p. 26).

Miller, G. A. (2003). The cognitive revolution: a historical perspective. *Trends in Cognitive Sciences*, *7*(3), 141–144. https://doi.org/10.1016/S1364-6613(03)00029-9 (cit. on p. 15).

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 (cit. on p. 170).

Miltner, W. H. R., Braun, C. H., & Coles, M. G. H. (1997). Event-Related Brain Potentials Following Incorrect Feedback in a Time-Estimation Task: Evidence for a

"Generic" Neural System for Error Detection. *Journal of Cognitive Neuroscience*, *9*(6), 788–798. https://doi.org/10.1162/jocn.1997.9.6.788 (cit. on p. 39).

Miltner, W. H. R., Lemke, U., Weiss, T., Holroyd, C., Scheffers, M. K., & Coles, M. G. H. (2003). Implementation of error-processing in the human anterior cingulate cortex: a source analysis of the magnetic equivalent of the error-related negativity. *Biological Psychology*, *64*(1), 157–166. https://doi.org/10.1016/S0301-0511(03)00107-8 (cit. on p. 22).

Miyamoto, K., Osada, T., Setsuie, R., Takeda, M., Tamura, K., Adachi, Y., & Miyashita, Y. (2017). Causal neural network of metamemory for retrospection in primates. *Science (New York, N.Y.)*, *355*(6321), 188–193. https://doi.org/10.1126/science.aal0162 (cit. on p. 18).

Moreno-Bote, R. (2010). Decision Confidence and Uncertainty in Diffusion Models with Partially Correlated Neuronal Integrators. *Neural Computation*, *22*(7), 1786–1811. https://doi.org/10.1162/neco.2010.12-08-930 (cit. on p. 19).

Morris, S. E., Yee, C. M., & Nuechterlein, K. H. (2006). Electrophysiological analysis of error monitoring in schizophrenia. *Journal of Abnormal Psychology*, *115*(2), 239–250. https://doi.org/10.1037/0021-843X.115.2.239 (cit. on p. 22).

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 (cit. on pp. 24, 124).

Murphy, P. R., Robertson, I. H., Allen, D., Hester, R., & O'Connell, R. G. (2012). An electrophysiological signal that precisely tracks the emergence of error awareness. *Frontiers in Human Neuroscience*, *6*, 65. https://doi.org/10.3389/fnhum.2012.00065 (cit. on p. 45).

Nakamura, K., Roesch, M. R., & Olson, C. R. (2005). Neuronal Activity in Macaque SEF and ACC During Performance of Tasks Involving Conflict. *Journal of Neurophysiology*, *93*(2), 884–908. https://doi.org/10.1152/jn.00305.2004 (cit. on p. 39).

Nelson, T. O. (1990). Metamemory: A Theoretical Framework and New Findings. In *Psychology of Learning and Motivation* (pp. 125–173, Vol. 26). Elsevier. https://doi.org/10.1016/S0079-7421(08)60053-5 (cit. on pp. 16, 17, 122, 123).

Nelson, T. O., & Dunlosky, J. (1991). When People's Judgments of Learning (JOLs) Are Extremely Accurate at Predicting Subsequent Recall: The "Delayed-JOL Effect". *Psychological Science*, *2*(4), 267–270 (cit. on p. 17).

Niessen, E., Fink, G. R., Hoffmann, H. E. M., Weiss, P. H., & Stahl, J. (2017). Error detection across the adult lifespan: Electrophysiological evidence for age-related deficits. *NeuroImage*, *152*, 517–529. https://doi.org/10.1016/j.neuroimage.2017.03.015 (cit. on p. 45).

Nieuwenhuis, S., Holroyd, C. B., Mol, N., & Coles, M. G. H. (2004). Reinforcement-related brain potentials from medial frontal cortex: origins and functional significance. *Neuroscience & Biobehavioral Reviews*, *28*(4), 441–448. https://doi.org/10.1016/j.neubiorev.2004.05.003 (cit. on p. 39).

Nieuwenhuis, S., Ridderinkhof, K. R., Blom, J., Band, G. P., & Kok, A. (2001). Error-related brain potentials are differentially related to awareness of response errors: Evidence from an antisaccade task. *Psychophysiology*, *38*(5), 752–760. https://doi.org/10.1111/1469-8986.3850752 (cit. on pp. 21, 23, 44, 45, 123).

Niki, H., & Watanabe, M. (1979). Prefrontal and cingulate unit activity during timing behavior in the monkey. *Brain Research*, *171*(2), 213–224. https://doi.org/10.1016/0006-8993(79)90328-7 (cit. on p. 39).

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 (cit. on p. 15).

Norman, E., Pfuhl, G., Sæle, R. G., Svartdal, F., Låg, T., & Dahl, T. I. (2019). Metacognition in Psychology. *Review of General Psychology*, *23*(4), 403–424. https://doi.org/10.1177/1089268019883821 (cit. on p. 16).

Notebaert, W., Houtman, F., Opstal, F. V., Gevers, W., Fias, W., & Verguts, T. (2009). Post-error slowing: an orienting account. *Cognition*, *111*(2), 275–279. https://doi.org/10.1016/j.cognition.2009.02.002 (cit. on p. 23).

Nunez, P. L., Silberstein, R. B., Cadusch, P. J., Wijesinghe, R. S., Westdorp, A. F., & Srinivasan, R. (1994). A theoretical and experimental study of high resolution EEG based on surface Laplacians and cortical imaging. *Electroencephalography and Clinical Neurophysiology*, *90*(1), 40–57. https://doi.org/10.1016/0013-4694(94)90112-0 (cit. on p. 42).

O'Connell, R., Dockree, P., & Kelly, S. (2012). A supramodal accumulation-to-bound signal that determines perceptual decisions in humans. *Nature neuroscience*, *15*. https://doi.org/10.1038/nn.3248 (cit. on p. 167).

O'Connell, R. G., Dockree, P. M., Bellgrove, M. A., Kelly, S. P., Hester, R., Garavan, H., Robertson, I. H., & Foxe, J. J. (2007). The role of cingulate cortex in the detection of errors with and without awareness: a high-density electrical mapping study: Error awareness. *European Journal of Neuroscience*, *25*(8), 2571–2579. https://doi.org/10.1111/j.1460-9568.2007.05477.x (cit. on pp. 23, 45).

Orr, C., & Hester, R. (2012). Error-related anterior cingulate cortex activity and the prediction of conscious error awareness. *Frontiers in Human Neuroscience*, *6*. https://doi.org/10.3389/fnhum.2012.00177 (cit. on p. 45).

Overbeek, T. J., Nieuwenhuis, S., & Ridderinkhof, K. R. (2005). Dissociable Components of Error Processing: On the Functional Significance of the Pe Vis-à-vis the ERN/Ne. *Journal of Psychophysiology*, *19*(4), 319–329. https://doi.org/10.1027/0269-8803.19.4.319 (cit. on pp. 22, 45).

Overgaard, M. (2006). Introspection in Science. *Consciousness and Cognition*, *15*(4), 629–633. https://doi.org/10.1016/j.concog.2006.10.004 (cit. on pp. 15, 171).

Öztel, T., & Balci, F. (2022). Temporal Error Monitoring: Does Ownership of Errors Matter? *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4237109 (cit. on pp. 29, 30).

Öztel, T., & Balci, F. (2023). Temporal Error Monitoring Does Not Depend on Working Memory. *Psychological Reports*, 00332941231187121. https://doi.org/10.1177/00332941231187121 (cit. on pp. 29, 30, 158).

Öztel, T., & Balcı, F. (2020). Stimulus-Induced Temporal Illusions: When Awareness Is Mesmerized by Time. *Timing & Time Perception*, *8*(3-4), 363–380. https://doi.org/10.1163/22134468-bja10015 (cit. on p. 27).

Öztel, T., Eskenazi, T., & Balcı, F. (2021). Temporal error monitoring with directional error magnitude judgements: a robust phenomenon with no effect of being watched. *Psychological Research*, *85*(5), 2069–2078. https://doi.org/10.1007/s00426-020-01379-0 (cit. on pp. 29, 30).

