# Representations of feedback in human learning

Wai Ying Chung

Université Paris Cité


École doctorale Cerveau, Cognition, Comportement - ED 158

Centre neuroscience intégrative et cognition - UMR 8002


# Representations of feedback in human learning


**Wai Ying CHUNG**


Thèse de doctorat de Neurosciences

Dirigée par **Florian WASZAK**

Soutenue publiquement le 26 Février 2024

**Membres du jury**

| | | |
|---|---|---|
| Florian Waszak. DR | Université Paris Cité | Directeur de thèse |
| Andrea Kiesel, Professor | Albert-Ludwigs-Universität Freiburg | Rapporteur |
| Marc Buehner, Professor | Cardiff University | Rapporteur |
| Arnaud Badets, CR | Université de Bordeaux | Examinateur |
| Karine Doré-Mazars, Professor | Université Paris Cité | Pésidente du jury / Examinateur |
| Simone Schütz-Bosbach, Professor | Ludwig-Maximilians-Universität München | Examinateur |

# Abstract

## Representation of feedback in human learning

The acquisition of cognitive and motor skills is vital at all stages of our life and it depends critically on feedback we received from the environment, providing information on whether and how improvement can be made. The dominant theoretical framework characterising feedback learning is provided by reinforcement learning. However, the primary focus of the reinforcement learning framework lies in how prediction errors guide learning, emphasising a relatively straightforward aspect of feedback – whether it is better or worse than expected. In reality, the context in which feedback is given and individuals' knowledge of their own actions significantly influence how we extract information from feedback. Furthermore, they play a pivotal role in determining the extent to which this information is utilised to facilitate the learning process.

The objectives of this PhD project are to expand current conceptions of learning from feedback by exploring the rich forms of information that can be extracted from feedback to drive rapid and flexible learning. To achieve these objectives, we have conducted experiments focused on examining how feedback is utilized to improve learning and its application in decision-making, specifically in determining whether to continue learning the same task and when to stop exploration behaviours. This also involves taking into account the impact of individuals' internal estimates of their own performance and the reliability of feedback on the feedback evaluation and decision-making processes. Additionally, we investigated the generation of action-effect predictions in different types of intention-based actions and the role of these predictions in error detection, as well as their impact on the neural processing of feedback and subsequent behavioural adjustments.

Our findings demonstrated the presence of action-effect prediction in intention-based actions, regardless of whether the action choice was based on the selection of action, the timing of action, or the decision to perform or withdraw from an action. This finding, the first of its kind, reveals that while different types of intentional actions may have distinct neurobiological underpinnings, these differences do not significantly influence the learning and prediction of action-associated effects. Having established a robust foundation regarding human ability to anticipate the consequences of

intentional behaviour, we then illustrated that the decision to give up on a task is not solely influenced by the valence of the feedback – whether positive or negative – but is also controlled by an individual's estimation of their own performance. Furthermore, information about feedback reliability significantly impacts how feedback is processed, particularly in terms of its effect on people's confidence in their learning and the adaptation of behavioural strategies.

Altogether, this thesis sheds light on the remarkable flexibility with which humans use feedback to facilitate learning, a process that extends far beyond the simple categorization of feedback as positive or negative. It reveals how individuals integrate a wealth of information about their behaviour and the feedback itself to make informed decisions on how and whether to utilise the feedback information. These findings enhance our understanding of how people learn from feedback in real-life scenarios and underscore the importance of incorporating contextual and individual factors, such as confidence, error awareness, and feedback reliability, into our research. These factors are crucial in determining how the brain processes feedback information and can significantly influence individual learning progress.

Keywords: Feedback processing, Reinforcement learning, Action-effect predictions, Decision making

# Résumé


## Représentations du feedback dans l'apprentissage humain

L'acquisition de compétences cognitives et motrices est vitale à tous les stades de notre vie et dépend de manière critique des retours que nous recevons de l'environnement, fournissant des informations sur si et comment des améliorations peuvent être apportées. Le cadre théorique dominant caractérisant l'apprentissage par retour d'information est fourni par l'apprentissage par renforcement. Cependant, l'accent principal du cadre d'apprentissage par renforcement réside dans la manière dont les erreurs de prédiction guident l'apprentissage, soulignant un aspect relativement simple du retour d'information – s'il est meilleur ou pire que prévu. En réalité, le contexte dans lequel le retour d'information est donné et la connaissance qu'ont les individus de leurs propres actions influencent considérablement la manière dont nous extrayons des informations du retour d'information. De plus, ils jouent un rôle pivot dans la détermination de la mesure dans laquelle cette information est utilisée pour faciliter le processus d'apprentissage.

Les objectifs de ce projet de doctorat sont d'élargir les conceptions actuelles de l'apprentissage à partir du retour d'information en explorant les formes riches d'informations qui peuvent être extraites du retour d'information pour favoriser un apprentissage rapide et flexible. Pour atteindre ces objectifs, nous avons mené des expériences axées sur l'examen de la manière dont le retour d'information est utilisé pour améliorer l'apprentissage et son application dans la prise de décision, spécifiquement dans la détermination de savoir si continuer à apprendre la même tâche et quand arrêter les comportements d'exploration. Cela implique également de prendre en compte l'impact des estimations internes des individus de leur propre performance et la fiabilité du retour d'information sur les processus d'évaluation du retour d'information et de prise de décision. De plus, nous avons étudié la génération de prédictions d'effet d'action dans différents types d'actions basées sur l'intention et le rôle de ces prédictions dans la détection des erreurs, ainsi que leur impact sur le traitement neuronal du retour d'information et les ajustements comportementaux ultérieurs.

Nos découvertes ont démontré la présence de prédiction d'effet d'action dans les actions basées sur l'intention, que le choix de l'action soit basé sur la sélection de l'action, le moment de l'action, ou la décision d'effectuer ou de se retirer d'une action. Cette découverte, la première en son genre, révèle que bien que différents types d'actions intentionnelles puissent avoir des bases neurobiologiques distinctes, ces différences n'influencent pas significativement l'apprentissage et la prédiction des effets associés à l'action. Ayant établi une base solide concernant la capacité humaine à anticiper les conséquences du comportement intentionnel, nous avons ensuite illustré que la décision d'abandonner une tâche n'est pas seulement influencée par la valence du retour d'information – qu'il soit positif ou négatif – mais est également contrôlée par l'estimation qu'a un individu de sa propre performance. De plus, les informations sur la fiabilité du retour d'information influencent considérablement la manière dont le retour d'information est traité, en particulier en termes de son effet sur la confiance des gens dans leur apprentissage et l'adaptation des stratégies comportementales.

Dans l'ensemble, cette thèse met en lumière la remarquable flexibilité avec laquelle les humains utilisent le retour d'information pour faciliter l'apprentissage, un processus qui va bien au-delà de la simple catégorisation du retour d'information comme positif ou négatif. Elle révèle comment les individus intègrent une richesse d'informations sur leur comportement et le retour d'information lui-même pour prendre des décisions éclairées sur comment et si utiliser les informations de retour d'information. Ces découvertes améliorent notre compréhension de la manière dont les gens apprennent à partir du retour d'information dans des scénarios de la vie réelle et soulignent l'importance d'incorporer des facteurs contextuels et individuels, tels que la confiance, la conscience des erreurs, et la fiabilité du retour d'information, dans notre recherche. Ces facteurs sont cruciaux pour déterminer comment le cerveau traite les informations de retour d'information et peuvent influencer de manière significative les progrès individuels dans l'apprentissage.

Mots-clés: Traitement des retours, Apprentissage par renforcement, Prédictions d'effet d'action, Prise de décision

# Résumé substantiel

L'acquisition de compétences cognitives et motrices est essentielle à toutes les étapes de notre vie et dépend de manière critique des retours que nous recevons de l'environnement. Ces retours fournissent des informations sur la possibilité et la manière d'apporter des améliorations. Actuellement, le concept d'apprentissage par renforcement est considéré comme le cadre théorique prédominant pour comprendre l'apprentissage par feedback. Ce cadre se compose essentiellement de modèles computationnels conçus pour capturer les mécanismes d'apprentissage de comportements optimaux à travers des résultats positifs (récompenses) et négatifs (punitions) au fil du temps et dans différents contextes. Bien qu'il puisse exister de nombreuses variantes parmi les différents modèles d'apprentissage par renforcement, le concept central est que l'apprentissage est motivé par l'erreur de prédiction. Plus précisément, la valeur d'un stimulus/action n'est mise à jour que lorsque la récompense reçue ne correspond pas à la récompense prédite. Parmi les nombreux modèles d'apprentissage par renforcement, le temporal-difference (TD) learning model, proposé par Sutton et Barto en 1998, est considéré comme le précurseur de la vision neuroscientifique cognitive moderne de l'apprentissage. Ceci est largement attribué à la découverte que le tir phasique des neurones dopaminergiques du mésencéphale ressemble à un signal très similaire à l'erreur de prédiction temporelle décrite dans le modèle (Bayer & Glimcher, 2005 ; Dayan & Sejnowski, 1996 ; Roesch et al., 2007 ; Schultz et al., 1997).

Sous la formule de TD learning model, la valeur d'un état/événement est prédite non seulement par la récompense immédiate reçue, mais aussi par la somme de toutes les récompenses futures attendues à partir de cet état et vers d'autres états dans le futur. De plus, les récompenses futures sont actualisées en fonction de leur éloignement dans le temps. Ces caractéristiques du TD learning model correspondent bien à ce que nous observons habituellement dans les comportements humains, où les récompenses futures sont considérées comme moins précieuses que les récompenses immédiates, et la valeur d'un état/événement ne repose pas seulement sur la récompense qu'il fournit immédiatement, mais aussi sur son potentiel à nous conduire vers un autre état/événement associé à une récompense plus élevée ou plus faible. De plus, TD learning fournit une explication directe de la manière dont les

humains et les animaux peuvent apprendre l'action optimale ou la séquence d'actions qui nous mène à des états avec des récompenses maximales. Étant donné que la valeur prédictive des états est apprise par TD learning, une personne doit simplement choisir l'action connue pour mener à l'état avec la récompense la plus élevée. Cela nécessite cependant de connaître les conséquences exactes de chaque action en termes d'état résultant. Dans ce cas, nous utilisons la même erreur de prédiction à différence temporelle pour comparer les valeurs de deux états consécutifs après qu'une action ait été choisie. Si l'action a conduit à un état ayant une valeur plus élevée que l'état précédent, cette erreur de prédiction est positive ; si l'état a une valeur inférieure à l'état précédent, alors l'erreur de prédiction est négative.

Le cadre d'apprentissage par renforcement a très bien caractérisé les propriétés de base de l'apprentissage en termes de sa nature motivée par la récompense. Cependant, lorsqu'il s'agit d'appliquer ce modèle à l'apprentissage par feedback chez les humains, certaines limitations deviennent évidentes. Le modèle emploie une approche plutôt simpliste pour caractériser le feedback – positif si la récompense est supérieure aux attentes, négatif lorsque la récompense est inférieure aux attentes. L'influence du feedback sur l'apprentissage est ainsi examinée exclusivement sous cet angle binaire. Bien que ce concept semble raisonnable pour des tâches simples et dans les études sur les animaux, où une récompense principale sert souvent de feedback, il peut ne pas être directement applicable aux humains.

Dans notre vie quotidienne, nous recevons des retours d'information sous diverses formes. Parfois, ils se contentent d'indiquer le résultat comme étant positif ou négatif. D'autres fois, ils peuvent offrir des perspectives sur notre performance et la manière de l'améliorer. De plus, nous recevons des retours internes de nos actions, qui jouent un rôle crucial dans la détection d'erreurs et dans notre performance d'apprentissage. Par conséquent, les informations issues des retours peuvent remplir plusieurs fonctions. Elles peuvent être utilisées pour apprendre la valeur des options comme suggéré dans le modèle d'apprentissage par renforcement, pour améliorer la performance lorsqu'elles contiennent des informations détaillées sur le comportement lui-même, ou même pour guider des décisions pertinentes pour l'apprentissage. Par exemple, nous nous demandons parfois si nous devons persister dans une tâche, chercher des conseils externes ou explorer des stratégies alternatives, et nous

fondons ces décisions sur les retours que nous avons reçus. La manière dont les humains utilisent les retours dans l'apprentissage est flexible et dynamique, variant selon le contenu spécifique du retour et les objectifs de l'individu.

De plus, dans le cadre du modèle d'apprentissage par renforcement, nous mettons à jour nos connaissances sur l'association stimuli/actions – résultat en fonction de la différence entre les résultats attendus et réels. Par conséquent, la valeur de l'association actions – résultat devrait augmenter, et l'action devrait être répétée lorsque la récompense est meilleure que prévu, et diminuer lorsque la récompense est pire que prévu. Cependant, ce qui semble être négligé dans le modèle est la mesure dans laquelle la valeur devrait augmenter ou diminuer pour chaque résultat positif et négatif. L'incertitude peut être le facteur qui contrôle le degré de mise à jour de la valeur du choix de comportement et impacte les ajustements futurs du comportement. Dans le monde réel, l'incertitude est un facteur constant, surtout en ce qui concerne les relations entre les stimuli, les actions et leurs résultats. Cette incertitude peut provenir de diverses sources : elle peut être due à la variabilité inhérente de l'environnement (Behrens et al., 2007; McGuire et al., 2014), à la nature des retours que nous recevons (Di Gregorio et al., 2019; Schiffer et al., 2017; Walsh & Anderson, 2011), aux erreurs dans l'exécution des actions (Akdoğan & Balcı, 2017; Gehring et al., 2018; Kononowicz & Van Wassenhove, 2019; McDougle et al., 2019), ou même au manque de précision dans nos prédictions internes sur l'effet sensoriel de nos actions (Frömer et al., 2021). La présence d'incertitude joue un rôle pivot dans la détermination de la perception de l'informativité des retours, de leur interprétation et de la probabilité de tout ajustement de comportement ultérieur basé sur eux.

Par conséquent, si nous souhaitons obtenir une image plus cohérente de la manière dont les humains utilisent les retours d'information pour améliorer l'apprentissage dans des scénarios de la vie réelle, il est essentiel de prendre en compte tous les facteurs mentionnés ci-dessus dans notre recherche et d'étudier comment leurs effets peuvent varier de manière prévisible d'un individu à l'autre. Dans ce projet de doctorat, nous avons exploré l'utilisation flexible des retours dans l'apprentissage et la prise de décision chez les humains. Nous avons mené une enquête systématique sur la manière dont divers facteurs contextuels et individuels, y compris la fiabilité des retours, la confiance subjective dans la progression de

l'apprentissage, la performance objective et la capacité d'un individu à prédire les résultats des actions, influencent le traitement et l'interprétation des retours, et l'effet des retours sur l'ajustement comportemental ultérieur.