Pashler, H. (1994). Dual-task interference in simple tasks: Data and theory. *Psychological Bulletin*, *116*(2), 220–244. https://doi.org/10.1037/0033-2909.116.2.220 (cit. on p. 131).

Pashler, H., & Johnston, J. C. (1989). Chronometric evidence for central postponement in temporally overlapping tasks. *The Quarterly Journal of Experimental Psychology Section A*, *41*(1), 19–45. https://doi.org/10.1080/14640748908402351 (cit. on p. 35).

Pereira, M., Faivre, N., Iturrate, I., Wirthlin, M., Serafini, L., Martin, S., Desvachez, A., Blanke, O., Van De Ville, D., & Millán, J. d. R. (2020). Disentangling the origins of confidence in speeded perceptual judgments through multimodal imaging. *Proceedings of the National Academy of Sciences*, *117*(15), 8382–8390. https://doi.org/10.1073/pnas.1918335117 (cit. on p. 160).

Pereira, M., Megevand, P., Tan, M. X., Chang, W., Wang, S., Rezai, A., Seeck, M., Corniola, M., Momjian, S., Bernasconi, F., Blanke, O., & Faivre, N. (2021). Evidence accumulation relates to perceptual consciousness and monitoring. *Nature Communications*, *12*(1), 3261. https://doi.org/10.1038/s41467-021-23540-y (cit. on p. 19).

Perquin, M. N., Heed, T., & Kayser, C. (2024, May). Variance (un)explained: Experimental conditions and temporal dependencies explain similarly small proportions of reaction time variability in linear models of perceptual and cognitive tasks. https://doi.org/10.1101/2022.12.22.521656 (cit. on p. 158).

Persaud, N., McLeod, P., & Cowey, A. (2007). Post-decision wagering objectively measures awareness. *Nature Neuroscience*, *10*(2), 257–261. https://doi.org/10.1038/nn1840 (cit. on p. 18).

Plato. (390). Apology. In *Plato:Complete Works* (Cooper, John M., pp. 17–36). Hackett Publishing Company. (Cit. on p. 13).

Pleskac, T. J., & Busemeyer, J. R. (2010). Two-stage dynamic signal detection: A theory of choice, decision time, and confidence. *Psychological Review*, *117*(3), 864–901. https://doi.org/10.1037/a0019737 (cit. on pp. 19, 21, 24).

Possamaï, C.-A., Burle, B., Osma, 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 (cit. on p. 33).

Pourtois, G., Vocat, R., N'Diaye, K., Spinelli, L., Seeck, M., & Vuilleumier, P. (2010). Errors recruit both cognitive and emotional monitoring systems: Simultaneous intracranial recordings in the dorsal anterior cingulate gyrus and amygdala

combined with fMRI. *Neuropsychologia*, *48*(4), 1144–1159. https://doi.org/10.1016/j.neuropsychologia.2009.12.020 (cit. on p. 38).

Pouthas, V., George, N., Poline, J.-B., Pfeuty, M., VandeMoorteele, P.-F., Hugueville, L., Ferrandez, A.-M., Lehéricy, S., LeBihan, D., & Renault, B. (2005). Neural network involved in time perception: An fMRI study comparing long and short interval estimation. *Human Brain Mapping*, *25*(4), 433–441. https://doi.org/10.1002/hbm.20126 (cit. on p. 26).

Purcell, B. A., & Kiani, R. (2016). Neural Mechanisms of Post-error Adjustments of Decision Policy in Parietal Cortex. *Neuron*, *89*(3), 658–671. https://doi.org/10.1016/j.neuron.2015.12.027 (cit. on p. 123).

Purcell, B. A., Weigand, P. K., & Schall, J. D. (2012). Supplementary Eye Field during Visual Search: Salience, Cognitive Control, and Performance Monitoring. *Journal of Neuroscience*, *32*(30), 10273–10285. https://doi.org/10.1523/JNEUROSCI.6386-11.2012 (cit. on p. 41).

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 (cit. on p. 24).

Questienne, L., Van Opstal, F., van Dijck, J.-P., & Gevers, W. (2018). Metacognition and cognitive control: behavioural adaptation requires conflict experience. *Quarterly Journal of Experimental Psychology (2006)*, *71*(2), 411–423. https://doi.org/10.1080/17470218.2016.1251473 (cit. on pp. 124, 159).

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 (cit. on pp. 21, 23, 44, 123).

Rabbitt, P. M. A., & Vyas, S. (1981). Processing a display even after you make a response to it. How perceptual errors can be corrected. *Quarterly Journal of Experimental Psychology*, *33A*, 223–239 (cit. on p. 21).

Rabbitt, P. (1979). How old and young subjects monitor and control responses for accuracy and speed. *British Journal of Psychology*, *70*(2), 305–311. https://doi.org/10.1111/j.2044-8295.1979.tb01687.x (cit. on pp. 34, 123).

Ramdani, C., Hasbroucq, T., & Vidal, F. (2024). Why is there an error negativity on correct trials? A reappraisal. *Neuroscience Letters*, *828*, 137731. https://doi.org/10.1016/j.neulet.2024.137731 (cit. on p. 46).

Ratcliff, R., & McKoon, G. (2008). The Diffusion Decision Model: Theory and Data for Two-Choice Decision Tasks. *Neural computation*, *20*(4), 873–922. https://doi.org/10.1162/neco.2008.12-06-420 (cit. on pp. 19, 32).

Resulaj, A., Kiani, R., Wolpert, D. M., & Shadlen, M. N. (2009). Changing your mind: a computational mechanism of vacillation. *Nature*, *461*(7261), 263–266. https://doi.org/10.1038/nature08275 (cit. on pp. 19, 21).

Ridderinkhof, K. R., Ullsperger, M., Crone, E. A., & Nieuwenhuis, S. (2004). The Role of the Medial Frontal Cortex in Cognitive Control. *Science*, *306*(5695), 443–447. https://doi.org/10.1126/science.1100301 (cit. on p. 22).

Riemer, M., Kubik, V., & Wolbers, T. (2019). The effect of feedback on temporal error monitoring and timing behavior. *Behavioural Brain Research*, *369*, 111929. https://doi.org/10.1016/j.bbr.2019.111929 (cit. on p. 158).

Riès, S., Janssen, N., Burle, B., & Alario, F.-X. (2013). Response-Locked Brain Dynamics of Word Production. *PLOS ONE*, *8*(3), e58197. https://doi.org/10.1371/journal.pone.0058197 (cit. on p. 37).

Riesel, A., Weinberg, A., Endrass, T., Kathmann, N., & Hajcak, G. (2012). Punishment has a lasting impact on error-related brain activity. *Psychophysiology*, *49*(2), 239–247. https://doi.org/10.1111/j.1469-8986.2011.01298.x (cit. on p. 44).

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 (cit. on pp. 21, 23).

Roger, C., Bénar, C. G., Vidal, F., Hasbroucq, T., & Burle, B. (2010). Rostral Cingulate Zone and correct response monitoring: ICA and source localization evidences for the unicity of correct- and error-negativities. *NeuroImage*, *51*(1), 391–403. https://doi.org/10.1016/j.neuroimage.2010.02.005 (cit. on pp. 22, 42, 43).

Rosenbaum, D., Glickman, M., Fleming, S. M., & Usher, M. (2022). The Cognition/Metacognition Trade-Off. *Psychological Science*, *33*(4), 613–628. https://doi.org/10.1177/09567976211043428 (cit. on pp. 24, 169).

Rouault, M., Dayan, P., & Fleming, S. M. (2019). Forming global estimates of self-performance from local confidence. *Nature Communications*, *10*(1), 1141. https://doi.org/10.1038/s41467-019-09075-3 (cit. on p. 158).

Sackett, A. M., Meyvis, T., Nelson, L. D., Converse, B. A., & Sackett, A. L. (2010). You're Having Fun When Time Flies: The Hedonic Consequences of Subjective Time Progression. *Psychological Science*, *21*(1), 111–117. https://doi.org/10.1177/0956797609354832 (cit. on p. 27).

Sackur, J. (2009). L'introspection en psychologie expérimentale. *Revue d'histoire des sciences*, *62*(2), 349–372. https://doi.org/10.3917/rhs.622.0349 (cit. on p. 14).

Sajad, A., Godlove, D. C., & Schall, J. D. (2019). Cortical Microcircuitry of Performance Monitoring. *Nature neuroscience*, *22*(2), 265–274. https://doi.org/10.1038/s41593-018-0309-8 (cit. on pp. 41, 43).

Sanford, A. J. (1970). Rating the speed of a simple reaction. *Psychonomic Science*, *21*(6), 333–334. https://doi.org/10.3758/BF03335809 (cit. on pp. 34, 168).

Scangos, K. W., Aronberg, R., & Stuphorn, V. (2013). Performance monitoring by pre-supplementary and supplementary motor area during an arm movement countermanding task. *Journal of Neurophysiology*, *109*(7), 1928–1939. https://doi.org/10.1152/jn.00688.2012 (cit. on p. 41).

Schaeffner, S., Chevalier, N., Kubota, M., & Karbach, J. (2021). Metacognitive Training. In T. Strobach & J. Karbach (Eds.), *Cognitive Training: An Overview of Features and Applications* (pp. 255–270). Springer International Publishing. https://doi.org/10.1007/978-3-030-39292-5_18 (cit. on p. 133).

Schall, J. D., Stuphorn, V., & Brown, J. W. (2002). Monitoring and Control of Action by the Frontal Lobes. *Neuron*, *36*(2), 309–322. https://doi.org/10.1016/S0896-6273(02)00964-9 (cit. on p. 41).

Scheffers, M. K., & Coles, M. G. H. (2000). Performance monitoring in a confusing world: Error-related brain activity, judgments of response accuracy, and types of errors. *Journal of Experimental Psychology: Human Perception and Performance*, *26*(1), 141–151. https://doi.org/10.1037/0096-1523.26.1.141 (cit. on pp. 22, 44, 45).

Scheffers, M. K., Coles, M. G. H., Bernstein, P., Gehring, W. J., & Donchin, E. (1996). Event-related brain potentials and error-related processing: An analysis of incorrect responses to go and no-go stimuli. *Psychophysiology*, *33*(1), 42–53. https://doi.org/10.1111/j.1469-8986.1996.tb02107.x (cit. on p. 21).

Schlichting, N., Fritz, C., & Zimmermann, E. (2023). Motor variability modulates calibration of precisely timed movements. *iScience*, *26*(7). https://doi.org/10.1016/j.isci.2023.107204 (cit. on p. 31).

Schneider, S. L. (1995). Item difficulty, discrimination, and the confidence-frequency effect in a categorical judgment task. *Organizational Behavior and Human Decision Processes*, *61*(2), 148–167. https://doi.org/10.1006/obhd.1995.1012 (cit. on p. 27).

Schultz, W., Apicella, P., & Ljungberg, T. (1993). Responses of monkey dopamine neurons to reward and conditioned stimuli during successive steps of learning a delayed response task. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *13*(3), 900–913. https://doi.org/10.1523/JNEUROSCI.13-03-00900.1993 (cit. on p. 40).

Schulz, L., Fleming, S. M., & Dayan, P. (2023). Metacognitive computations for information search: Confidence in control. *Psychological Review*, *130*(3), 604–639. https://doi.org/10.1037/rev0000401 (cit. on p. 20).

Schwartz, B. L. (1994). Sources of information in metamemory: Judgments of learning and feelings of knowing. *Psychonomic Bulletin & Review*, *1*(3), 357–375. https://doi.org/10.3758/BF03213977 (cit. on p. 17).

Seabold, S., & Perktold, J. (2010). Statsmodels: Econometric and Statistical Modeling with Python, 92–96. https://doi.org/10.25080/Majora-92bf1922-011 (cit. on pp. 112, 127).

Servant, M., Logan, G. D., Gajdos, T., & Evans, N. J. (2021). An integrated theory of deciding and acting. *Journal of Experimental Psychology. General*, *150*(12), 2435–2454. https://doi.org/10.1037/xge0001063 (cit. on p. 168).

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 (cit. on pp. 33, 160, 168).

Shalgi, S., Barkan, I., & Deouell, L. Y. (2009). On the positive side of error processing: error-awareness positivity revisited. *European Journal of Neuroscience*, *29*(7), 1522–1532. https://doi.org/10.1111/j.1460-9568.2009.06690.x (cit. on p. 45).

Shalgi, S., & Deouell, L. Y. (2012). Is any awareness necessary for an Ne? *Frontiers in Human Neuroscience*, *6*. https://doi.org/10.3389/fnhum.2012.00124 (cit. on p. 45).

Sigman, M., & Dehaene, S. (2005). Parsing a Cognitive Task: A Characterization of the Mind's Bottleneck. *PLOS Biology*, *3*(2), e37. https://doi.org/10.1371/journal.pbio.0030037 (cit. on p. 34).

Silvetti, M., Alexander, W., Verguts, T., & Brown, J. W. (2014). From conflict management to reward-based decision making: Actors and critics in primate medial frontal cortex. *Neuroscience & Biobehavioral Reviews*, *46*, 44–57. https://doi.org/10.1016/j.neubiorev.2013.11.003 (cit. on p. 40).

Silvetti, M., Seurinck, R., & Verguts, T. (2011). Value and Prediction Error in Medial Frontal Cortex: Integrating the Single-Unit and Systems Levels of Analysis. *Frontiers in Human Neuroscience*, *5*. https://doi.org/10.3389/fnhum.2011.00075 (cit. on p. 40).

Simen, P., Balci, F., deSouza, L., Cohen, J. D., & Holmes, P. (2011). A Model of Interval Timing by Neural Integration. *The Journal of Neuroscience*, *31*(25), 9238–9253. https://doi.org/10.1523/JNEUROSCI.3121-10.2011 (cit. on pp. 28, 164).

Skinner, B. F. (1938). *The behavior of organisms: an experimental analysis*. Appleton-Century. (Cit. on p. 15).

Slifkin, A. B., & Newell, K. M. (1999). Noise, information transmission, and force variability. *Journal of Experimental Psychology: Human Perception and Performance*, *25*(3), 837–851. https://doi.org/10.1037/0096-1523.25.3.837 (cit. on p. 132).

Smith, J. G., Harper, D. N., Gittings, D., & Abernethy, D. (2007). The effect of Parkinson's disease on time estimation as a function of stimulus duration range and modality. *Brain and Cognition*, *64*(2), 130–143. https://doi.org/10.1016/j.bandc.2007.01.005 (cit. on p. 26).

Smith, M. C. (1968). Repetition effect and short-term memory. *Journal of Experimental Psychology*, *77*(3), 435–439. https://doi.org/10.1037/h0021293 (cit. on p. 124).

Spieser, L., Servant, M., Hasbroucq, T., & Burle, B. (2017). Beyond decision! Motor contribution to speed–accuracy trade-off in decision-making. *Psychonomic Bulletin & Review*, *24*(3), 950–956. https://doi.org/10.3758/s13423-016-1172-9 (cit. on p. 33).

Stahl, J. (2010). Error detection and the use of internal and external error indicators: An investigation of the first-indicator hypothesis. *International Journal of Psychophysiology*, *77*(1), 43–52. https://doi.org/10.1016/j.ijpsycho.2010.04.005 (cit. on p. 47).

Stanton, J. D., Sebesta, A. J., & Dunlosky, J. (2021). Fostering Metacognition to Support Student Learning and Performance. *CBE—Life Sciences Education*, *20*(2), fe3. https://doi.org/10.1187/cbe.20-12-0289 (cit. on p. 18).

Steinhauser, M., & Yeung, N. (2010). Decision Processes in Human Performance Monitoring. *Journal of Neuroscience*, *30*(46), 15643–15653. https://doi.org/10.1523/JNEUROSCI.1899-10.2010 (cit. on pp. 45, 111).

Stemmer, B., Segalowitz, S. J., Dywan, J., Panisset, M., & Melmed, C. (2007). The error negativity in nonmedicated and medicated patients with Parkinson's disease. *Clinical Neurophysiology*, *118*(6), 1223–1229. https://doi.org/10.1016/j.clinph.2007.02.019 (cit. on p. 40).