Nous avons mené trois expériences au cours de ce projet. Dans Experiment 1, nous avons étudié la génération de prédictions d'effet-action sous des actions intentionnelles. Comme mentionné précédemment, la capacité de prédire les conséquences sensorielles de notre action est cruciale pour la détection des erreurs et l'affinement de la performance motrice. Nous avons reconnu que la recherche antérieure traitait souvent les actions intentionnelles comme un concept unifié, malgré le fait qu'elles peuvent être catégorisées en trois types principaux, en fonction des décisions internes concernant le fait d'agir, quelle action effectuer et quand la réaliser. En utilisant l'EEG, nous avons mesuré si la prédiction des effets d'action était générée indépendamment du fait que l'effet soit lié au choix de l'action, au moment de l'action, ou à la décision de réaliser ou de se retirer d'une action. Plus précisément, nous avons accédé à la présence de prédictions d'effet-action indirectement en observant la réponse d'erreur de prédiction dans les données EEG lorsqu'un effet-action attendu est violé. Nos découvertes ont révélé que la prédiction d'effet-action se produit dans tous les trois types de décisions d'action. Ce résultat suggère que, malgré que des études de neuroimagerie antérieures (Krieghoff et al., 2009; Mueller et al., 2007) indiquaient qu'il existe des mécanismes neurobiologiques distincts sous-jacents à différents types d'actions intentionnelles, ces différences ne semblent pas influencer de manière significative le processus de formation des associations action – résultat. Les résultats de cette expérience enrichissent notre compréhension de la capacité humaine à prédire les conséquences de leurs propres actions. En même temps, cela ouvre la porte à des enquêtes plus détaillées sur la manière dont les prédictions issues de différents aspects de l'action peuvent interagir entre elles. Par exemple, le résultat de nos actions est généralement prédit conjointement par le choix de l'action et le moment de son exécution.

Dans Experiment 2, nous avons étudié comment les gens utilisent les retours pour prendre des décisions concernant l'apprentissage, en particulier lorsqu'ils décident de poursuivre ou d'abandonner une tâche. Nous reconnaissons que les retours, qu'ils soient positifs ou négatifs, peuvent influencer les individus à persister

ou à arrêter l'apprentissage. Alors que les retours positifs peuvent motiver à continuer l'engagement dans une tâche, l'effort nécessaire pour de nouvelles tentatives sur la même tâche peut conduire à une décision de ne pas continuer. En revanche, les retours négatifs peuvent décourager de nouvelles tentatives, mais la perspective d'une récompense future possible peut motiver la persévérance. En combinant des mesures comportementales avec l'EEG, nous avons testé si la décision de continuer ou d'abandonner l'apprentissage peut être prédite par la valence des retours, les signatures neuronales du traitement des retours en EEG (FRN and P3), et la performance objective des participants. Plus précisément, nous avons conçu une tâche sensori-motrice où les participants devaient reproduire le temps total de présentation de stimuli visuels en maintenant enfoncée une touche du clavier, puis recevaient des retours positifs ou négatifs pour leur réponse. Par la suite, ils devaient décider de retenter ou non le même essai, où ils pouvaient recevoir une récompense ou une pénalité en fonction de leur performance.

Nous avons constaté que la décision de retenter une tâche dépendait significativement du retour reçu. Comme prévu, les retours positifs encourageaient la répétition du comportement. De manière importante, la performance objective des participants a également eu un impact significatif sur la décision. Les gens étaient capables de prendre leur décision en fonction de leur performance : si la performance était bonne, ils étaient enclins à réessayer ; si ce n'était pas le cas, ils étaient plus susceptibles de passer à l'essai suivant. Puisque les retours dans cette tâche ne signalaient pas la performance, cela suggère que les participants étaient capables d'estimer leur performance de manière interne et d'appliquer cette information dans la décision de retenter ou non la tâche. De manière intéressante, l'impact de la performance sur la décision se reflétait dans l'amplitude du FRN, un marqueur neuronal typique pour l'erreur de prédiction en apprentissage. Plus l'amplitude du FRN est grande, plus les participants sont susceptibles de prendre leur décision, qu'il s'agisse de réessayer l'essai actuel ou de passer au suivant, conformément à leur performance. Une grande amplitude du FRN pourrait indirectement indiquer une prédiction de résultat générée à partir d'un suivi interne du comportement. Cela est dû au fait que la présence d'une attente est supposée déclencher une plus grande réponse neuronale face au résultat réel, comparativement à une situation où aucune attente ne peut être formée (Holroyd et al., 2009 ; Hsu et al., 2015 ; Wurm et al., 2022).

De plus, nous avons observé qu'une augmentation de l'amplitude du P3 prédisait la décision de réessayer malgré une mauvaise performance, ce qui peut refléter une mise à jour du modèle cognitif et une augmentation de la valeur motivationnelle (Nieuwenhuis et al., 2005 ; Polich, 2007). Nos découvertes dans cette étude sont précieuses pour montrer que la décision de répéter un comportement ou de continuer à apprendre ne repose pas uniquement sur le fait que les retours reçus soient positifs ou négatifs. Les gens sont capables de générer une estimation interne de leur performance motrice et de l'utiliser pour optimiser leurs décisions.

Dans Experiment 3, nous avons étudié comment la fiabilité des retours, ainsi que leur nature positive ou négative, affecte les taux d'apprentissage individuels et les adaptations comportementales. De plus, nous avons obtenu une mesure subjective des évaluations de confiance des individus concernant leur progression d'apprentissage durant la tâche et avons évalué son impact sur l'évaluation des retours. Nous avons employé une tâche d'apprentissage probabiliste où les participants se fiaient aux retours pour améliorer progressivement leur performance et pour décider de continuer à explorer ou de s'engager dans leur choix comportemental afin de gagner des récompenses potentielles. De manière cruciale, nous avons manipulé la fiabilité des retours à travers différents blocs expérimentaux, la fixant à un niveau élevé (autour de 80 %) et bas (autour de 70 %). Les participants étaient informés de la fiabilité des retours au début de chaque bloc. De plus, ils étaient invités à évaluer leur niveau de confiance dans leur progression d'apprentissage au début de chaque essai de chaque bloc.

Les résultats de cette étude sont intrigants. Nous avons découvert que les retours n'affectaient les évaluations de confiance des participants que lorsqu'ils n'avaient pas encore démontré un apprentissage réussi. Cependant, une fois que la performance est restée constamment bonne, les retours n'influençaient plus leurs évaluations de confiance. Leur confiance était alignée sur leur performance, restant relativement élevée lorsque la performance était bonne, qu'ils reçoivent des retours positifs ou négatifs. Alors que les retours influençaient significativement la confiance pendant les périodes de mauvaise performance, cet effet était médiatisé par la fiabilité des retours. Les participants évaluaient leur confiance plus hautement après avoir reçu des retours positifs d'un bloc de haute fiabilité de retours comparé au même

retour dans un bloc de faible fiabilité. Cette ligne de résultats est cohérente avec des études précédentes qui ont montré une réduction de la réponse neuronale aux retours dans les dernières étapes d'une tâche d'apprentissage (Bellebaum et Daum 2008; Eppinger et al., 2008; Hajcak et al., 2007; Krigolson et al., 2009; Pietschmann et al., 2008). De plus, d'autres études ont trouvé que la réduction de l'information de retour est corrélée à une amélioration de la performance d'apprentissage (Frömer et al., 2021; Sewell et al., 2018). Nos résultats démontrent en outre que la réduction des retours est également évidente dans une mesure explicite de la confiance du sujet.

Dans cette étude, nous avons également analysé la tendance à l'exploration. Nous avons quantifié les comportements exploratoires par le degré de changement de réponse entre les essais et avons trouvé que, en général, l'évaluation de la confiance est un bon prédicteur du degré d'ajustement de la réponse. Typiquement, une confiance plus faible entraînait de plus grands changements. Intéressant, bien que nous ayons trouvé que la performance influençait également le degré de changement de réponse (avec un degré de changement plus petit lorsque la réponse précédente était exacte), cet effet n'était observable que dans les blocs de haute fiabilité de retours. Cela pourrait suggérer que les participants étaient généralement moins conscients de l'exactitude de leurs réponses dans des conditions de faible fiabilité de retours, et cette incertitude pourrait en même temps encourager un comportement exploratoire (Cavanagh et al., 2012; Nassar et al., 2010, 2016). Concernant la décision de s'engager dans une réponse pour gagner une récompense potentielle, nous avons trouvé que les participants étaient plus enclins à le faire après avoir reçu des retours positifs ou lorsque leur évaluation de confiance était élevée. Intéressant, la probabilité de s'engager dans une réponse était plus élevée dans les blocs de faible fiabilité de retours par rapport aux blocs de haute fiabilité. Nous soupçonnons que cette tendance peut représenter un comportement exploratoire, où les participants choisissent de vérifier s'ils ont reçu des retours précis plus tôt, même si cela comporte un coût mineur étant donné les chances limitées de gagner une récompense. Nos découvertes démontrent la flexibilité de la manière dont les apprenants humains utilisent les retours pour ajuster l'apprentissage, où l'impact des retours sur l'ajustement comportemental se réduit progressivement à mesure que l'apprentissage progresse, et la connaissance de la fiabilité des retours est prise en compte dans le traitement de l'information de retour.

Pour résumer, l'objectif de cette thèse est de fournir une meilleure compréhension de la nature complexe de l'apprentissage par feedback chez les humains. À travers notre travail, nous avons souligné l'importance d'intégrer des facteurs qui régissent le niveau d'incertitude dans les informations dans l'étude de l'apprentissage humain. De plus, nous avons démontré la capacité des apprenants humains à intégrer des informations provenant de diverses sources (par exemple, l'estimation interne de l'effet-action, la confiance subjective dans la performance, la connaissance de la crédibilité du feedback) pour les aider à interpréter les retours qu'ils reçoivent et décider comment appliquer cette information pour optimiser dynamiquement la performance. Nos découvertes sont précieuses pour contribuer au développement d'un cadre d'apprentissage plus efficace qui reflète avec précision la dynamique complexe des environnements d'apprentissage du monde réel.

# Acknowledgments

First of all, I would like to thank Florian Waszak for being my PhD supervisor. I am grateful for the trust and confidence you have shown in me over the years. Thank you for being such a comforting and reassuring figure whenever I felt stressed or lost. I know I can be difficult to talk to or to reach out to sometimes, and I appreciate your patience in this regard. You are one of the most caring people I know, as you always try your best to ensure the well-being of everyone. I have learned a lot from you, both professionally and personally. I hope I will be more like you in the future and inspire trust in the people around me.

I would also like to thanks Álvaro Darriba for the support he shows in my works during every moment in this PhD project. I couldn’t have managed and get to this point without you. It was a real pleasure to work together and we should definitely keep it going in the future too!

The greatest thanks also go to all the people who helped me get to where I am today. Hermann, Simone, Jakob, without the opportunities these individuals have given me, I simply wouldn't have made it here.

Also, the deepest thanks to all the friends I made in this lab, Rongrong, Hanna, Weiwei, Léa, Diane, Lucie, Robert. All of you have made the laboratory life awesome. Thanks also to my family and friends back home for their unconditional support in whatever I do in my life.

Lastly, I would like to extend my heartfelt thanks to all the member of the jury for having accepted to review the present work: Simone Schütz-Bosbach, Marc Buehner, Andrea Kiesel, Arnaud Badets and Karine Doré-Mazars.

# Table of Contents

# Table of illustrations

# List of abbreviations

ACC: Anterior cingulate cortex

CS: Conditioned stimuli

EEG: Electroencephalography

ERP: Event-related potentials

fMRI: Functional magnetic resonance imaging

FRN: Feedback-Related Negativity

ICA: Independent component analysis

MMN: Mismatch negativity

PE: Prediction error

SNc: Substantia nigra pars compacta

TD: Temporal-difference

US: Unconditioned stimuli

VTA: Ventral tegmental area

# 1. INTRODUCTION

## 1.1. Background

The acquisition of cognitive and motor skills is vital at all stages of our life and critically depends on the feedback we receive from the environment. Feedback provides information on whether and how improvements can be made. Currently, the dominant theoretical framework characterizing feedback learning is reinforcement learning. According to this framework, learning critically depends on prediction error (PE), which concerns whether outcomes are better or worse than expected. While the reinforcement learning framework has proven to be a powerful tool for understanding human learning — as the existence and use of prediction errors in driving learning have been repeatedly documented in research over decades at both behavioural and neurophysiological levels (Daw & Doya, 2006; Pessiglione et al., 2006; Schultz et al., 1997). However, in real-life situations, using and interpreting feedback involves more than just the knowledge of whether it is better or worse than expected.

Human learners possess the ability to integrate information from various sources, to help us interpret feedback and decide how to apply it to optimize performance. One key factor that significantly influences feedback processing is – the feeling of uncertainty. In our daily lives, we all consistently experience some level of uncertainty as there is seldom a deterministic relationship between stimuli/actions and outcomes. The level of uncertainty we experience conditioned the informativeness of feedback. Take environmental uncertainty as an example: in an environment that is consistently changing, unexpected feedback holds greater informational value as it is more likely to indicate the occurrence of real changes in the context. In contrast, in a stable environment, unexpected outcomes are more likely to be seen as exceptions and less likely to prompt any behavioural changes (Behrens et al., 2007). Another dimension of uncertainty is linked to our understanding of the feedback's properties, such as its reliability. Feedback reliability can be assessed based on experiences, explicit instructions, or the credibility of the source providing the feedback (e.g., the trustworthiness of the person offering the feedback). Previous studies have shown that the perceived reliability of feedback has a direct impact on the learning rate and on the modification of behavioural strategies in individuals (Carlebach & Yeung, 2023; Pescetelli et al., 2021; Schiffer et al., 2017).

In addition, humans possess metacognitive abilities, allowing us to consistently perform second-order evaluations regarding the accuracy of our estimations about the external world (i.e. what have we learned ?) (Meyniel et al., 2015; Yeung & Summerfield, 2012). The outcome of this evaluation is expressed as *confidence*. Confidence is a rational measure of performance during the learning process, as it has been reported that subjective confidence increases linearly with objective performance (Bounmy et al., 2023; Meyniel, Schlunegger, et al., 2015), and it has also been found to share similar neural markers with the processes of error detection and error monitoring (Boldt & Yeung, 2015; Yeung & Summerfield, 2012a). In the context of learning, the estimation of confidence modulates the effect of prediction error/surprise on learning adjustments. Smaller updates occur after a surprising outcome when confidence is high, conversely, larger updates occur when confidence is low (Meyniel, 2020; Meyniel & Dehaene, 2017). This modulation effect has been demonstrated in a variety of studies across different task settings, e.g., probabilistic learning task with both visual and auditory stimuli (Meyniel, 2020; Meyniel, Schlunegger, et al., 2015), value-based decision making (E. Payzan-LeNestour et al., 2013) and motor learning (Frömer et al., 2021). Therefore, subjective confidence, along with the general volatility of the external context, conditions the effect of feedback on learning, beyond simply the degree of error indicated by the feedback.