Sternberg, S. (1984). Stage models of mental processing and the additive-factor method. *Behavioral and Brain Sciences*, *7*(1), 82–84. https://doi.org/10.1017/S0140525X00026285 (cit. on p. 32).

Stevenson, L. M., & Carlson, R. A. (2020). Consistency, not speed: temporal regularity as a metacognitive cue. *Psychological Research*, *84*(1), 88–98. https://doi.org/10.1007/s00426-018-0973-z (cit. on p. 24).

Stoll, F. M., Fontanier, V., & Procyk, E. (2016). Specific frontal neural dynamics contribute to decisions to check. *Nature Communications*, *7*(1), 11990. https://doi.org/10.1038/ncomms11990 (cit. on p. 18).

Stuphorn, V., Taylor, T. L., & Schall, J. D. (2000). Performance monitoring by the supplementary eye field. *Nature*, *408*(6814), 857–860. https://doi.org/10.1038/35048576 (cit. on p. 41).

Swick, D., & Turken, A. U. (2002). Dissociation between conflict detection and error monitoring in the human anterior cingulate cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(25), 16354–16359. https://doi.org/10.1073/pnas.252521499 (cit. on p. 39).

Tandonnet, C., Burle, B., Vidal, F., & Hasbroucq, T. (2003). The influence of time preparation on motor processes assessed by surface Laplacian estimation. *Clinical Neurophysiology: Official Journal of the International Federation of Clinical Neurophysiology*, *114*(12), 2376–2384. https://doi.org/10.1016/s1388-2457(03)00253-0 (cit. on p. 33).

Tandonnet, C., Burle, B., Vidal, F., & Hasbroucq, T. (2006). Knowing when to respond and the efficiency of the cortical motor command: A Laplacian ERP study. *Brain Research*, *1109*(1), 158–163. https://doi.org/10.1016/j.brainres.2006.06.052 (cit. on p. 33).

Tomassini, A., & Morrone, M. C. (2016). Perceived visual time depends on motor preparation and direction of hand movements. *Scientific Reports*, *6*(1), 27947. https://doi.org/10.1038/srep27947 (cit. on p. 30).

Treisman, M. (1963). Temporal discrimination and the indifference interval. Implications for a model of the "internal clock". *Psychological Monographs*, *77*(13), 1–31. https://doi.org/10.1037/h0093864 (cit. on p. 25).

Ullsperger, M., Danielmeier, C., & Jocham, G. (2014). Neurophysiology of Performance Monitoring and Adaptive Behavior. *Physiological Reviews*, *94*(1), 35–79. https://doi.org/10.1152/physrev.00041.2012 (cit. on pp. 37, 38, 161).

Ullsperger, M., Harsay, H. A., Wessel, J. R., & Ridderinkhof, K. R. (2010). Conscious perception of errors and its relation to the anterior insula. *Brain Structure & Function*, *214*(5), 629–643. https://doi.org/10.1007/s00429-010-0261-1 (cit. on p. 22).

Ullsperger, M., & Von Cramon, D. Y. (2006). The Role of Intact Frontostriatal Circuits in Error Processing. *Journal of Cognitive Neuroscience*, *18*(4), 651–664. https://doi.org/10.1162/jocn.2006.18.4.651 (cit. on p. 40).

Vallat, R. (2018). Pingouin: statistics in Python. *Journal of Open Source Software*, *3*(31), 1026. https://doi.org/10.21105/joss.01026 (cit. on pp. 112, 126).

Valt, C., & Stürmer, B. (2017). On the correct side of performance: Processing of internal and external signals in response speed evaluation. *International Journal of Psychophysiology*, *117*, 26–36. https://doi.org/10.1016/j.ijpsycho.2017.04.005 (cit. on p. 47).

Valt, C., & Stürmer, B. (2021). On the processing of optimal performances: Studying arousal evoked by being correct and fast. *Brain and Behavior*, *11*(6), e02162. https://doi.org/10.1002/brb3.2162 (cit. on pp. 47, 117).

van Gaal, S., Ridderinkhof, K. R., van den Wildenberg, W. P. M., & Lamme, V. A. F. (2009). Dissociating consciousness from inhibitory control: evidence for unconsciously triggered response inhibition in the stop-signal task. *Journal of Experimental Psychology. Human Perception and Performance*, *35*(4), 1129–1139. https://doi.org/10.1037/a0013551 (cit. on p. 23).

van Veen, V., Cohen, J. D., Botvinick, M. M., Stenger, V. A., & Carter, C. S. (2001). Anterior Cingulate Cortex, Conflict Monitoring, and Levels of Processing. *NeuroImage*, *14*(6), 1302–1308. https://doi.org/10.1006/nimg.2001.0923 (cit. on p. 39).

van den Berg, R., Anandalingam, K., Zylberberg, A., Kiani, R., Shadlen, M. N., & Wolpert, D. M. (2016). A common mechanism underlies changes of mind about decisions and confidence (T. E. Behrens, Ed.). *eLife*, *5*, e12192. https://doi.org/10.7554/eLife.12192 (cit. on p. 21).

Van 't Ent, D., & Apkarian, P. (1999). Motoric response inhibition in finger movement and saccadic eye movement: a comparative study. *Clinical Neurophysiology: Official Journal of the International Federation of Clinical Neurophysiology*, *110*(6), 1058–1072. https://doi.org/10.1016/s1388-2457(98)00036-4 (cit. on p. 37).

Vickers, D., & Packer, J. (1982). Effects of alternating set for speed or accuracy on response time, accuracy and confidence in a unidimensional discrimination task. *Acta Psychologica*, *50*(2), 179–197. https://doi.org/10.1016/0001-6918(82)90006-3 (cit. on p. 24).

Vidal, F., Burle, B., Bonnet, M., Grapperon, J., & Hasbroucq, T. (2003). Error negativity on correct trials: a reexamination of available data. *Biological Psychology*, *64*(3), 265–282. https://doi.org/10.1016/S0301-0511(03)00097-8 (cit. on p. 47).

Vidal, F., Hasbroucq, T., Grapperon, J., & Bonnet, M. (2000). Is the 'error negativity' specific to errors? *Biological Psychology*, *51*(2-3), 109–128. https://doi.org/10.1016/S0301-0511(99)00032-0 (cit. on pp. 42, 43).

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 (cit. on p. 32).

Vidal, F., Burle, B., & Hasbroucq, T. (2022). On the Comparison Between the Nc/CRN and the Ne/ERN. *Frontiers in Human Neuroscience*, *15*, 788167. https://doi.org/10.3389/fnhum.2021.788167 (cit. on p. 43).

Vidal, F., Burle, B., Spieser, L., Carbonnell, L., Meckler, C., Casini, L., & Hasbroucq, T. (2015). Linking EEG signals, brain functions and mental operations: Advantages of the Laplacian transformation. *International Journal of Psychophysiology*, *97*(3), 221–232. https://doi.org/10.1016/j.ijpsycho.2015.04.022 (cit. on p. 43).

Vidal, F., Grapperon, J., Bonnet, M., & Hasbroucq, T. (2003). The nature of unilateral motor commands in between-hand choice tasks as revealed by surface Laplacian estimation. *Psychophysiology*, *40*(5), 796–805. https://doi.org/10.1111/1469-8986.00080 (cit. on pp. 110, 111, 168).

Visalli, A., Montefinese, M., Viviani, G., Finos, L., Vallesi, A., & Ambrosini, E. (2024). lmeEEG: Mass linear mixed-effects modeling of EEG data with crossed random effects. *Journal of Neuroscience Methods*, *401*, 109991. https://doi.org/10.1016/j.jneumeth.2023.109991 (cit. on p. 171).

Vocat, R., Pourtois, G., & Vuilleumier, P. (2008). Unavoidable errors: A spatio-temporal analysis of time-course and neural sources of evoked potentials associated with error processing in a speeded task. *Neuropsychologia*, *46*(10), 2545–2555. https://doi.org/10.1016/j.neuropsychologia.2008.04.006 (cit. on pp. 22, 43).