Furthermore, the processing of external feedback in motor learning is relatively underexplored in the reinforcement learning literature, as the tasks typically employed are often related to learning the underlying probabilities of events through feedback and observations. Motor learning tasks are more complex because they involve a credit assignment problem: how do we determine if the absence of a reward reflects an extrinsic property of the environment, an incorrect estimation of task parameters due to an insufficient amount of samples, or an intrinsic error in motor execution? Previous studies have demonstrated that humans can estimate the magnitude of motor errors reasonably accurately (Akdoğan & Balcı, 2017; Kononowicz et al., 2019; Kononowicz & Van Wassenhove, 2019). It has also been shown that when a motor execution error is detected, it is discounted in decision-making (McDougle et al., 2016), and the prediction error response is attenuated when an unexpected outcome is associated with an execution error compared to when the execution is successful (McDougle et al., 2019). These findings indicate that the impact of an observed

outcome on updating learning and decision-making can be modulated by internal motor feedback.

Altogether, while the framework of reinforcement learning provides a strong foundation for understanding how feedback drives learning through prediction error. We are in need for a more complex framework that would take into account of the impact of uncertainty (whether if it is from the environment, the feedback itself or from our internal estimation of self-performance) and to understand in what way the effect of feedback in learning is affect by the level of uncertainty. Such a framework will eventually help us to better translate laboratory findings into real-world applications.

## 1.2. Thesis objectives

The objective of this PhD project is to address the limitations of the reinforcement model, where value estimates are updated solely based on prediction errors. We aim to demonstrate that human learners can dynamically and optimally adjust their use of feedback in learning, based on information beyond prediction error, such as knowledge about the quality of the feedback and individuals' internal estimation of performance. Meanwhile, we are interested in exploring the impact of feedback on subsequent decision-making, specifically regarding whether to continue or give up on a task, and the decision to explore. These decisions are critical in the context of learning, as the ultimate goal of learning is to maximize potential rewards over the long term. To achieve this, learners must consistently use all available information to decide whether the effort required for further attempts is justified by the prospect of a possible future reward, and whether to explore other possible options, even at the expense of temporarily choosing less rewarding actions. We believe that the information from feedback plays a significant role in this decision-making process.

To address these objectives, we conducted three experiments during the period of this project. In our first experiment, we examined the human ability to generate predictions about their own action effects across different types of actions, establishing the basics of human ability to monitor their action outcomes. In our second experiment, we explored the impact of feedback and motor performance on the decision to either

give up or continue in a motor learning task. This was the first time any study has investigated the interactive effect between internal motor feedback and external feedback in terms of the decision to give up or continue in a motor learning task. This study addresses previous findings that internal monitoring of motor performance affects decision-making and the effect of outcomes/feedback on learning. (Frömer et al., 2021; McDougle et al., 2016, 2019).We expected that external feedback would significantly impact the decision to continue learning, but the effect of the feedback would be modulated by the internal monitoring of motor performance.
In the third experiment, we investigated how uncertainty regarding the feedback itself (by controlling feedback reliability) and subjective confidence regarding learning performance modulate the effect of feedback in learning. Additionally, we explored how the decision to continue exploring other possible options or remain with the chosen option is made in relation to feedback and uncertainty (about the quality of the feedback itself and about learning performance). Altogether, our findings will shed light on the role of feedback in learning adjustment and decision-making, and to gain a better understanding of how we use feedback to aid learning and decision-making in a flexible way. This involves taking into account other available information in the environment and our internal estimation of the context.

## 1.3. Structure of the thesis

The first part of this thesis offers an overview of the relevant literature for this PhD project, divided into two main sections. In the first section, we outline the basic concepts of reinforcement learning and present results from neuroimaging studies that support the reinforcement learning model. This provides readers with insights into how the reinforcement learning framework, initially a computational approach in machine learning, has become the leading framework for understanding human learning in cognitive neuroscience today. In the second section, we focus specifically on feedback learning. We discuss the neural basis of feedback processing and provide a detailed account of the different sources of uncertainty we consistently encounter in our daily lives. Furthermore, we describe how each source impacts the way feedback is interpreted and used in learning and decision-making at both the behavioural and neural levels, supported by previous findings.

In the second part of the thesis, we present the work conducted throughout this PhD project. We will present the results from three experiments. In the first experiment, we investigated the formation of action-effect predictions in intentional actions. Previous studies have repeatedly demonstrated that the execution of different types of actions—the selection of the action (what), the timing of the action (when), and the decision to perform or withdraw from an action (whether)—can be partly dissociated in different brain regions (Brass & Haggard, 2008; Krieghoff et al., 2009; Kühn & Brass, 2010; Mueller et al., 2007; Zapparoli et al., 2017). However, no study has yet investigated the formation of action-effect predictions separately in these different types of action. The results of this study inform us whether action-effect predictions are generated under all three types of action and whether the strength of those predictions may differ depending on the type of action choice. After an in-depth examination of the human ability to monitor their action outcomes and learn the associations between their action and the sensory effects its caused. In our second experiment, we examined the role of external feedback and internal motor feedback in the decision of whether to give up or continue learning, using a sensorimotor task. It is well-established in the literature that external feedback can help improving motor learning performance, but less is known about the extent to which the decision to reattempt a task or not depends on the received feedback. Additionally, how this decision may also be informed by the internal estimations of motor performance, especially if we decorrelate the performance from the external feedback. The result of this study would illustrate the role of external feedback in decision-making, as well as the possible interaction effect between the external and internal motor feedback on both learning and decision-making. In the third experiment, we investigated how uncertainty regarding the feedback itself, along with subjective confidence about learning performance, modulates the effect of feedback on response adjustment and the decision to explore. We employed a learning task where participants had to rely on the feedback received on every trial to gradually improve their performance while we induced different levels of feedback reliability in the task. This approach allowed us to examine the degree of learning adjustment from one trial to the next, with varying levels of confidence and feedback reliability. Additionally, participants had the option to decide to stop sampling evidence and just remain with their chosen option, and we examined how this decision is made in relation to feedback, feedback reliability, and confidence.

In the final part of the thesis, we discuss the experimental results obtained from the three experiments and offer insights into potential directions for future research.

# 2. LITERATURE REVIEW

## 2.1. What is reinforcement learning?

Imagine driving to the supermarket every day. If you always go from home to the same supermarket, the activity becomes largely effortless. Although there may be an overwhelming number of sensory cues in the environment, your responses to those cues are well-established: turn right at Saint-Germain Street, make a left turn when you see the bakery, etc. Based on the received cue, you execute a pre-programmed response without the need to think about it or pay any attention to it. But how did you learn to execute this specific response? Why this response? Out of all the possible actions you could take (like scratching your nose or waving your hand), you learned to turn right at Saint-Germain Street.

Formal research into conditioned responses can be traced back to the early 19th century, marked by the seminal works of Pavlov (1927) and Thorndike (1911). In the typical setup of what is now termed classical or Pavlovian conditioning, a conditioned stimulus (e.g., the sound of a bell) is paired with an unconditioned stimulus (e.g., food). The unconditioned stimulus typically signals a primary reward and elicits an observable biological reaction, referred to as the unconditioned response. The most intriguing observation within this framework is that, after repeated pairings of the conditioned stimulus with the unconditioned stimulus, the unconditioned response begins to occur ahead of time, triggered solely by the presentation of the conditioned stimulus. This phenomenon occurs because the organism now has learned to *predict* the reward upon the presentation of a previously neutral cue. This framework illustrates how an organism learns to predict events controlled by the environment through cues or the context.

Nonetheless, as active agents in the world, we are capable of using actions to change the environment with the goal of maximising reward. Using the same example as earlier, food does not generally appear magically after the ringing of a bell, animals

in the wild has to go out and look for food. In this case, animals need to learn to produce the right behaviour in the given context in order to achieve their goals. This learning process was first formulated and described by Thorndike (1905, 1911, 1913) under the theory of the 'Law of Effect', which essentially states that behaviours followed by pleasant consequences are likely to be repeated, while those followed by unpleasant consequences are less likely to be repeated. One of his most famous experiments involved placing cats in a puzzle box, from which they could escape by triggering a mechanism, such as pulling a string or pressing a lever. Initially, the cats would move randomly and aimlessly, but over time, they learned to associate the specific action with escape and the subsequent reward. This observation demonstrates how animals learn and adapt their behaviours based on the outcomes they experienced. These data obtained from behavioural experiments in psychology have provoked great interest among researchers in computer science. They were inspired to build compact computational models that can capture this mechanism of learning optimal behaviours through positive and negative outcomes over time, and under a variety of contexts. These models are the basis for machine learning, and nowadays they are also powerful tools for understanding human learning mechanism within the field of neuroscience. Due to the discovery that the computational learning models are not only capable of predicting behaviours in animal and human but also parallel the actual neural processes observed in primates. Specifically, the dopaminergic neurons activities in the midbrain regions (will be described in more detail in later paragraphs).

In the following section, I would provide examples of reinforcement learning model. Although the primary focus of this thesis is not on the computational aspect of reinforcement learning (for details, see Sutton & Barto, 2018) and it is also impossible to cover the exhaustive amount of learning models that exist to date. Two models, however, deserve special mention because they are widely considered to be the progenitors of the modern cognitive neuroscientific view of learning.

### 2.1.1. Reinforcement learning model

Rescorla and Wagner (1972) described a model for learning the predictive value of conditioned stimuli (CS) in the case of Pavlovian conditioning. Here, the update rule for the value of the stimulus $CS_i$ is according to

$$V_{new}(CS_i) = \alpha\beta[\lambda(US) - \sum_i V_{old}(CS_i)]$$

The critical part of the Rescorla-Wagner model is that learning/the update of the value of the CSs is driven by prediction error – the discrepancy between what was predicted $\Sigma_i V(CS_i)$ (the sum of all the predictive value of the CSs presents in the trial) and what actually happened $\lambda(US)$. $\alpha$ and $\beta$ are learning rates related to the conditional and unconditional stimuli. While this model explains how the updating of stimulus values in an experimental trial could take place, there are two major shortcomings that make it difficult to apply to real-life learning. Firstly, the model treats conditioned and unconditioned stimuli as separate entities. This created the problem of it not being able to account for second-order conditioning, where if stimulus A predicts a primary reward and stimulus B predicts stimulus A, then stimulus B should also gain predictive value. This is exactly how monetary outcomes can be seen as reward to human leaners. Secondly, the Rescorla-Wagner model fails to account for the temporal evolution of events. The model explains the learning of the value for a series of conditional stimuli in a 'trial', without specifying the temporal relations between the conditioned and the unconditioned stimuli within the trial, or the effects of those temporal relations may have on the value of the stimuli.

To address those limitations, Sutton and Barto (1998) proposed a model named – Temporal-difference (TD) learning. There are many variants of the TD learning model but the basis formula is as follow:

$$V(S_t) = V(S_t) + \alpha[r_{t+1} + \gamma V(S_{t+1}) - V(S_t)]$$

This formula is somewhat similar to the Rescorla-Wagner model in that it still relies on prediction error to update learning. However, it is now sensitive to the temporal structure of events within a trial. In TD learning, the value of a state/event $V(S_t)$ is predicted not only by the immediate reward received but also by the sum of all future rewards expected to be obtained, starting from that state and to other states in the future. However, these future rewards are discounted with the factor $\gamma$ based on how distant they are in the future. The power of TD learning lies in the fact that even without any knowledge of the true reward probabilities in the environment, simply by recursively updating the value of the immediate reward received from the current state,

and the estimated rewards for the next state to which the current state leads to, the value estimates for each state gradually converge to their true values. The true value of a state is the expected sum of return when starting from that state.

Simply by understanding the basic formula, we can already see how TD learning fits better with the learning behaviours of humans and animals that we observe in real life. Rewards in the future are usually seen as less valuable than immediate rewards, and the value of a state/event is not only based on the reward it provides immediately but also on its potential in leading us toward another state/event that is associated with a higher or lower reward. Again, we are not passive agents in the world. In some cases, it is true that we cannot do much more than just observe the transition between events/states one after another and try to learn the reward probability associated with each state and the transition probability between states through our experience and observation. However, most of the time, we are capable of taking action, and we can control the transition of states with our actions. Then, the question arises: how do we learn the action or action sequence (also termed the 'policy') that leads us to states with maximum rewards?

The rule of TD learning can also be applied in this regard, with only minor modifications. Given that the predictive value of states is learned through TD learning, a person simply needs to select the action that is known to lead to the state with the highest reward. However, this requires knowing the exact consequences of each action in terms of the resulting state. Just like we usually do not have any idea about the true reward probabilities in an environment, we also need to learn the value of each action performed in each state – essentially, the value of all possible state-action pairs. In this case, we use the same temporal-difference prediction error $\gamma V(S_{t+1}) - V(S_t)$, to compare the values of two consecutive states after an action is chosen. If the action has led to a state with a higher value than the previous state, this prediction error is positive, if the state has a lower value than the previous state, then the prediction error is negative. Just as previously described in the Law of Effect (Thorndike 1905; 1911; 1913), the likelihood of performing actions that lead to positive outcomes should increase, while the probability of taking actions that lead to negative outcomes should decrease. Consequently, the agent can learn an explicit policy of how to act (e.g., always choosing the action that leads to the highest probability of

reward) by creating a probability distribution over all available actions at each state, assuming there has been sufficient exploration of all state-action pairs.

While it is impressive how complex learning problems can be so elegantly solved by relatively simple computational models, the real interest in reinforcement learning models within cognitive neuroscience began to flourish for a specific reason. This interest was sparked when researchers discovered that the phasic firing of midbrain dopaminergic neurons resembles a signal very similar to the reward prediction error described in reinforcement learning models (Dayan & Sejnowski, 1996; Schultz et al., 1997). In the following section, we will review the neural data on reinforcement learning, from single-unit recordings of dopaminergic neuron activity to whole-brain imaging studies using fMRI.

### 2.1.2. Neural data on reinforcement learning

Dopamine has long been linked with reward processing, as evidenced by studies showing that blocking dopamine receptors reduces pleasure in response to previously rewarding stimuli in humans and animals (Wise, 2004; Wise et al., 1978). The pioneering work of Schultz et al. (1997) highlighted a direct correlation between the activity of dopaminergic neurons in the ventral tegmental area (VTA) of monkeys' midbrains and the prediction error signal described in the reinforcement models using a simple conditioning task. In the experiment, monkeys received a sip of juice as the unconditioned stimulus (US), and they learned to associate a tone or light (CS) with this reward. As depicted in Figure 1, before the association between the CS and the US was established, the occurrence of the unpredicted US elicited a positive prediction error. Once the monkey learned to predict the occurrence of the US by the CS, the positive prediction error signal shifted to the moment of the CS presentation. Conversely, when the US, anticipated by the CS, was omitted, a negative prediction error was recorded at the expected delivery time of the US. These characteristics of the dopaminergic neurons closely resemble what would be expected of a prediction error signal and have since been replicated in many studies (Bayer et al., 2007; Bayer & Glimcher, 2005; Fiorillo et al., 2003; Tobler et al., 2003).

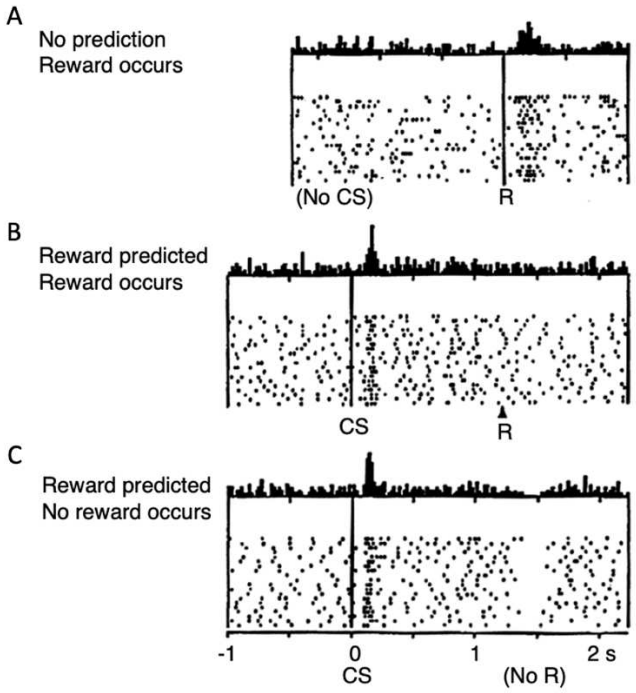



Figure 1. Dopamine activity under conditioning.
(A). Before learning, dopaminergic neurons respond with a phasic burst of firing to the US. (B). After the occurrence Us is associated with a cue, the phasic response now accompanies the presentation of the predictive cue. (C). In "catch" trials, in which the food reward was unexpectedly omitted, dopaminergic neurons showed a below baseline firing rate at the time of the expected reward delivery. Adapted from Schultz, Dayan & Montague (1997).

In later studies, it has been further demonstrated that the firing pattern of midbrain dopaminergic neurons also fit well with the characteristics of temporal-difference prediction errors, where the prediction error signal is calculated based on not only the immediate reward but also the expected rewards in the future, with delayed rewards being discounted in their value. In a study of Bayer and Glimcher (2005), they trained monkeys to perform a saccade task where the amount of reward varied depending on the timing of making a saccade to the target. The longer the monkeys waited before making a saccade, the more juice they received, up to a certain deadline. This task design enabled the monkeys to learn the optimal timing to make a saccade to maximise their rewards. In this context, a wide range of prediction errors were observed during the learning process. They found that the dopaminergic neurons firing rates for the current reward were predicted by the difference between the value of the current reward and the weighted average of previous rewards in the last ten trials. Notably, when the current reward's value fell below the weighted average of past rewards, the dopamine neurons exhibited no response to the present reward. Also, in Roesch et al. (2007), they trained rats to perform an odor-discrimination task, where the rats learned that a specific odor cue can predict either a short or long delay reward.

Their results showed that the cue-evoked firing activity of VTA dopaminergic neurons was significantly lower for the cue that predict long-delay reward compared to short-delay reward, even though the objective reward value was the same (Figure 2A). This effect was also mirrored behaviorally, with rats showing a preference to go to location that offer a shorter delay reward over a long-delay reward when given a choice (Figure 2B).

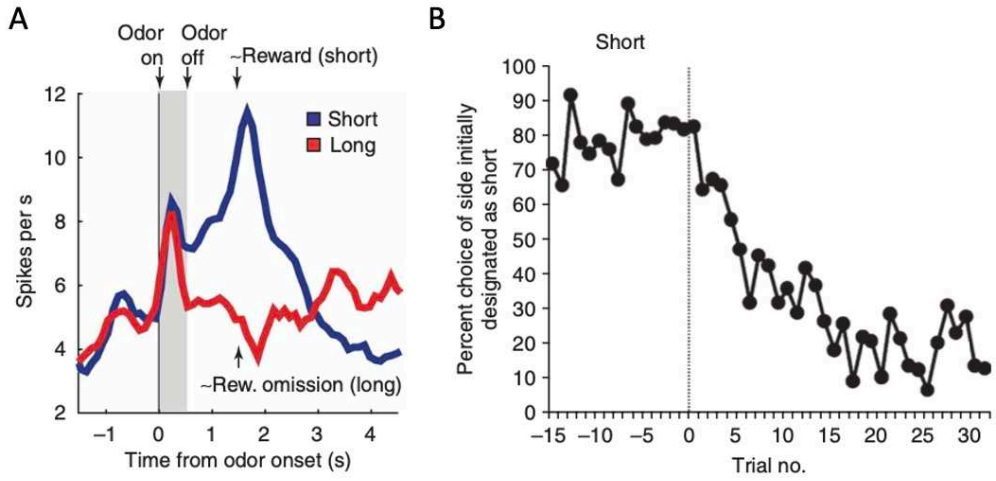



Figure 2. The impact of delay reward in dopaminergic neurons activity and behavioural choice.
(A). Average dopaminergic neurons activity of short and long delay reward delivery (B). Choice behavior before and after the reward switch from short to long delay. Adapted from Roesch et al. (2007).

So far, we have mainly focused on neural data collected from animals using single-unit recording. To study the impact of dopaminergic neuron activity on human learning and decision-making, nonetheless, would require techniques that are much less invasive. fMRI is the common choice of technique for human studies. Although it may not provide information as precise as single-unit recording, nor offer such excellent temporal resolution, it has the advantages of being non-invasive and capable of recording activity at the whole-brain level.

Using fMRI, previous studies have identified several pathways of midbrain dopaminergic neuron projection (Figure 3). Dopaminergic neurons from the substantia nigra pars compacta (SNc) primarily target the striatum, while those from the VTA were mainly projected to the prefrontal cortical areas, including the nucleus accumbens, dorsal, ventral lateral prefrontal cortex, and orbitofrontal cortex (Berns et al., 2001; Knutson et al., 2001; Pagnoni et al., 2002). The pathway between the SNc and

striatum (termed the nigrostriatal pathway) is found to be primarily involved in motor control (Deumens et al., 2002; Rodríguez et al., 2000). The projections from the VTA to the nucleus accumbens and prefrontal cortex regions, constituting the mesocorticolimbic system. This system is believed to be responsible for the evaluation and updating of reward/motivational values associated with different action options (Björklund & Dunnett, 2007; Daw et al., 2011; Joel et al., 2002). Altogether, the functions of the dopaminergic neuron pathways are believed to be crucial for reward-seeking behaviour, guiding the selection of actions based on their reward value (Arias-Carrión et al., 2010; Pessiglione et al., 2006; Schönberg et al., 2007).

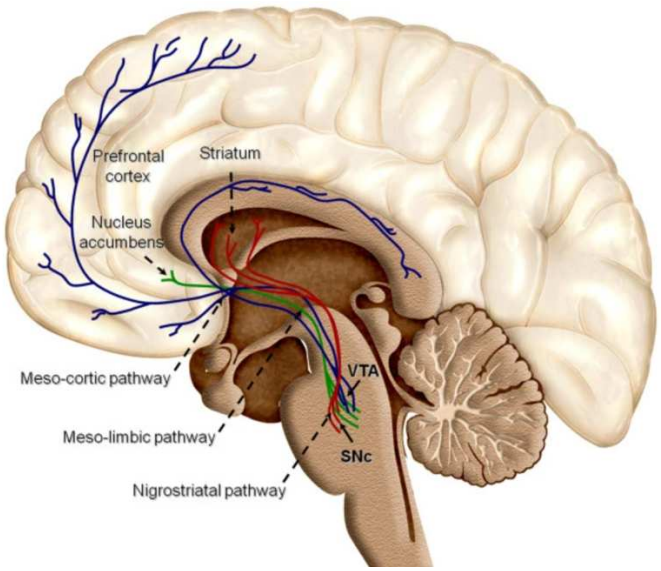



Figure 3. Pathways of dopamine projection.
Dopaminergic neurons are primary located in the midbrain structures substantia nigra (SNc) and the ventral tegmental area (VTA). Their axons project to the striatum, the dorsal and ventral prefrontal cortex. Adapted from Arisa-Carrión et al. (2010).

### 2.1.3. Summary

In summary, the integration of reinforcement learning models with modern neuroscience techniques has markedly enhanced our grasp of learning and decision-making processes in human and animal. The virtue of computational models is that all the parameters and computational processes required are explicated stated. This clarity has been instrumental in driving forward neuroscience research, allowing for a more sophisticated, model-based analysis of complex neural data.

## 2.2. Feedback learning

When considering how feedback is represented within the reinforcement learning framework, the concept is quite straightforward: any reward that exceeds expectations equates to positive feedback, while any situation where no reward is received or the reward is less than expected equates to negative feedback. Learning is then updated in proportion to the degree of mismatch between the received feedback and the expectation. This framework provides a solid foundation for understanding the basics of how feedback can drive learning through prediction error. However, the effect of feedback on learning is not solely modulated by the degree of prediction error, especially in the case of human learners. In this chapter, we aim to provide an overview of how the sense of uncertainty shapes learning. We will first briefly review the neural correlates of feedback processing. Then, we will discuss several forms of certainty that are commonly present in our daily lives, including external uncertainty that comes from the stochastic nature of the environment and internal uncertainties that are related to the state of knowledge and the precision of motor execution. We will focus on the impact of uncertainties on learning and the modulation effect they have on the processing of feedback, as well as subsequent behavioral decision-making, especially regarding the decision between exploration and exploitation.

### 2.2.1. Neural processing of feedback

The study of the neural correlates of feedback processing is primarily based on findings from EEG studies. The high temporal resolution of EEG makes it an ideal technique for capturing the brain's rapid response to feedback events and tracking the neural dynamics that occur during the process of evaluating and responding to feedback. The Feedback-Related Negativity (FRN) and P3 (or P300) ERP components have been identified as crucial neural markers in feedback processing (Holroyd et al., 2003; Holroyd & Coles, 2002; Nieuwenhuis et al., 2004; Yeung & Sanfey, 2004).

The FRN, typically emerging about 200-300 milliseconds after feedback presentation, is observed as a negative deflection in the ERP signal and is most prominently detected at frontocentral scalp sites (Figure 4). The amplitude of the FRN,

believed to be linked to the brain's reward prediction error signal, is sensitive to the valence of the feedback. It is larger for negative feedback compared to positive feedback (Holroyd et al., 2003; Nieuwenhuis et al., 2004; Williams et al., 2021; Yeung & Sanfey, 2004, but see Holroyd et al., 2008). It is believed to reflect the brain's rapid evaluation of the outcome as worse than expected and also acts as a signal for increased cognitive control (Cohen et al., 2011; van de Vijver et al., 2011; Van Der Helden et al., 2010). The anterior cingulate cortex (ACC) is identified as the most likely neural generator of the FRN, given the frontocentral location of it, and it is also found to be more sensitive to negative than to positive outcomes in neuroimaging studies of reward processing (Holroyd & Coles, 2002; Knutson et al., 2000). On the other hand, the P3 component, also known as the P300, manifests as a positive ERP deflection typically occurring around 300-600 milliseconds post-feedback, often observed at parietal scalp locations. The P3 is associated with the allocation of attentional resources (Karayanidis et al., 2000; Luck & Kappenman, 2012) and the processing of the motivational significance of an event (Briggs & Martin, 2009; Carrillo-de-la-Peña & Cadaveira, 2000; Franken et al., 2011). In the context of feedback processing, studies have shown that the amplitude of the P3 component is larger when the feedback is more unexpected, carries higher motivational significance, or is considered to be more task-relevant (Donaldson et al., 2016; Walentowska et al., 2016). The P3 component is also believed to reflect the updating of cognitive models and behavioural strategies based on the feedback information (Chase et al., 2011; Schiffer et al., 2017). The generators of the P3 include a broad network of brain regions, encompassing the medial temporal and subcortical structures (such as the hippocampus, amygdala, and thalamus) and the lateral prefrontal cortex (Nieuwenhuis et al., 2005; Polich, 2007). In addition to the ERPs, neural oscillations have also been suggested to reflect specific aspects of feedback processing (Cohen et al., 2011). Theta band activity (around 4-8 Hz) at the frontocentral electrodes has been suggested to signal the need for cognitive control, showing a significant decrease in power for negative feedback compared to positive feedback(Cavanagh, Zambrano-Vazquez, et al., 2012; Cohen, 2011). Furthermore, the degree of this power desynchronization is found to be related to behavioural adaptation, being correlated with learning success in the subsequent trial (Cavanagh, Figueroa, et al., 2012; Cohen, 2011).

To summarise, EEG studies have identified the Feedback-Related Negativity (FRN) and P3 components as key neural markers of feedback processing. The FRN signals the brain's response to feedback valence and prediction error, and its amplitude is linked to the increase in cognitive control. Meanwhile, the P3 component is associated with the allocation of attention and the processing of motivational significance, as well as the updating of the cognitive model, which may later lead to adjustments in behavioural strategies. Furthermore, theta band activity within neural oscillations has been observed to signal the need for cognitive control and correlates with subsequent behavioural successes.

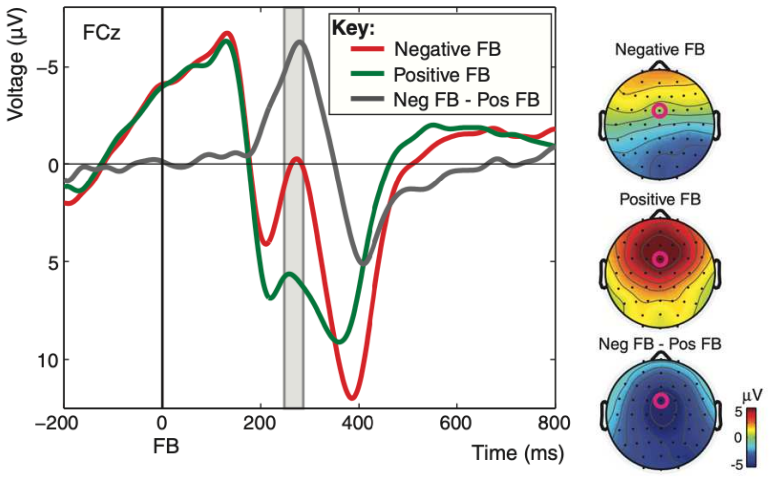



Figure 4. Example of the FRN waveform.
A typical FRN waveform which displayed lager amplitude in response to negative feedback compared to positive feedback. Data from Fcz is commonly used for the analyse of the FRN due to its effect being the most prominent in the frontal-central region. Adapted from Cohen et al. (2011).

### 2.2.2. The impact of uncertainty in learning

In our daily experiences of interacting with the environment, outcomes do not always have a deterministic relationship with the stimuli and actions that precede them, owing to the inherent unpredictability and complexity of the real world. Under the framework of the reinforcement learning model, we update our knowledge about the stimuli/actions – outcome association based on the difference between expected and actual outcomes. Consequently, the value of the actions – outcome association should increase, and the action should be repeated when the reward is better than expected, and decrease when the reward is worse than expected. However, what seems to be overlooked in the model is the extent to which the value should increase or decrease for each positive and negative outcome. Uncertainty may be the factor that controls

the degree of updating of the value of the behaviour choice and impacts future behaviour adjustments. The general feeling of uncertainty can come from many different sources, such as the statistical regularities of the reward environment, the top-down knowledge of whether the feedback is reliable or not, it also depended on our ability to have a more or less precise prediction on our action execution and the sensory consequence of the action.