Vocat, R., Pourtois, G., & Vuilleumier, P. (2011). Parametric modulation of error-related ERP components by the magnitude of visuo-motor mismatch. *Neuropsychologia*, *49*(3), 360–367. https://doi.org/10.1016/j.neuropsychologia.2010.12.027 (cit. on p. 46).

Wagenmakers, E.-J., & Brown, S. (2007). On the linear relation between the mean and the standard deviation of a response time distribution. *Psychological Review*, *114*(3), 830–841. https://doi.org/10.1037/0033-295X.114.3.830 (cit. on pp. 32, 132).

Watson, J. B. (1913). Psychology as the behaviorist views it. *Psychological Review*, *20*(2), 158–177. https://doi.org/10.1037/h0074428 (cit. on p. 14).

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 (cit. on pp. 33, 168).

Weindel, G., Gajdos, T., & Burle, B. (2021). The Decisive Role of Non-Decision Time for Interpreting Decision Making Models, 21 (cit. on p. 160).

Wessel, J., Danielmeier, C., & Ullsperger, M. (2011). Error Awareness Revisited: Accumulation of Multimodal Evidence from Central and Autonomic Nervous Systems. *Journal of cognitive neuroscience*, *23*, 3021–36. https://doi.org/10.1162/jocn.2011.21635 (cit. on pp. 23, 45, 119, 123).

Wessel, J. R. (2012). Error awareness and the error-related negativity: evaluating the first decade of evidence. *Frontiers in Human Neuroscience*, *6*. https://doi.org/10.3389/fnhum.2012.00088 (cit. on pp. 23, 45).

Wickelgren, W. A. (1977). Speed-accuracy tradeoff and information processing dynamics. *Acta Psychologica*, *41*(1), 67–85. https://doi.org/10.1016/0001-6918(77)90012-9 (cit. on p. 123).

Wiener, M., Turkeltaub, P., & Coslett, H. B. (2010). The image of time: A voxel-wise meta-analysis. *NeuroImage*, *49*(2), 1728–1740. https://doi.org/10.1016/j.neuroimage.2009.09.064 (cit. on p. 47).

Wokke, M. E., Achoui, D., & Cleeremans, A. (2020). Action information contributes to metacognitive decision-making. *Scientific Reports*, *10*, 3632. https://doi.org/10.1038/s41598-020-60382-y (cit. on p. 19).

Wolpert, D. M., & Kawato, M. (1998). Multiple paired forward and inverse models for motor control. *Neural Networks*, *11*(7), 1317–1329. https://doi.org/10.1016/S0893-6080(98)00066-5 (cit. on p. 168).

Woodman, G. F. (2010). Masked targets trigger event-related potentials indexing shifts of attention but not error detection. *Psychophysiology*, *47*(3), 410–414. https://doi.org/10.1111/j.1469-8986.2009.00948.x (cit. on p. 45).

Wundt, W. (1873). *Principles of Physiological Psychology"*. (Cit. on p. 14).

Yallak, E., & Balcı, F. (2021). Metric error monitoring: Another generalized mechanism for magnitude representations? *Cognition*, *210*, 104532. https://doi.org/10.1016/j.cognition.2020.104532 (cit. on p. 29).

Yallak, E., & Balcı, F. (2022). Metric error monitoring for a cleaner record of timing. *Journal of Experimental Psychology: Human Perception and Performance*, *48*(10), 1130–1136. https://doi.org/10.1037/xhp0001050 (cit. on p. 29).

Yeung, N., Botvinick, M. M., & Cohen, J. D. (2004). The Neural Basis of Error Detection: Conflict Monitoring and the Error-Related Negativity. *Psychological Review*, *111*(4), 931–959. https://doi.org/10.1037/0033-295X.111.4.931 (cit. on p. 39).

Yeung, N., Holroyd, C. B., & Cohen, J. D. (2005). ERP Correlates of Feedback and Reward Processing in the Presence and Absence of Response Choice. *Cerebral Cortex*, *15*(5), 535–544. https://doi.org/10.1093/cercor/bhh153 (cit. on p. 39).

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 (cit. on pp. 16, 23, 111, 119).

Yokosaka, T., Kuroki, S., Nishida, S., & Watanabe, J. (2015). Apparent Time Interval of Visual Stimuli Is Compressed during Fast Hand Movement. *PLoS ONE*, *10*(4), e0124901. https://doi.org/10.1371/journal.pone.0124901 (cit. on p. 30).

Yon, D., Edey, R., Ivry, R. B., & Press, C. (2017). Time on your hands: Perceived duration of sensory events is biased toward concurrent actions. *Journal of Experimental Psychology. General*, *146*(2), 182–193. https://doi.org/10.1037/xge0000254 (cit. on p. 30).

Zheng, Y., Wang, D., Ye, Q., Zou, F., Li, Y., & Kwok, S. C. (2021). Diffusion property and functional connectivity of superior longitudinal fasciculus underpin human metacognition. *Neuropsychologia*, *156*, 107847. https://doi.org/10.1016/j.neuropsychologia.2021.107847 (cit. on p. 20).

Zylberberg, A., Barttfeld, P., & Sigman, M. (2012). The construction of confidence in a perceptual decision. *Frontiers in Integrative Neuroscience*, *6*, 79. https://doi.org/10.3389/fnint.2012.00079 (cit. on p. 24).

APPENDICES

A. Behavioral results of EEG experiment

A.1. Dissociating RT and experimental factors contributions to iRT

We reproduced mixed model analyses of chapter 2 to dissociate RT and biases due to other factors contributions to iRT. For each model, we used iRT as dependent variable and participants as random effect, with both random intercepts and random slopes for each predictor.

The first model with RT, stimulus angle (1 or 7 degrees) and stimulus orientation (clockwise or counterclockwise) showed an effect of RT ($\beta = 0.42, t = 7.3, p < 0.001$) and an effect of orientation ($\beta = 0.021, t = 3.28, p = 0.033$), but no effect of angle ($\beta = -0.001, t = -1.3, p = 0.12$).

The second model with RT and accuracy as predictors showed effects of RT ($\beta = 0.41, t = 7.0, p < 0.001$) and accuracy ($\beta = 0.004, t = 0.85, p = 0.39$), with higher iRTs for correct trials, although no effect of accuracy was observed on RT ($F(1,19) = 0.49, p = 0.49$).

A last model revealed effects of RT($\beta = 0.48, t = 10.27, p < 0.001$) and number of EMG bursts ($\beta = 0.021, t = 4.02, p < 0.001$), with higher iRT for trials with multiple EMG bursts, beyond the effect on RT ($F(2,36) = 127.53, p < 0.001$).

A.2. Dissociating force and experimental factors contributions to iForce

Similar mixed model analyses were done on data of the Force-group, to dissociate force and biases due to other factors contributions to iForce.

The first model showed an effect of force ($\beta = 0.46, t = 9.43, p < 0.001$) and an effect of angle ($\beta = 0.39, t = 3.52, p < 0.001$), but no effect of orientation ($\beta = -0.57, t = -0.85, p = 0.39$). For the same produced force, iForce were higher for larger angles (see Fig.A.1).

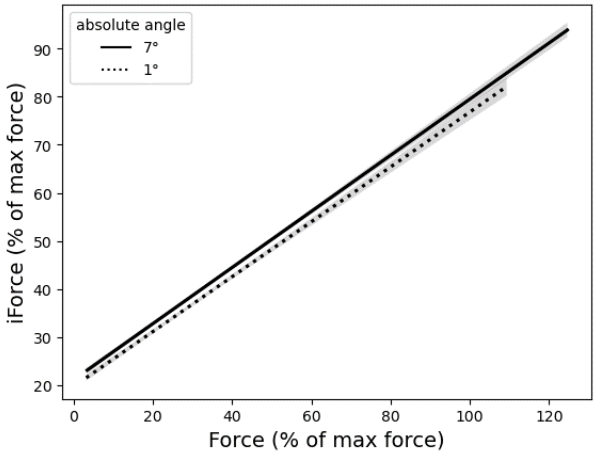


Figure A.1. *Linear regression between trial-by-trial forces and iForces for the two stimulus angles.*

The second model showed effects of force ($\beta = 0.47$, $t = 10.73$, $p < 0.001$) and accuracy ($\beta = 1.8$, $t = 3.32$, $p = 0.001$), with higher iForce for correct trials, beyond the effect of accuracy on force ($F(1,18) = 28.58$, $p < 0.001$).

The last model showed effects of force ($\beta = 0.47$, $t = 10.26$, $p < 0.001$) and number of EMG bursts ($\beta = -1.32$, $t = -3.01$, $p = 0.003$), with lower iForce for trials with multiple EMG bursts, beyond the similar effect of number of EMG bursts on produced force ($F(2,30) = 8.15$, $p = 0.005$).

A.3. Impact of PMT and MT on iRT

To assess to what extent premotor and motor components of RT contribute to iRT with the data of the RT-group of the EEG experiment, we reproduced mixed model comparisons performed in chapter 2. We compared three models with either RT, PMT or PMT+MT as predictors of iRT, with likelihood ratio tests. Participants were included as random effect and we used both random intercept and slopes for each factors. PMT+MT model provided a better fit than PMT model ($\chi^2(4) = 176.66$, $p < 0.001$), with effects of both PMT ($\beta = 0.42$, $t = 7.67$, $p < 0.001$) and MT ($\beta = 0.5$, $t = 5.94$, $p < 0.001$) on iRT. PMT+MT model also provided a better fit than RT model ($\chi^2(4) = 11.41$, $p < 0.02$). We hence completely reproduced findings of chapter 2.

B. Supplementary EEG analyses

B.1. Ne and iForce

We analyzed the relationship between Ne amplitude and iForce in the Force-group. We performed a two-way ANOVA with bin number and dimension, force or iForce, as factors, to assess potential difference of sensitivity between both. There was a significant effect of bin number ($F(4,72) = 9.69$, $p < 0.001$), but no effect of dimension

($F(1,18) = 0.025, p < 0.88$), nor interaction ($F(4,72) = 0.87, p < 0.49$). Ne amplitude increased with both increasing force and iForce (see Fig.B.2).

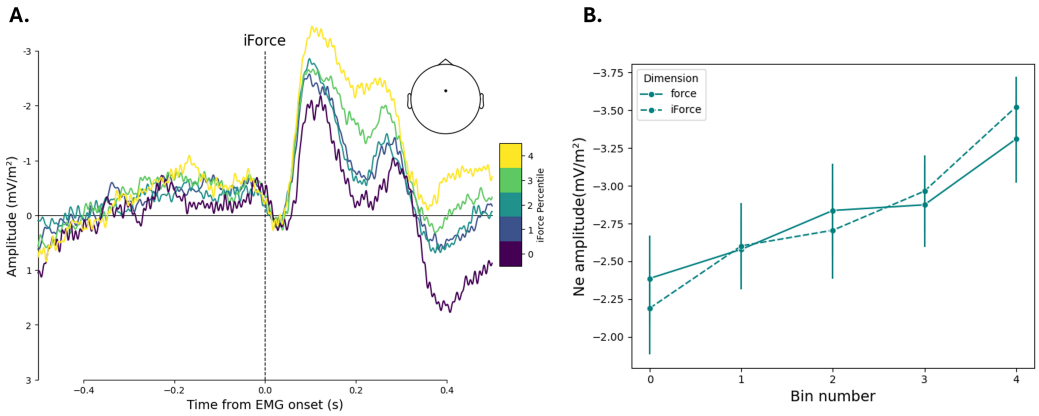


Figure B.2. *(A) Grand averages per iForce bins at electrode FCz. (B) Ne amplitude as a function of force (solid line) and iForce (dashed line) bins.*

B.2. Ne sensitivity to RT as a function of RT estimation ability

To explore whether Ne sensitivity to RT would differ depending on participants' ability to estimate their RT (see Fig.C.5), we divided them into two groups on the basis of their RT/iRT correlation. We performed a median split to obtain a low correlation group (mean r=0.29) and a high correlation group (mean r=0.61).

We performed a mixed ANOVA with high or low correlation group as between factor and RT bin as within factor. Ne amplitude was affected by RT bins ($F(4,72) = 2.76, p = 0.034$), but did not differ between both groups ($F(1,18) = 0.73, p = 0.4$). No interaction was observed between both factors ($F(4,72) = 0.56, p = 0.69$), suggesting a similar sensitivity of the Ne to RT.

Although Fig.C.5 suggests a more flattened pattern between Ne amplitude and RT for the high correlation group, no statistical difference was observed, probably due to a high interindividual variability. In addition, few participants presented low correlations or no correlation, so the corresponding group certainly includes participants with high correlations. Properly doing such analysis would require two have two groups of the same size based on fixed criteria concerning their RT/iRT correlation.

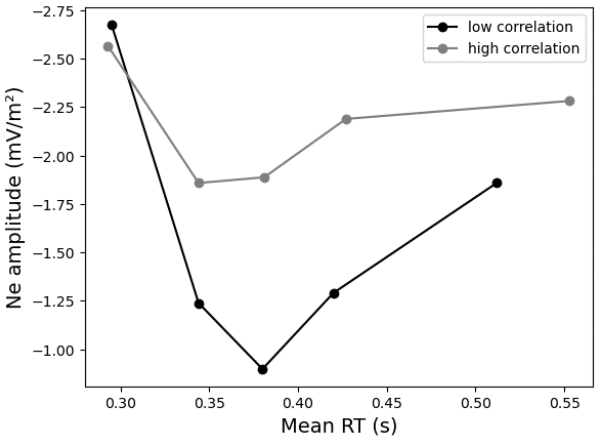


Figure B.3. *Ne amplitude as a function of mean RT per bins for both groups: low (in black) and high (in grey) RT/iRT correlations.*

B.3. Ne trial-by-trial analyses

We attempted to extract Ne amplitude for each individual trials of each participants of the RT-group. To do that, we segmented EEG data on the EMG onset. Each obtained epoch was filtered between 5 and 10 Hz, as the Ne activity lies in this range of frequency. To extract amplitudes, we used onsets and peak indices measured on averages across all trials for each participant, at electrode FCz. We extracted onset and peak positions of each individual epoch by finding the indices of maximum and minimal values in 50-samples windows around participants onsets and peaks. Amplitude was then calculated as the difference between values at peak and values at onset. We could then run a mixed model analysis with individual Ne amplitudes as dependent variable and individuals RT and RT-squared as predictors. We entered participants as random effect, with both random intercepts and slopes for each predictor. With this analysis, we reproduced results of the ERP image analysis of section 3.2. Both RT ($\beta = 0.013, t = 2.38, p = 0.018$) and RT-squared ($\beta = -0.014, t = -2.28, p = 0.023$) predicted Ne amplitude, indicating a non-linear decrease in Ne amplitude with increasing RT, with higher amplitude for both faster and slower RTs (see Fig.B.4).

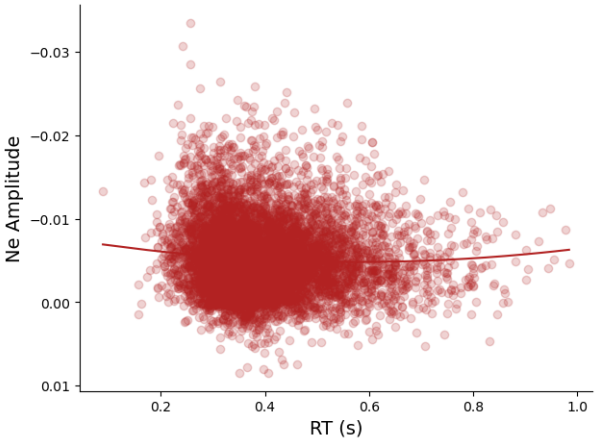


Figure B.4. *Ne amplitudes extracted on individuals filtered epochs as a function of RT with the fitted quadratic function.*

C. Supplementary behavioral analyses

C.1. Dissociation between effects of block on RT and iRT

To investigate whether iRT was sensitive to slow variations of RT, we looked at their evolutions through the experimental blocks. We ran repeated-measure ANOVAs with block numbers as factors. We first did this analysis on the data of the EMG experiment of chapter 2, in which participants performed 12 experimental blocks. Block number had an effect on RT ($F(11,319) = 7.41$, $p < 0.001$), with a slow decrease during the first blocks (see Fig.C.1.A), but no effect on iRT ($F(11,319) = 0.68$, $p = 0.76$). We reproduced these results with data of the EEG experiment of chapter 3, in which participants performed 10 experimental blocks. RT decreased through the blocks ($F(9,171) = 4.38$, $p < 0.001$), but is was not the case for iRT ($F(9,171) = 0.38$, $p = 0.95$, see Fig.C.1.B).