In a study by Behrens et al. (2007), they demonstrated how the activation of the anterior cingulate cortex (ACC) and, behaviourally, the learning rate, change depending on whether the reward probability of the environment was stable or volatile. In their task, participants had to choose between a blue or green card, with only the correct choice being rewarded. During the stable phase, the probability of the blue card being the correct choice was 75% and remained stable. In the volatile phase, reward probabilities switched between 80% blue and 80% green every 30 or 40 trials. It is assumed that in a more volatile environment, the learning rate should be higher because it would be more important to keep track of every recent outcome in order to adapt behaviour flexibly. However, in a stable environment, the learning rate should be lower because behaviour adjustment is expected to be based on experiences from a more extended period of time. Their results showed that the average learning rate of the participants was significantly higher during the volatile phase, and the BOLD signal of the ACC was significantly larger during the trial outcomes monitoring period in the volatile phase, where the outcome is believed to have a greater influence on future actions (Figure 5). The role of the ACC in learning is believed to be related to the consistent monitoring of action outcomes, to represent and update decision values (Holroyd & Yeung, 2012). This finding highlights the impact of environmental volatility on the evaluation of feedback and showing that the updating of action value is significantly control by the environmental context in which the feedback is given.

Subsequent studies have also obtained similar findings when environmental volatility is manipulated in their experimental design (McGuire et al., 2014; Nassar et al., 2010, 2019; Schiffer et al., 2017). In Schiffer et al. (2017), EEG activity was measured during feedback processing in a probabilistic learning task where participants needed to learn the correct mapping between two images and two response keys using the feedback they received on every trial, which was correct 75%

of the time. In the first experiment, environmental volatility was manipulated by providing instructions at the beginning of each experimental block about whether the rule was likely to remain stable or likely to change within the block.

Rule reversals occurred in two-thirds of the blocks that were instructed as likely to change and in one-third of the blocks that were instructed as stable. A significantly larger FRN amplitude was observed in blocks labelled as volatile, in contrast to those marked as stable. Moreover, participants demonstrated quicker behavioural adjustments following a rule reversal in blocks with the volatile instruction. Similar to Behrens et al. (2007), the authors suggested that the increased FRN amplitude reflects the higher informativeness of feedback under a volatile environment compared to a stable environment. In a second experiment using the same task, the focus shifted to controlling feedback reliability instead of environmental volatility. Blocks were given either reliable or unreliable feedback instructions. In half of the blocks with unreliable feedback instruction, the feedback reliability was 62.5%, and in half of the blocks with reliable feedback instruction, the feedback reliability was 87.5%. The rest of the blocks had a feedback reliability of 75% but came with different instructions (either reliable or unreliable feedback). The results once again showed that the FRN amplitude was larger for feedbacks believed to be more informative, which occurred in the reliable feedback blocks, and this effect was evident even when the objective feedback reliability was the same but only accompanied by different instructions (for similar findings, see Di Gregorio et al., 2019). Altogether, these findings suggest a flexible learning system that integrates the environmental regularities and external information about feedback reliability (when available) to guide adaptive behaviour.

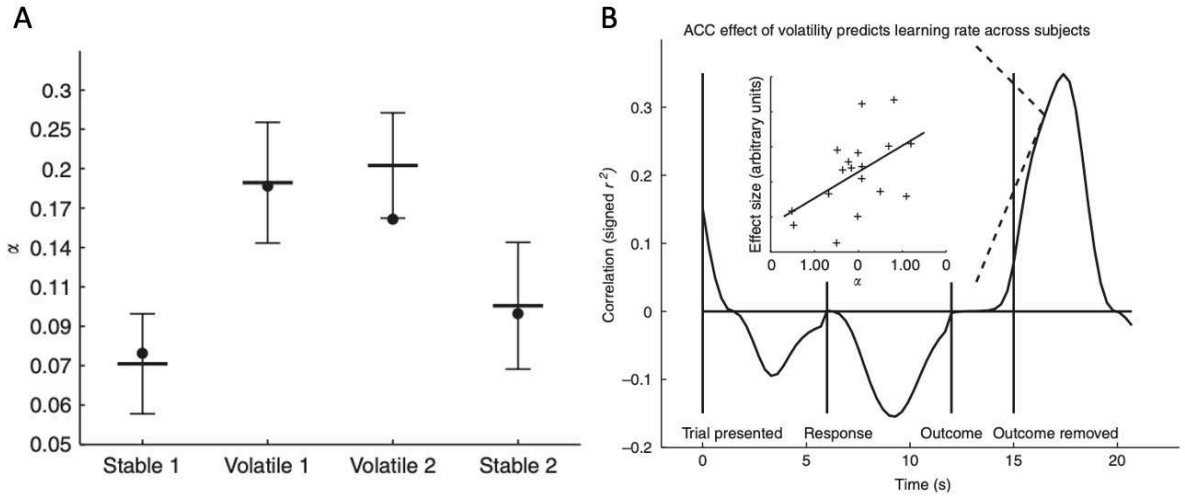



Figure 5. The impact of environmental volatility on the processing of behaviour outcome.
(A). Average learning rates during the stable and volatile phases of the experiment (stable-first and volatile-first, respectively). (B). Volatility related activity in the ACC and a time series of correlations (signed r2) between the effect size in the ACC and the mean learning rate of the subjects. Adapted from Behrens et al. (2007).

### 2.2.3. Action-effect prediction

While external information is used to help us adjust our learning, we also rely on the internal representation of our actions to improve learning, especially in motor tasks. The two most prominent frameworks that describe the mechanisms behind the internal monitoring of one's action consequences are the forward model (Wolpert et al., 1995, 2011) and the ideomotor principle (Greenwald, 1970; Prinz, 1990, 1997). In the forward model, it is suggested that the actual effect of an action is systematically compared with an internal prediction of the action-effect, which is generated by an efference copy of the motor command (Figure 6). By comparing the actual sensory feedback with the internal prediction, any sensory effect that was predicted is cancelled out, and any discrepancies generate a prediction error that can be used to adjust motor commands accordingly. This model is commonly used to explain how one is able to distinguish action-effects that are self-generated or caused by others, and the phenomenon of one being less sensitive to self-caused effects, e.g., not being able to tickle oneself (Blakemore et al., 2000; Wolpert & Flanagan, 2016). On the other hand, the ideomotor principle posits that action is selected based on the anticipation of its sensory consequences and that actions and their sensory effects share a common representation in the brain once the association is formed (Hommel et al., 2001; Prinz, 1990). In this way, we guide action selection by the intended effects we

aim to produce in the environment, and a prediction error is generated if the actual effect is not as intended.

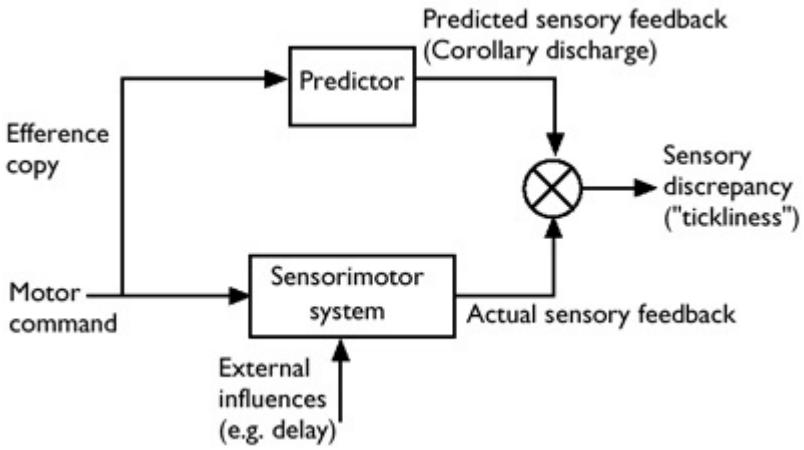



Figure 6. The forward model of action control.
The actual sensory feedback of an action is compared with the internal prediction generated from the motor command. Adapted from Blakemore, Wolpert & Frith (2000).

Both theories have received substantial empirical support from studies, and it is likely that both mechanisms are responsible for guiding action selection and the learning of action-effect associations depending on the situation. The crucial point here is that having an internal model of our action outcomes (both the sensory input in the body and a higher-order prediction regarding the effect it will cause in the environment) allows us to interpret feedback more optimally, extracting useful information for adjusting learning while discounting what may not be informative. Imagine that the same feedback can be caused by many different sources; sometimes, we may have a wrong idea about the correct action to take, but at other times, we may know exactly what to do but still make an execution error. In the former case, the feedback will be useful because it may indicate the correct move, but in the latter case, the feedback should be discounted. Since people are usually aware of their response errors and able to report them with reasonable accuracy (Akdoğan & Balcı, 2017; Kononowicz & Van Wassenhove, 2019; Maier & Steinhauser, 2013; Riesel et al., 2013; Yeung & Summerfield, 2012). It is possible to obtain an explicit measure of the internal representation of action-effect by simply asking individuals to estimate their motor errors. Furthermore, we can examine the precision of one's action-effect predictions using a subjective confidence measure on their error estimation (Meyniel, Sigman, et al., 2015; Nassar et al., 2010; Pouget et al., 2016).

Previous research has shown that error detection and confidence share the same neural correlates (Boldt et al., 2019; Boldt & Yeung, 2015; Desender et al., 2019) and that people use confidence to regulate learning, as well as to make decisions, particularly regarding information-seeking/exploration behaviour (Meyniel et al., 2015; Meyniel & Dehaene, 2017).

In a study by Frömer et al. (2021), they employed a modified time-estimation task where participants had to learn a prespecified time for holding down a key on the keyboard. On every trial, participants reported their estimation of motor error (ranging from too short to too long), their confidence in this estimation, and then received graded feedback on the same scale. They found that the FRN amplitude reflected prediction errors unaccounted for by the reported execution error, rather than the magnitude of the objective error itself. Moreover, they observed that individuals who were better at calibrating their confidence to the precision of their outcome predictions learned more quickly. The authors suggested that higher confidence amplifies the surprise experienced in any mismatch between feedback and error estimation as reflected by the increased P3, thus prompting learning. Altogether, it is evident that our ability to predict action-outcomes, coupled with our awareness of errors, significantly impacts the interpretation and processing of received feedback. More importantly, these factors influence the extent to which feedback information is used to enhance performance.

### 2.2.4. Summary

To conclude, some degree of uncertainty about stimuli/action outcomes is inevitable in real life, whether it comes from the environment, the properties of the feedback itself, or even within ourselves. The degree of uncertainty about the received feedback renders it more or less informative and significantly impacts the interpretation, the neural processing of the feedback, and subsequent behavioural adjustments. Therefore, it is important to incorporate factors that influence uncertainty into our research and study how their effects vary predictably across individuals.

# 3. EXPERIMENTAL CONTRIBUTIONS

In this PhD project, we aim to explore how feedback can be used flexibly and to investigate factors that are important for feedback interpretation, processing, and subsequent adjustment of behaviour. We conducted three experiments during this project, and the research questions, along with the logical basis behind them, are explained as follows:

**Experiment 1 – 'what', 'When', and 'Whether' intentional actions:** In this experiment, we investigated the generation of action-effect predictions under intentional actions. We recognised prior research often treated intentional actions as a unified concept despite they can be separated into three major types depend on the internal decisions regarding *whether* to act, *what* action to perform and *when* to perform it. Our objective here is to ascertain if action-effect predictions exist across all types of intentional actions and whether their strength varies depending on the action type. The result of this investigation aims will provide us a better understand of the extent of human's ability to predict the outcome of intentional behaviour.

**Experiment 2 – The influence of external and internal feedback in learning decision:** Here, we investigated how the decisions to engage in learning for potential rewards are influenced by the feedback we received, as well as, by our performance. The basic behind this experiment is that we are interested in understanding not only how people may use information from feedback to improve learning, but also how they make decisions that are relevant for learning based on it. While positive feedback might motivate continued engagement in a task, the effort required for further attempts at the same task can lead to a decision not to continue. In contrast, negative feedback may discourage further attempts at a task, but the prospect of a possible future reward may motivate perseverance. By combined behavioural measure with EEG, we tested if the decision to continuous or give up learning can be predicted by the valence of feedback, the neural signatures of feedback processing (FRN and P3), also by the objective performance of the participants.

**Experiment 3 – Impacts of confidence and feedback reliability in learning:** In this experiment, we employed a learning task where participants can gradually improve their performance by trial and error. They need to rate their confidence level about

their learning progress during the task, also after they received feedback about their response, they need to make a decision of whether to commit to their selected response for the chance of earning potential reward. The reliability of the feedback (high/low) is explicitly control and the information was given to the participants. By using this design, we aim to investigate how the impact of feedback in behavioural adjustment depend on people's confidence in their learning progress, as well as the reliability of the feedback itself. At the meantime, we address the limitation in experiment 1 by having a subjective measure of learning performance, as well as a learning task that would allow us to observed gradual leaming improvement in extend period.

## 3.1. Experiment 1 – Action-effect predictions in 'what', 'when' and 'whether' intentional action

This chapter is based on: **Chung, W. Y.**, Darriba, Á., Korka, B., Widmann, A., Schröger, E., & Waszak, F. (2022). Action effect predictions in 'what', 'when' and 'whether' intentional actions. *Brain Research*, *1791*, 147992.

### 3.1.1. Introduction

In everyday life, behaviour is usually goal-directed. Performing an action that aims at a desired state presupposes the knowledge of action-effect relationships. Accordingly, the ideomotor principle assumes that action is selected based on the anticipation of its sensory consequences (Greenwald, 1970; Prinz, 1997). The common coding theory has taken the assumption of the ideomotor principle even further by suggesting that actions and their sensory effects share a common representation in the brain. In this way, the anticipation of a desired sensory effect can be used for, or rather is part of, action selection (Hommel et al., 2001; Prinz, 1990).

Over the years, a number of studies have provided empirical evidence in support of the ideomotor principle and the common coding theory. For example, Elsner and Hommel (2001) showed that once participants acquired the associations between the action (right/left key press) and a certain tone (high/low pitch), the action that was previously associated with a specific tone was performed faster when the tone was

played. This finding suggested that the perception of the learned sensory effect primed the action that was previously associated with it, as reflected by the significantly faster reaction time of the performed action. In a neuroimaging study of Kühn and colleagues (2010), the authors showed an action-induced activity in the fusiform face area (FFA) and the parahippocampal place area (PPA) by simply having participants perform actions that were previously associated with the presentation of faces or house stimuli. This finding demonstrated that activation in traditionally called 'perceptual' networks were involved in action control, as the common coding theory suggests.

The ideomotor principle postulates that the intention for action involves a prediction of the upcoming action-effect (Elsner & Hommel, 2001; Prinz, 1997). According to predictive coding models, an internal prediction mechanism compares the incoming sensory input with a model of the expected sensory input generated on the basis of previous experience/learning. The difference between top-down prediction and bottom up sensory data is then translated into prediction error (PE), which is used to correct the prediction from the higher level via a continuous process aimed at minimizing the PE (Friston, 2005). Many previous EEG experiments on ideomotor action have focused on this notion of PE. A common finding is that compared with externally generated stimuli or stimuli triggered by non-intentional actions (for example, TMS triggered movement), the amplitudes of the N1 and sometimes, also the amplitude P2 event-related potentials (ERP) components are found to be attenuated for self-generated stimuli (Bendixen et al., 2012; Horváth, 2015; Timm et al., 2014). These effects are usually interpreted as a consequence of reduced PE, since sensory effects resulting from self-generated actions can be better anticipated than sensory effects externally generated or resulting from involuntary action. There are other studies (Hughes et al., 2013; Korka et al., 2019; Le Bars et al., 2019) in which rather than comparing voluntary and involuntary action effects, expected and unexpected effects of voluntary actions are compared in an active oddball paradigm. In those studies, an attenuation of the N1 amplitude toward predicted action-effects (Hughes et al., 2013; Hughes & Waszak, 2011) and a smaller amplitude in the time range of P2 for mispredicted action-effects has been described (Korka et al., 2019; Korka, Schröger, et al., 2021). This difference in the P2 amplitude between predicted and mispredicted action-effects has been linked to the mismatch negativity (MMN), an ERP component frequently reported in the literature of sensory predictive process in the

auditory system (Näätänen, 1990; Näätänen et al., 2007, 2011; Winkler et al., 1996) and considered to be an index of the learning/updating of the predictive model (Garrido et al., 2009).

In most research on action control, intentional action was treated as a unitary concept. However, intentional action has been categorised into three different types based on the decision process in action planning, namely what to do, when to do it and whether to do it or not (see Brass & Haggard, 2008). Previous neuroimaging studies have provided supporting evidence of potentially dissociated neuroanatomical networks underlying the 'what' and 'when' actions. For example, Mueller et al. (2007) found that activity in the rostral cingulate zone (RCZ) was stronger when participants had to select what to do, relative to being instructed to perform an identical action. These authors also suggested that there could be a possible role for the presupplementary motor area (preSMA) in the internal timing of action (the 'when' component), as they found increased activity in preSMA in both internally and externally selected conditions in which participants were required to internally control the timing of their action based on the bisection point between the interval of the two visual stimuli in both conditions. In a later study, Krieghoff et al. (2009) presented visual cues to indicate if participants had to perform a particular action or were free to decide between two possible actions, and if they had to perform the action with a particular timing or were free to decide between two possible timings. In this design, they managed to separate the what and when actions in different trials, and they found that RCZ is involved in the decision of which action to perform, while an area of the superior frontal gyrus (SFG) in the left paramedian frontal cortex was shown to be involved in the decision of when to act. Regarding the 'whether' action, Kühn and colleagues (2009; 2010) have demonstrated that voluntary non-action/intentionally not to act closely resembles intentional action, as they found that intentional non-action activated brain areas that were involved in the processing of the auditory effect that the decision of not to act was previously associated with, as it could be expected in intentional action. A recent meta-analytic study from Zapparoli and colleagues (2017) have also identified different activation patterns of brain activity in regard to the what, when and whether action component.

So far, no study has investigated and compared the three types of intentional actions (what, when, whether) in the context of action-effect prediction. In the current work, we aim to study (1) whether action-effect predictions are generated under all three types of intentional actions, and (2) whether there are any differences between the three types of actions with regard to the strength of those predictions. In order to achieve these goals, we ran an experiment in which we used ERPs to measure the PE response provoked by the violation of the predicted effect resulting from each type of action. Participants underwent a learning and a test phase for each condition (what, when, whether). In the learning phase, they were required to learn the associations between certain actions (keypresses on a computer keyboard), and certain auditory stimuli (tones with different frequencies). There were two possible actions for each condition and each action was associated with a specific effect (tone). These associations held 100% valid during the learning phase. In the "What" condition, participants decided between two keys to make a response to a 'Go' signal presented on the screen; in the "When" condition, there were two 'Go' signals presented on each trial at different time points, and participants needed to make a decision on which of the 'Go' signals they wished to respond to (the earlier or the later one); in the "Whether" condition, participants had to decide whether to make a response or not to a single 'Go' signal. In the test phase, the association between actions and tones acquired in the learning phase held valid only in 80% of the trials (standard tones). In the remaining 20%, however, a tone different from that associated to each action (or non-action) was presented instead (deviant tones). We expected that the existence of action-effect predictions would be reflected on significant differences between the predicted (standard) and the mispredicted (deviant) tones in the time range of N1 and/or P2, with relatively larger N1 and smaller P2 amplitudes in response to mispredicted than predicted stimuli indicating a prediction error response. More importantly, if the action-effect predictions in the three types of intentional actions are different from each other, those differences should reflect in the degree of the PE response indexed by those components. In this regard, we expect possible differences in terms of how well participants can predict the effects of their actions depending on the type of action performed, firstly, due to the partially isolated neuroanatomical network responsible for the 'what' , 'when' 'whether' action reported in previous studies (Krieghoff et al., 2009; Mueller et al., 2007; Zapparoli et al., 2017), and secondly, due to the possible influence of our daily life experience, where we frequently encounter situations in

which choosing between two different actions results in two different consequences, while expecting different consequences depending on the timing with which an action is performed, or on whether we perform an action or not, may be less commonly experienced. More specifically, the more frequent daily experience of choosing between two different actions resulting in two different consequences may have a bottom-up influence in how well we can learn to predict the consequences of our own actions, and result in a larger degree of PE response in the 'what' action in relative to the 'when' and 'whether' action.

### 3.1.2. Methods

#### *Participants*

Data were collected from 30 participants who received monetary compensation for their participation. The number of participants was determined on the basis of previous studies on action-effect prediction in which significant statistical effects in the time range of N1-P2 were observed with smaller sample sizes (Korka et al., 2019; Timm et al., 2014). Two participants were excluded from the analyses due to the large amount of noise in the EEG signal. Thus, 28 participants remained in the final analyses (16 females, 12 males; 26 right-handed, 2 left-handed, mean age = 26.5, age range = 19-39 years). All participants reported normal hearing and normal or corrected-to-normal vision, and none reported any history of neurological conditions. Written informed consent was obtained and experimental procedures were undertaken in accordance with the Declaration of Helsinki and with the approval by the Comité de Protection des Personnes Ile de France II.

#### *Procedures*

The experiment consisted of three conditions, "What", "When", and "Whether", representing the three different types of intentional action (Figure 7). Each condition included a learning phase followed by a test phase. In the learning phase (20 trials), participants were required to learn the associations between two possible actions (keypresses on a computer keyboard), and two different auditory stimuli (two tones, A and B, with different frequencies). These associations held 100% valid throughout the learning phase. The details of these associations are explained below. The learning

phases were designed to familiarize participants with the tasks as well as to allow them to build up the action-effect contingencies. The test phase of each condition consisted of 6 blocks in total, with 150 trials in each block, resulting in 900 trials overall. Participants completed in succession all the 6 blocks corresponding to each condition before switching to a different condition. In the test phases, the associations between actions and tones that were acquired as 100% reliable in the learning phase, held valid in 80% of the trials only (standard tones); in the remaining 20%, a tone different from that associated to each action (or non-action) was presented instead (deviant tones). All the conditions were designed with this 80/20 vs. 20/80 pattern for each of the two possible actions. The order of the conditions was counterbalanced across participants. The whole experiment took approximately 2.5 hours. Participants had short breaks between blocks within each experimental condition, and also at the beginning of each condition.

In the “When”condition (earlier response vs. later response), each trial began with the word “Ready” presented at the centre of the screen for 200 ms, followed by a fixation cross presented for 500 ms, and the word “Go” presented for 200 ms (Go1) afterwards. After the Go1 signal the fixation cross was presented for another 1000 ms, and followed again by the word “Go” presented for 200 ms (Go2). Participants’ task was to decide whether to respond to the first/earlier “Go” signal (Go1) or to the second/later “Go” signal (Go2) by pressing a key with their right hand (the “L” key on the keyboard). The two “Go” signals were presented in every trial, regardless of participants’ choice. Earlier responses triggered tone A in 80% of the trials (standard, frequent tone), while in the remaining 20% of the trials tone B was presented (deviant, rare tone). Later responses triggered tone B in 80% of the trials (standard, frequent tone), while in the remaining 20% of the trials tone A was presented (deviant, rare tone).

In the “What” (left vs right key) and “Whether” (press vs. no press) conditions each trial began with the word “Ready” presented at the centre of the screen for 200 ms, followed by a fixation cross for either 500 ms or 1500 ms (randomized across trials with equal probability), and the word “Go” presented on the screen for 200 ms to indicate the time for participants to make a response. The reason for using two different fixation cross durations was to make these conditions more easily

comparable to the ‘When’ condition, where the interval between the Ready signal and the action could be either 500 ms or 1500 ms, since participants had to choose between an earlier action (Go1, 500 ms after the Ready signal), and a later action (Go2, presented 1500 ms after the Ready signal).

In the “What” condition (left vs right key), participants’ task was to respond with a key press on the computer keyboard with either their left or right hand (using the “S” and the “L” keys of the keyboard, respectively) whenever a “Go” signal appeared on the screen. Left-hand responses triggered tone A in 80% of the trials (standard, frequent tone), while in the remaining 20% of the trials tone B was presented (deviant, rare tone). Right-hand responses triggered tone B in 80% of the trials (standard, frequent tone), while in the remaining 20% of the trials tone A was presented (deviant, rare tone).

In the “Whether” condition (press vs no press), the procedure was the same as in the “What” condition described above, with the only difference being that instead of choosing between two actions, participants chose whether to press or not a key with their right hand (the “L” key on the keyboard) whenever a ‘go’ signal was presented. If no action was performed before the time for response was exceeded (1000 ms), it was considered that participants have intentionally chosen not to act, and a tone was played after the timeout (1000 ms from the “Go” signal onset) while the tone was played immediately after the keypress in all other conditions. Key presses triggered tone A in 80% of the trials (standard, frequent tone), while in the remaining 20% of the trials tone B was presented (deviant, rare tone). No key-press triggered tone B in 80% of the trials (standard, frequent tone), while in the remaining 20% of the trials tone A was presented (deviant, rare tone).

All visual stimuli were presented on the central of a 27-inch, 60 Hz LCD monitor against a grey background. The distance between the monitor and participants was one meter. Two tones with different frequency were used in this experiment, a high pitch tone (440 Hz) and a low pitch tone (261.63 Hz). The high/low frequency tone – standard/deviant tone patterns were counterbalanced among participants. In every condition, participants had 1000 ms to respond after the “Go” signals and were free to choose which action (depending on the condition) they would like to perform,

but they were asked to aim at selecting each of the two possible actions approximately the same number of times in each task. The inter trial interval was 1000 ms.

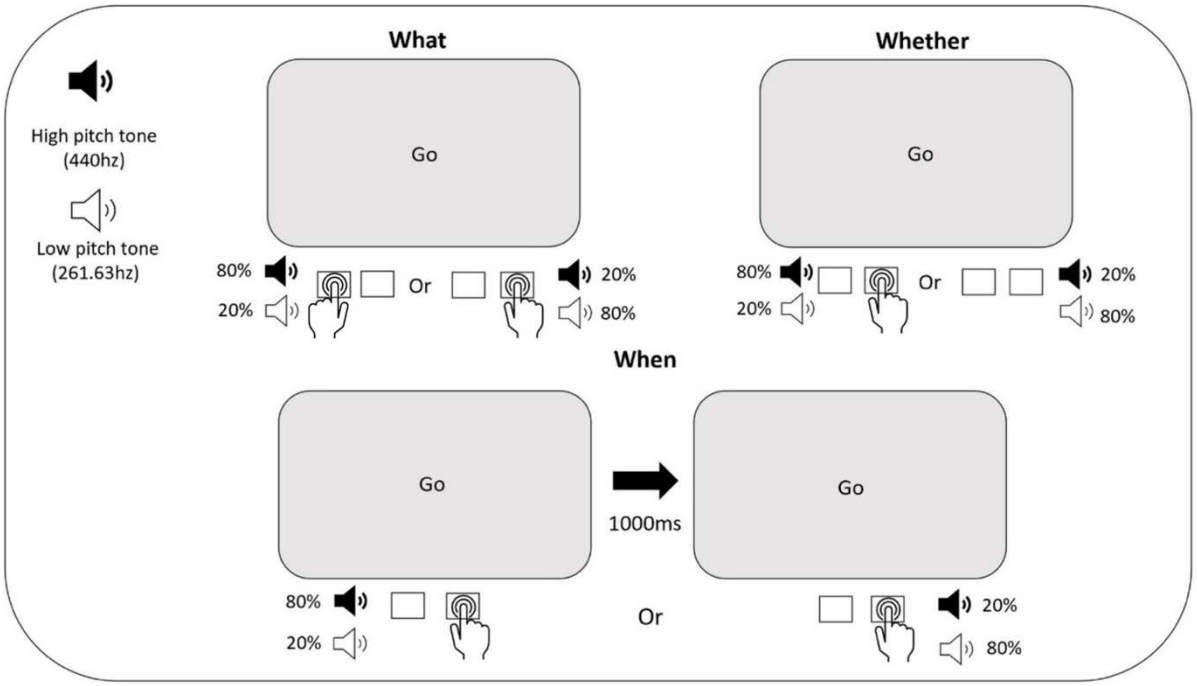



Figure 7. *Schematic representations of the three experimental conditions – what, when and whether.*

### *Data analysis*

*EEG recording and pre-processing.* The experimental task was delivered with Psychtoolbox-3 (Kleiner et al., 2007) running on MATLAB. We recorded the EEG using the PyCorder system and actiCHamp amplifiers (BrainProducts GmbH, Gilching, Germany) in DC recording mode with a sampling rate of 2000 Hz. Continuous EEG data were collected from 60 actiCAP EEG electrodes (BrainProducts GmbH) mounted on an elastic cap and referenced to right mastoid. EEG electrodes were arranged following the extended 10-10 position system (Acharya et al., 2016). Additional electrodes were placed on the right/left mastoid and on the outer canthi of both eyes.

EEGLAB (Delorme & Makeig, 2004) was used for the pre-processing of the EEG data. The EEG data was filtered offline (high pass: 0.1 Hz, low pass: 40 Hz), and re-referenced offline to linked mastoids. Bad channels were identified by visual inspection of the EEG raw data and the channels' power spectra, and excluded from the next processing steps. Epochs were extracted from −200 ms to +1000 ms time-locked to the stimuli onset, and were inspected for non-stereotyped artifacts and

removed if present. Stereotyped artifacts, including blinks, eye movements, and muscle artifacts were deleted via independent component analysis (ICA) using the extended infomax algorithm (Bell & Sejnowski, 1995). Components containing those artifacts were rejected by visual inspection and measures computed with the EEGLAB plug-in functions SASICA (Chaumon et al., 2015) and ADJUST (Mognon et al., 2011). Finally, channels that were deemed bad were reintroduced by interpolating data between neighbouring electrodes using spherical spline interpolation (Perrin et al., 1987).