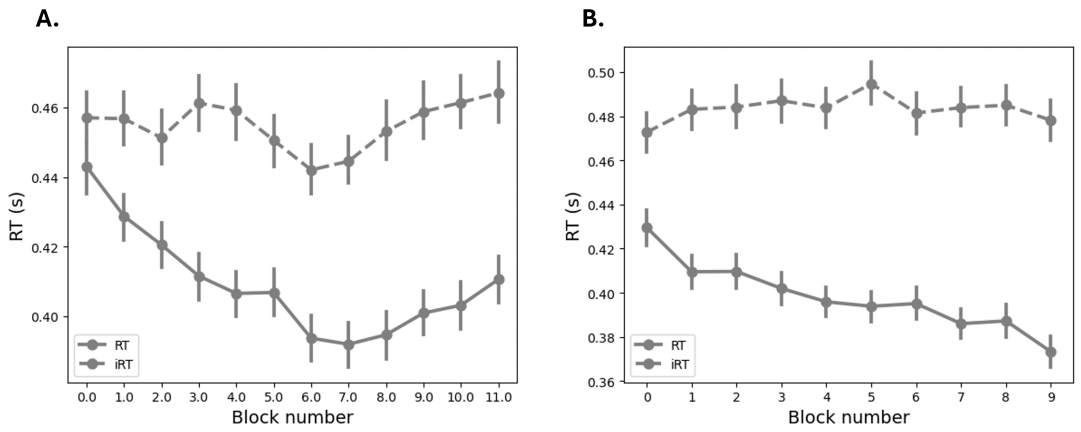


Figure C.5. *RT (solid line) and iRT (dashed line) as a function of blocks in (A) the EMG experiment and (B) the EEG experiment.*

C.2. Individual models with factor effects and timing ability

With data of chapter 2, we ran individual linear models with iRT as dependent variable, and RT, force (weak or strong) and difficulty (indexed by stimulus angle) as predictors, to assess the interindividual variability in the use of non-temporal cues. Nine participants had no additionnal effects of factors, 18 had an effect of one factor (either force or difficulty), and 3 had an effect of both factors. In this experiment, participants performed a small task in which they had to estimate the duration of a visual stimulus. Correlation between objective and estimated durations can be taken as an index of their timing abilities. We hypothesized that participants who had lower timing abilities would rely more on other non temporal cues when estimating their RT. To verify this, we divided participants into 3 groups depending on the number of factors having an effect on iRT beyond RT, and compared their correlation coefficients in the duration estimation task. There was no difference in timing abilities between the different groups ($F(2,27) = 0.23, p = 0.79$, Fig.C.6.B). There was no difference either in their RT/iRT correlations in the RT estimation task ($F(2,27) = 0.4, p = 0.68$, Fig.C.6.C).

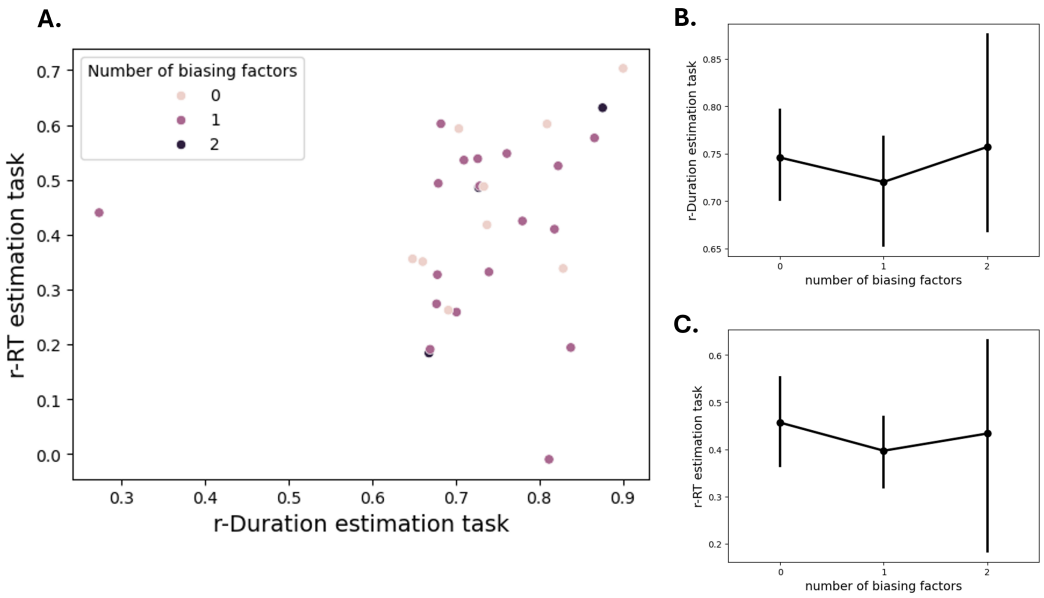


Figure C.6. *(A) Relationship between correlations in the RT estimation task and in the duration estimation task. Participants are color-coded according to their number of biasing factors. (B) Coefficients of correlation in the duration estimation task as a function of number of biasing factors. (C) Coefficients of correlation in the RT estimation task as a function of number of biasing factors.*

It is to be noted that the 3 groups were not well balanced with only 3 participants having an effect of both factors. In addition, the timing task used was really short and the study was not designed for this purpose.

List of Figures

1.1. *Schematic representation of the Nelson and Narens model, based on Nelson (1990).* 16
1.2. Schematic representation of the two-choice Drift Diffusion Model for several trials. The decision process starts from an initial point and evolves over time, with evidence accumulated towards one of two decision boundaries (upper or lower, representing correct and incorrect responses). The time taken to reach a boundary corresponds to the reaction time, including decision and non-decision times (t). The threshold (a) represents the amount of evidence required to make a decision. The drift rate (v) is the average slope of the evidence accumulation process across a large number of trials. 20
1.3. Monopolar grand-averages time-locked to the response, of correct and error trials. The Ne (in red) and the Pe (in green) are highlighted as large negative and positive deflecttions. Adapted from Falkenstein et al. (2000). 22
1.4. Pacemaker-accumulator model proposed by Gibbon et al., 1984. 26
1.5. A schematic presentation of the two-stage temporal error monitoring process. In the first-stage, two independent clocks accumulate temporal evidence, whose threshold first crossing times are then compared with each other to determine the direction of timing error. Upon the threshold crossing of the slower timing process, a second-stage evidence accumulation process starts from the earlier threshold (as illustrated with a horizontal dashed line) crossing point, which eventually leads participants to commit to a confidence judgment, with prolonged accumulation processes resulting in lower certainty (due to larger vertical deviation from starting point). Numbers in circles indicate distinct diffusion processes (shown in red), each used as a comparison clock in different trials. The dotted orange lines depict separate second-level conventional diffusion processes, which are in turn used for confidence ratings. Qualitative model predictions regarding the relationship between confidence ratings and timing error as well as the RTs associated with confidence ratings (RT2) are illustrated on the upper left-hand side of the figure. In the words of Akdoğan and Balcı, 2017. 29
1.6. RT fractioning using EMG. The premotor time (PMT) is the time between the stimulus and the EMG onset; the motor time (MT) is the time between the EMG onset and the mechanical response. 33