*Statistical analysis.* We investigated ERP effects related to prediction by comparing ERPs for standard and deviant stimulus (defined by learned action-tone contingencies) in the three experimental conditions (what, when, whether). Single trial EEG data were analysed with a Bayesian linear mixed-model (LMM) analysis using the package brms (Bürkner, 2017), a high-level interface on Stan (Carpenter et al., 2017) in R (RCore, 2016). Plots were made using brms and ggplot2 (Wickham, 2016). An advantage of LMMs over traditional approaches such as repeated measures ANOVA and paired sample t-tests is that a single model can take all sources of variance into account simultaneously. Furthermore, comparisons between conditions can be implemented in a single model. The coefficient estimates are expressed in credible intervals. Credible intervals reflect the intuitive notion of the value of a parameter falling within that interval with a given probability, 95% in this case.

The relevant time window and electrodes for the statistical analysis were first investigated by performing a clustered-based permutation analysis (Maris & Oostenveld, 2007), on aggregated data of the standard stimulus trials and deviant stimulus trials across the three conditions (what, when, whether) with the time window between 0 ms to 1000 ms from stimulus onset using Fieldtrip (Oostenveld et al., 2011) to explore the effect of prediction (deviant vs. standard tone). While this method does not permit to draw conclusions about the significance of specific timepoints and electrode locations, it allows for the identification of time windows and regions of interest for further investigation by providing evidence for a difference in the ERPs between conditions (Maris & Oostenveld, 2007; Sassenhagen & Draschkow, 2019). The clustered-based permutation test revealed a significant cluster ($p$ =.009) extended from approximately 180 to 340 ms (Figure 8) which corresponds to the time range of

the P2 component observed in the grand-averaged waveform. No significant cluster was observed for the latency range of N1. Taking into account the results from the clustered-based permutation test, we limited our analysis to the P2 effect, on a 40 ms window with regard to the peak of the P2 (220 ms) on six frontocentral electrodes (FC1, FC2, FCz, C1, C2, Cz) selected on the basis of the topographical distribution of the activity on the scalp (see Figure 3a) and on previous findings showing maximum amplitudes of P2 in the frontocentral region for auditory stimuli (Baess et al., 2009; Hughes et al., 2013; Näätänen et al., 2011).

We used a predefined model reflecting our experimental design (Barr et al., 2013). Participant amplitudes were normally distributed and did not need transformation to their logarithmic function (Baayen & Milin, 2010). Amplitudes were z-scaled for ease of interpretation and comparison. In the model, observations were predicted by Condition (What vs. When vs. Whether), under which Action (Action 1 vs. Action 2, representing, respectively, Right vs. Left, Go1 vs. Go2, Press vs. No press) and Stimulus (standard vs. deviant) were nested in a full interaction. The model additionally included individual participant intercepts and slopes of Action, Stimulus and their interaction in order to account for individual variation. Contrasts of all categorical factors were centred using sum contrasts (Baayen, 2008), so the intercept of the model represents the grand mean. Planned pairwise comparisons were conducted via Bayesian hypothesis testing using the function Hypothesis in brms with Bonferroni correction. We used a generic weakly informative prior with mean 0 and 1 SD over the fixed effects and kept all other priors at default. We used 4 chains of 2000 iterations each per model, of which 1000 per chain were used for warm-up only, a maximum tree depth of 15 and a target acceptance rate (adapt delta) of .95. Convergence was verified through visual inspection of trace plots, and the Rhat of 1.00 for each parameter.

The model was specified as follows,

brm ( formula : Scaled Amplitude ~ Condition / ( Action * Stimulus ) + (1 + Action * Stimulus | Participant)

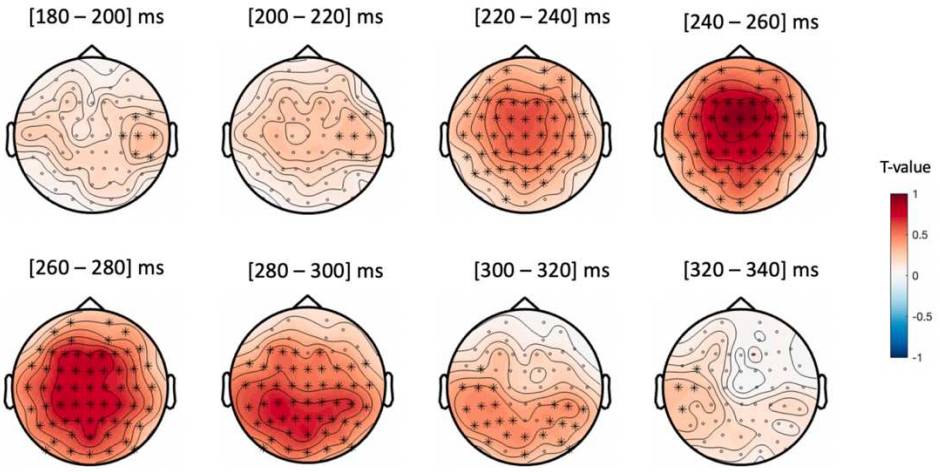



Figure 8. *Results of the cluster-based permutation analysis.*
Comparing the amplitude between standard to deviant stimuli (collapsed over conditions). Electrodes that are part of clusters with p-values < 0.05 are highlighted in the corresponding time windows.

## 3.1.3. Results

### *Behavioural results*

The number of generations of the high pitch (440 Hz) and low pitch tone (261.63 Hz) was not significantly different in any condition ("What" condition: high pitch tone, M = 439.14, SD = 31.74; low pitch tone, M = 440.71, SD = 35.08), $t(27) = 0.14$, $p = .885$; "When" condition: high pitch tone, M = 439.60, SD = 80.41; low pitch tone, M = 419.21, SD = 83.99), $t(27) = 0.86$, $p = .501$; "Whether" condition: high pitch tone, M = 460.75, SD = 93.46; low pitch tone, M = 439.07, SD = 93.60), $t(27) = 0.61$, $p = .545$). Hence, any observed effect in each experimental condition could not be due to the effect of tone frequency or to global differences in tone probability.

### *ERP results*

According to the result of the cluster-based permutation test computed on the difference between deviant and standard tone trials, there were no significant differences between the predicted (standard) and the mispredicted (deviant) tones on the amplitude of N1. All the significant effects were observed in the analysis of P2. Figure 22 depicts the topographic map and the ERP waveforms at the ROI composed of six electrodes in the frontal-central region (FC1, FC2, FCz, C1, C2, Cz).

The results obtained in the statistical analyses are graphically illustrated in Figures 10 and 11. The analysis of P2 amplitudes revealed effects of Stimulus in the "What", "When", and "Whether" conditions, indicating that standard stimuli elicited significantly larger amplitudes than the deviant ones in every condition (Figure 10), plus an Action x Stimulus interaction in the Whether condition (Figure 10). The planned comparisons showed that the size of the standard-deviant difference did not differ between conditions (Figure 11, upper panel). Finally, the planned comparisons ran to examine the Action x Stimulus interaction in the "Whether" condition showed that the difference between standard and deviant stimuli was observed only when participants press a key, but not when they did not (Figure 11, lower panel).

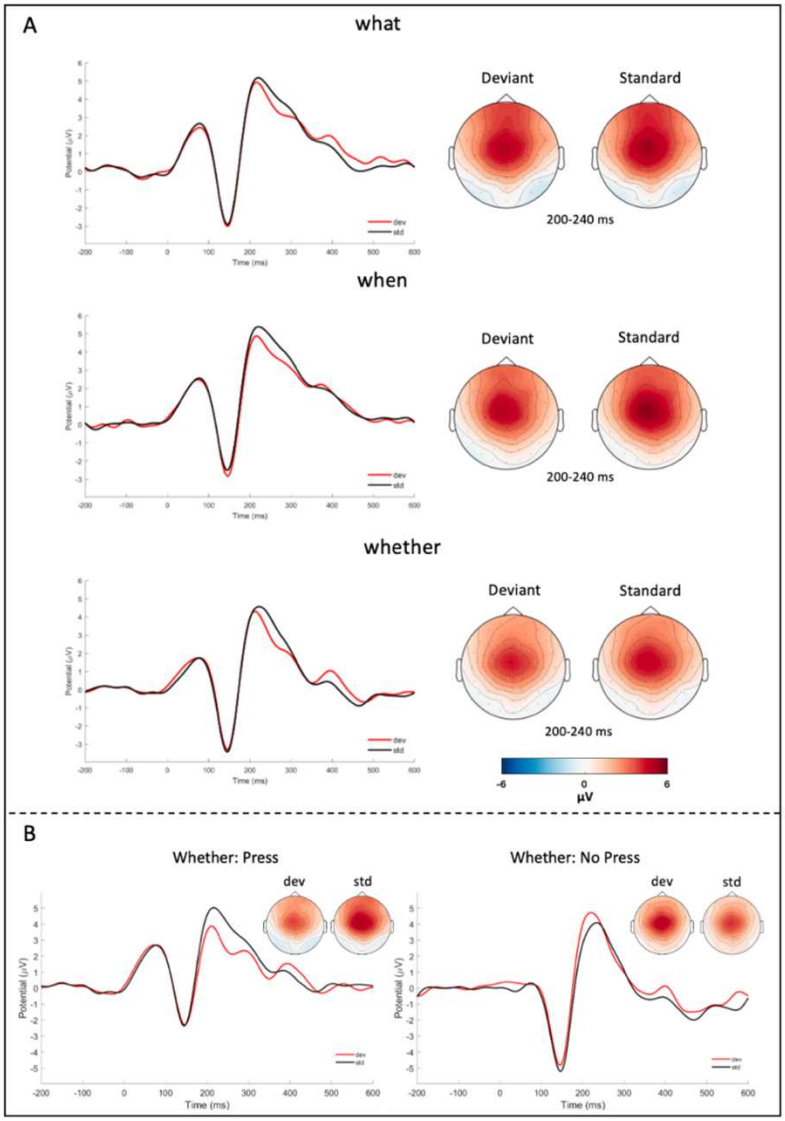
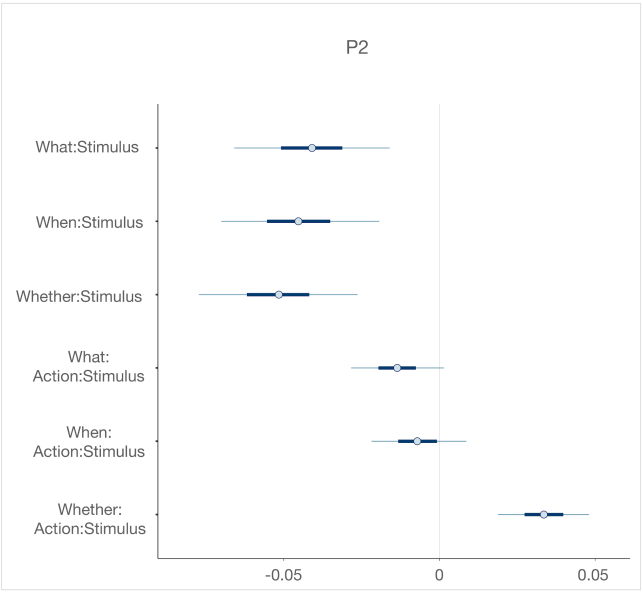



Figure 9. *ERP grand-averages waveforms and topographic maps.*
A) ERP grand-averages waveforms and topographic maps. From the 6 frontocentral electrodes (FC1, FC2, FCz, C1, C2, Cz) for the standard and deviant tones in the What, When, Whether condition. (B) ERP grand-averages

waveforms and topographic maps from the 6 frontocentral electrodes (FC1, FC2, FCz, C1, C2, Cz) for the standard and deviant tones in the Press/action, No Press/nonaction trials in the Whether condition.

A

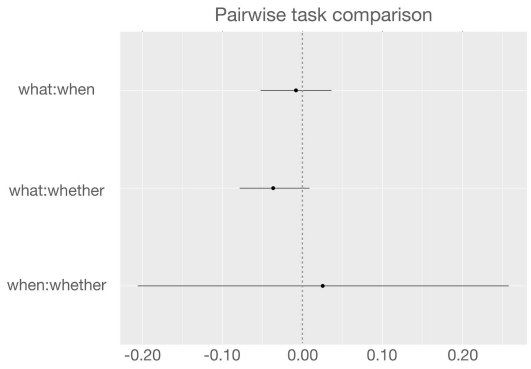



B

| P2 | Median | Lower-95% Confidence Interval | Upper-95% Confidence Interval |
|---|---|---|---|
| What:Stimulus | -0.04 | -0.07 | -0.01 |
| When:Stimulus | -0.05 | -0.08 | -0.01 |
| Whether:Stimulus | -0.05 | -0.08 | -0.02 |
| What: Action:Stimulus | -0.01 | -0.03 | 0.00 |
| When: Action:Stimulus | -0.01 | -0.03 | 0.01 |
| Whether: Action:Stimulus | 0.03 | 0.02 | 0.05 |

Figure 10. Result of the Bayesian linear mixed effect model at the P2 time window.
Medians and credible intervals (On the plot [A]: 50%, thick line; 90% thin line. On the table [B]: 95%) of parameter values in P2. Intervals that do not include zero have the denoted probability to be a true effect.

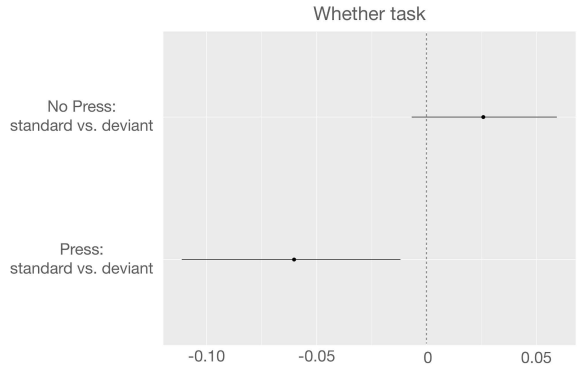
A




B

| P2 | Median | Lower-95% Confidence Interval | Upper-95% Confidence Interval |
|---|---|---|---|
| What:When | -0.01 | -0.05 | 0.04 |
| What:Whether | -0.04 | -0.08 | 0.17 |
| When: Whether | 0.03 | -0.21 | 0.26 |

A




B

| P2 | Median | Lower-95% Confidence Interval | Upper-95% Confidence Interval |
|---|---|---|---|
| No Press: Standard vs. deviant | 0.03 | -0.01 | 0.06 |
| Press: Standard vs. deviant | -0.06 | -0.11 | -0.01 |

Figure 11. Result of the pairwise comparisons of between conditions.
**Upper panel.** Medians and credible intervals (95%) of planned pairwise comparisons of condition estimates (what, when, whether), calculated with the function 'hypothesis' in the R package 'brms', for components in which significant effects were observed. Intervals that do not include zero have the denoted probability to be a true effect, i.e., differences between the indicated factors to be significant. **Lower panel.** Medians and credible intervals (95%) of planned pairwise comparisons of action estimates (press, no press) in the Whether condition, calculated with the function 'hypothesis' in the R package 'brms', for components in which significant effects were observed. Intervals that do not include zero have the denoted probability to be a true effect.