1.7. Important brain structures involved in performance monitoring and adaptive behavior, their function, and interaction. pMFC, posterior medial frontal cortex; SMA, supplementary motor area; MCC, midcingulate cortex; ACC, anterior cingulate cortex; dmPFC, dorsomedial prefrontal cortex; vmPFC, ventromedial prefrontal cortex; LPFC, lateral prefrontal cortex; IFJ, inferior frontal junction; IFG, inferior frontal gyrus; OFC, orbitofrontal cortex; m/lOFC, medial/lateral OFC; AIC, anterior insular cortex; fO, frontal operculum; BG, basal ganglia; Cd, caudate nucleus; Put, putamen; GP, globus pallidus; Th, thalamus; Hb, habenula; Mon, mesencephalic and pontine monoaminergic nuclei; Cbl, cerebellum. Main structures evoked in this section are outlined in the red rectangle. Adapted from Ullsperger et al. (2014). 38
1.8. The Predicted Response Outcome (PRO) model. The prediction units (in blue) generate a timed prediction of what outcomes are expected, with what probability, and when. Greater probabilities are associated with greater prediction activity, and the activity peaks at the time the outcome is expected to occur. The evaluation units (in red) compute negative surprise—that is, they detect when an expected outcome fails to occur. Negative surprise is computed simply by subtracting the actual outcome signal from the prediction-unit signal at each point in time. Events that are predicted yield greater surprise signals when they fail to occur (solid line) than when they do occur as expected (dashed line). Outcome units (in green) signal when an actual outcome occurs. An outcome unit is transiently activated at the moment a corresponding outcome actually occurs (dashed line); if no outcome occurs, the outcome unit remains inactive (solid line). Adapted from J. W. Brown (2013). 41
1.9. Laplacian-transformed grand-averages of the Ne for correct trials (red), partial errors (green) and errors (blue). From Roger et al. (2010). 42

3.1. (A) Grand averages per iRT bins at electrode FCz. (B) Ne amplitude as a function of RT (solid line) and iRT (dashed line) bins. (C) Grand averaged ERP image sorted by iRT. Epoch 0 corresponds to smallest iRT. (D) Ne amplitude of each grand averaged 'pseudo trial' is plotted as a function of corresponding iRT with fitted quadratic function. 113
3.2. (A) Grand averages per PMT bins at electrode FCz. (B) Grand averages per MT bins at electrode FCz. (C) Ne amplitude as a function of RT (red), PMT (green) and MT (grey) bins. (D) Ne amplitude plotted as a function of mean RT (red) and mean PMT (green) of each bin. (E) Ne amplitude plotted as a function of mean MT of each bin. 114
3.3. (A) Example of an individual ERP-image sorted by MT (top). Solid black line represents the response (end of MT). ERP obtained by averaging all trials is plotted (bottom). Vertical black line is plotted at averaged MT. (B) Ne peak latency as a function of MT bins. 115

3.4. (A) Grand averages per MT bins at electrode C3. Grey zone represents the mean window of analysis. Topography at mean peak latency of M1 activation is plotted at the bottom left. (B) M1 activation amplitude as a function of MT bins. (C) Grand averages per PMT bins at electrode C3. Grey zone represents the mean window of analysis. Topography at mean peak latency of M1 activation is plotted at the bottom left. (D) M1 activation amplitude as a function of PMT bins. 116
3.5. (A) Grand averages per RT bins (top) and iRT bins (bottom) at electrode Cz. Grey zone represents the window of analysis. (B) Slope of the late activity as a function of RT bins (solid line) and iRT bins (dashed line). (C) Slope of the late activity plotted as a function of mean RT (solid) and mean iRT (dashed) of each bin. 117

4.1. Visual representation of Granger Causality. A time serie x_1 is said to have a causal influence on another time serie x_2, if the past of x_1 ($x_{1,t-n}$) improves the prediction of x_2 ($x_{2,t}$) by its own past ($x_{2,t-n}$) 125
4.2. (A) Cumulative density functions of RTs and forces for both groups, obtained after averaging of vincentized distributions of each participant. (B) Comparison of mean RTs and forces between groups. (C) Comparison of RT and force standard deviations between groups. . . . 127
4.3. (A) Force and (B) RT coefficients of variation of both groups. 128
4.4. Individual trial-by-trial correlations between RT and force in RT-group (A) and Force-group (B). Each regression line represents a participant. (C) individual Pearson's r values of force-group (blue) and RT-group (red). Dashed grey horizontal lines represent critical r values for alpha=0.05 and n=400. 129
4.5. Results of mixed models 4.1 and 4.2 run on data of Experiment 2 (A) and Experiment 3 (B). Grey nodes represent predictors of the models. Left columns represent $trial_{n-1}$ and right columns $trial_n$. Arrows indicate predictors' influence on the dependant variables and are associated with the corresponding coefficient from the model. Green arrows represent positive link and red arrows negative ones. Asterisks are associated with significant predictors. 131

6.1. Different cues contributing to iRT judgment. 161
6.2. We propose that MPFC activity reflects the comparison between a predicted PMT and the actual one. Highest prediction errors for PMTs that deviate from the predicted value lead to larger MPFC activities. 163
6.3. Schematic representation of the proposed mechanism for PMT estimation for two trials. A temporal accumulation runs in parallel with the decision process to inform about PMT deviations from a predicted value. First accumulation process to reach its threshold triggers a second-stage accumulation informing about the size of the deviation. MPFC activity reflects the unsigned deviation from predicted PMT. 166

6.4. Relations between the studied processes and the different measured used.172

A.1. Linear regression between trial-by-trial forces and iForces for the two stimulus angles. 208
B.2. (A) Grand averages per iForce bins at electrode FCz. (B) Ne amplitude as a function of force (solid line) and iForce (dashed line) bins. 209
B.3. Ne amplitude as a function of mean RT per bins for both groups: low (in black) and high (in grey) RT/iRT correlations. 210
B.4. Ne amplitudes extracted on individuals filtered epochs as a function of RT with the fitted quadratic function. 211
C.5. RT (solid line) and iRT (dashed line) as a function of blocks in (A) the EMG experiment and (B) the EEG experiment. 212
C.6. (A) Relationship between correlations in the RT estimation task and in the duration estimation task. Participants are color-coded according to their number of biasing factors. (B) Coefficients of correlation in the duration estimation task as a function of number of biasing factors. (C) Coefficients of correlation in the RT estimation task as a function of number of biasing factors. 213

List of acronyms

2AFC

two-alternative forced choice. 18, 26, 32, 46

ACC

anterior cingulate cortex. 22, 37–46

dACC

dorsal anterior cingulate cortex. 40

DDM

Drift Diffusion Model. 19, 32, 33, 165

DMN

Default Mode Network. 30

dmPFC

dorso-medial prefrontal cortex. 37

EAT

error awareness task. 45

EMG

electromyography. 32, 46

ERP

event-related-potential. 38, 170

FRN

Feedback-Related Negativity. 39, 40

iRT

introspective reaction time. 34–36

LFP

local field potential. 38, 39, 41–43, 46

LIP

lateral intraparietal cortex. 32

MFC

medio-frontal cortex. 45

mPFC

medial prefrontal cortex. 37, 39–41, 161–163

MT

motor time. 33, 159, 160

Nc

Correct Negativity. 43, 44

Ne

Error Negativity. 21–23, 30, 37–47

Pe

Error Positivity. 22, 23, 30, 38, 45

PES

post-error slowing. 23, 34

PFC

prefrontal cortex. 17, 18, 20, 22, 26

PMT

premotor time. 32, 33, 159, 160

pre-SMA

pre-supplementary motor area. 37, 39, 41–44

PRO

Predicted-Response Outcome. 40, 162

PRP

psychological refractory period. 34, 35

RCZ

rostral cingulate zone. 43

RT

reaction time. 23, 24, 31–36, 47

RVPM

Reward Value and Prediction Model. 40

SEF

supplementary eye field. 41–43

SMA

supplementary motor area. 26, 30, 41–45, 47, 163, 167

SMG

supramarginal gyrus. 45

SOA

stimulus onset asynchrony. 34, 35

TEM

Temporal Error Monitoring. 28–30, 36, 164, 165

VAS

visual analog scale. 34, 35, 157