### 3.1.4. Discussion

In the current study, we used EEG to explore action-effect predictions in intentional actions based on three types of decision processes – what (selecting what to do), when (selecting when to act) and whether (to perform the action or not). We found evidence for action-effect prediction in all three types of action decision, as reflected by the significant P2 difference between standard and deviant tones (defined by the learned action-tone contingencies), while we did not observe significant difference between standard and deviant tones in the N1 ERP component in any of the conditions. Furthermore, when we compared the ERPs between the what, when and whether tasks, no significant difference was observed between any of them. Finally, when we looked more closely into the 'whether' action, we found that the P2

difference between standard and deviant tones was only significant when participants chose to execute an action, but not when they decided not to act, despite they have learned the consequence of the non-action. These results are discussed in the following paragraphs.

The P2 component has been previously suggested to reflect the neural process of comparison between sensory inputs and internal predictions (Baldeweg, 2007; Jacobsen & Schröger, 2001; Näätänen et al., 1989; Winkler et al., 1996). The ERP difference observed here between predicted and mispredicted stimuli in P2 is compatible with the time range in which a mismatch negativity (MMN) was reported when comparing standard and deviant stimuli employing auditory oddball paradigms in studies on sensory-based predictions (Näätänen, 1982; Näätänen et al., 2007). The larger P2 amplitude observed in response to predicted compared to mispredicted stimuli is also congruent with previous works in which P2 enhancements were observed for expected relative to unexpected outcomes (Costa-Faidella et al., 2011; Hsu et al., 2015). The observed significant difference in the P2 time range between the standard and deviant tones in our experiment could reflect the comparison between the present auditory input and the predicted input resulting from previous experiences/learning (Garrido et al., 2009). Hence, the significant P2 differences in all of the conditions indicate that effect prediction took place regardless of whether the effects resulted from actions that were based on ‘what’, ‘when’ or ‘whether’ decisions. That is, unlike what could be expected on the basis of previous neuroimaging studies showing that different action decisions elicited activity in different brain region (Krieghoff et al., 2009; Mueller et al., 2007), our results suggest that the effect of action-effect prediction reflected on the EEG activity did not differ between the ‘what’, ‘when’ and ‘whether’ conditions. At first sight, we may postulate that the effects we found in this study are action-type independent because they are based on an action-unspecific predictive mechanism. However, we found that the P2 difference between standard and deviant tones was not significant in the decision of non-action (whether condition) despite participants clearly learned the consequence of non-action. This result appears to suggest a specific role of action in the prediction effect and will be discussed in more detail below. It might be argued that there are possibly differences in allocation of attention between the what, when, and whether conditions. However, note that these differences are in the nature of the three types of decision we studied in the present research. Note also that the lack of statistically significant differences

between the conditions speaks against this possibility and that the ERP results presented here suggest clearly that the action-associated sensory effects could be predicted equally well in the three types of action decision (what, when, whether). Further details about any potential differences between the different types of action decision like, for example, the mechanisms involved in the decisional process, the relative cognitive effort required or the attentional demands imposed, will need to be addressed in future experiments.

The lack of auditory N1 attenuation effect in our results is somewhat surprising given several other experiments reporting such an effect in the literature (Bendixen et al., 2012; Horvath, 2015; Timm et al., 2014). However, there are previous studies in which also did not observe any N1 suppression effect toward predicted stimuli (e.g. Korka et al., 2019; Le Bars et al., 2019). For example, Korka and colleagues (2019), employing a paradigm similar to this study's, found that the N1 suppression effect was sensitive to global regularities rather than to the predictability of the tone, and reasoned that the N1 attenuation effect presumably reflects a stimulus specific adaptation of the neuronal responses. Specifically, they ran an experiment in which the global tone probability was different between three experimental conditions – the "Regularity" condition where the standard tone was being presented overall 80% of time; the "Both" condition where participants were instructed to press one of the buttons 80% of the time, which resulted in the standard tone of the frequently press button being presented more often compared to the other tone; the "Intention" condition which is similar to our experimental paradigm, the mapping of standards and deviants was inversely associated with the left and right keys, and both keys were pressed equally frequent, meaning that the two tones were overall presented with equal chances. They found that the N1 attenuation effect decreased as a function of global tone probability with no observable N1 effect in the "Intention" condition. In this case, our results replicated the findings of Korka et al. (2019) in terms of the N1 effect. As in their study, the two action choices in each condition of our paradigm (different action depending on the conditions) were inversely associated with a low and high pitch tone (the standard tone in one action is the deviant tone in the other action) and participants were instructed to select each of the two possible actions with the same approximately equally often. Therefore, there was not a global regularity for the stimuli in our study. Moreover, while most previous studies described the N1 attenuation effect in the context of Self-generated vs Externally generated stimulus (Sanmiguel et

al., 2013; Saupe et al., 2013; Timm et al., 2013), the comparison in our study was made between predicted and mispredicted sensory effects, and both were self-generated. In a study of Baess et al. (2008), the authors showed that when they compared self-generated and externally generated stimuli, the N1 suppression effect was evident as long as the brain identifies its owner as the agent of the respective sound, even if the frequency of the sound cannot be precisely predicted (ranging from 400 to 1990 Hz). However, they observed the N1 suppression effect to be the largest when both the frequency and the onset of the sound were predictable. In the case of our study, it is possible that mispredicted stimuli are also associated with some degree of N1 suppression, simply because the occurrence of the sound can be predicted via action. Although we would expect the predicted stimuli to show a stronger attenuation effect due to the more precise foreknowledge of the frequency of the sound, the difference in N1 amplitude between predicted and mispredicted stimuli would in the end turn out to be much smaller than the N1 difference observed in previous studies, where the comparisons were made between self-generated and externally generated stimuli.

We see several possible explanations for our result of not observing any prediction effect in the ERPs when no action was executed in the “Whether” condition. Firstly, it could be that there simply was no prediction in the non-action trials. We, however, consider this possibility unlikely since participants were clearly aware of the resulting sensory effect of the decision of not to act and it has been shown in the literature that we can learn to predict sensory effects quickly and efficiently no matter whether the effect is action-related or not (Vroomen and Stekelenburg, 2010). Secondly, we might fail to observe the prediction effect in non-action trials due to some intrinsic differences between the action and non-action trials. For instance, since there is no explicit response in non-action trials, the tone was played 1000 ms after the presentation of the go signal if no action was detected, while the effect tone was played immediately after the execution of action in other conditions. In this regard, previous work has shown that differences in temporal control and temporal proximity of the action resulting sensory effect could affect the predictability as well as the neural response to the predicted stimuli (see Waszak et al., 2012). The delayed presentation of the tone or the fact that no action was executed may also cause participants to pay less attention to the tone, which resulted in a lack of attentional enhancement for the deviant stimulus (Hillyard et al., 1973; Näätänen, 1982). Thirdly, the lack of significant

differences between stimuli when no action was executed in the “Whether” condition might suggest a special role of action-related processing in the generation of the action-effect prediction. Conceptually, the prediction that was learned in the non-action trials of the “Whether” condition can be seen as different from those learned in other (action involving) trials in the sense that the former is a result of learning the arbitrary regularities introduced in an experiment on a cognitive level and the latter were learned via action-effect learning. Based on this difference, there could be extra processing for the experienced sensory effect when the effect is a resulting consequence of voluntary action (Korka, Schröger, et al., 2021; Korka, Widmann, et al., 2021). One possible assumption about this extra processing could be that the action-effect was already activated in the perceptual network before the effect was actually presented. According to the ideomotor principle (Hommel et al., 2001; Prinz, 1997), voluntary action selection is guided by the activation of the action-associated effect in the perceptual network and this process is rendered possible when a common representation is shared between the motor and perceptual network via the learning of action-effect association (Elsner & Hommel, 2001; Kühn et al., 2010). The observed difference in the ERPs between predicted and mispredicted action-effects was found in some studies to be a result of the reduction in sensory processing for the predicted inputs (Hughes et al., 2013; Hughes & Waszak, 2011) and this reduction of processing could be explained as the consequence of the pre-activation of sensory effect by assuming that activity of neurons sensitive to the sensory effect was inhibited in action trials when the actual action-effect is perceived due to the previous pre-activation of the predicted action effect (Waszak et al., 2012). This account of the differences between predicted and mispredicted action-effects parallels explanations proposed to explain the MMN manifesting and getting larger along the number of repetitions of standard stimuli in repetition suppression studies (Baldeweg et al., 2004; Haenschel, 2005), in which predictions are based on sensory evidence. A neurophysiological model proposed by Näätänen (1990) and data from Javitt et al. (1996) have suggested that repeating stimuli lead to an increase in tonic inhibition of supragranular auditory neurons that are sensitive to the frequency of the standard stimulus while simultaneously decreasing the level of inhibition of neurons sensitive to other frequency. However, the similarity between the predictive mechanisms in action-effect based and sensory-based predictions, and their neurophysiological basis are beyond the scope of this study, and their relevance in this experiment is limited as we have no

clear evidence to indicate whether the prediction effect in the current study was resulted by attenuation of predicted inputs, enhancement of deviant events or both, and as we do not have an insight of the internal process of non-action decision in terms of how the prediction of sensory effect could be represented in those trials. Altogether, the lack of observable prediction effect in the ERPs in the non-action trials may be the result of the lack of timing information and temporal control of the stimuli in the non-action trial compared to other action-involved conditions or it can suggest that there are some unique properties in the motor system contributed to the predictive process of sensory effect. Future research is needed to understand whether the prediction of action effects is better explained with a general domain-nonspecific predictive mechanism or motor-specific framework.

To conclude, we showed that action-effect prediction is evident in intentional action, regardless of whether the action choice was based on the selection of action (what), the timing of action (when) or the decision to preform/withdrawal action (whether), and we did not observe any PE difference between the what, when, whether action in the ERPs. This finding suggests that despite different types of intentional action may have different underlying neurobiological underpinnings as shown in previous neuroimaging studies (Krieghoff et al., 2009; Mueller et al., 2007), those differences did not reflect on the learning and the prediction of action associated effect. We also found that the ERPs signature for a prediction effect was no longer observable when no action was performed, which may suggest that action-related process when performing voluntary actions provides extra information for the formation of action-effect predictions. This result, however, needs to be interpreted with caution as there are differences other than the movement itself between action and non-action trials such as, for example, temporal control and temporal proximity. More research is necessary to unambiguously separate the role of action in the predictive process of sensory effect.

## 3.2. Experiment 2 – Give it a second try? The influence of feedback and performance in the decision of reattempting

This chapter is based on**: Chung, W. Y.,** Darriba, Á., Yeung, N., & Waszak, F. (under review). Give it a Second Try? The Influence of Feedback and Performance in the Decision of Reattempting. Preprint available at: http://dx.doi.org/10.2139/ssrn.4580310

### 3.2.1. Introduction

Feedback on performance, either external or internal, is essential for cognitive and motor skill acquisition, since it provides information on whether and how improvement can be made. The neural mechanism behind feedback learning is well described within the framework of reinforcement learning theory. According to this view, learning depends fundamentally on prediction error (PE), i.e., the difference between the actual and the expected outcome of a given action, particularly regarding whether outcomes are better (more rewarding, less costly) or worse (less rewarding, more costly) than expected. PE can be used to form and adjust associations between actions/stimuli and their resulting effect. More importantly, PE can also guide decision making by signalling the need to adjust behaviour. Here, we aim to investigate how evaluation of action outcome as indexed by neural processing of external feedback, in particular, the FRN and P3 ERP components (Schiffer et al., 2017) could translated into appropriate adjustment of behaviour, in term of whether to give a second attempt on the same task or to move on to a new one.

Previous studies using scalp-recorded EEG have revealed an event-related component called feedback related negativity (FRN) that might reflect a feedback-related PE signal (Holroyd & Coles, 2002; Nieuwenhuis et al., 2004; Yordanova et al., 2004). The FRN is a negative frontocentral deflection, peaking around 200-400 ms after feedback presentation, larger for negative compared to positive feedback. It has shown to be sensitive to the size of PE (Holroyd et al., 2003, 2009; Yasuda et al., 2004) and importantly, it has also being linked to the possibility of future behaviour adjustment (Cavanagh et al., 2010; Cavanagh, Figueroa, et al., 2012; Cohen et al., 2011; Cohen & Ranganath, 2007; van de Vijver et al., 2011; Van Der Helden et al., 2010). In this regard, for instance, it has been previously reported that a larger FRN amplitude precedes behaviour switch on a trial to trial basis (Cohen & Ranganath, 2007; Sallet et al., 2013), and that individual differences in FRN magnitude are also predictive of the degree to which participants subsequently avoid decisions with negative outcomes (Frank et al., 2005). In other studies, nonetheless, the P3 ERP component, rather than the FRN, has been shown to be predictive of behaviour adjustment. Specifically, in those studies amplitude enhancements of P3 have been described in response to feedback, before behavioural switches, while the FRN only

reflected the degree of PE (Chase et al., 2011; Schiffer et al., 2017; Yeung & Sanfey, 2004).

In real life, any given feedback, be it positive or negative, can affect behavioural continuation or discontinuation. Positive feedback may encourage people to persevere in a given task, but the effort required for further attempting the same task, or the subjective estimate of their own performance, can make people decide not to continue. Likewise, negative feedback may discourage any further attempts in a task, but subjects can nevertheless be willing to persevere in order to obtain a possible future reward. Given that PE refers to the difference between the expected and received outcome, differences in expectation should modulate the size of PE. One critical factor that influences people's expectation of feedback is their internal estimation of motor performance. In Experiment 1, we demonstrated that individuals can quickly and accurately form predictions about the outcomes of their own actions, regardless of whether these predictions are associated with the selection of an action and/or the timing of an action. Other studies have also found that individuals can report the gradual error of motor performance with reasonable accuracy (Akdoğan & Balcı, 2017; Kononowicz et al., 2019; Kononowicz & Van Wassenhove, 2019). This ability may be supported by reliance on internal models to predict the outcomes of movements (Prinz, 1990, 1997; Wolpert et al., 1995, 2011). More importantly, it was reported in previous findings that individuals automatically discounted the sensorimotor error signal in the presence of prediction error. It was found that when the absence of a rewarding outcome can be attributed to a motor execution error, the value of the option that led to the undesired outcome is not updated based on the negative prediction error. Conversely, in the case of successful motor execution, the option that led to the undesired outcome is penalized, and behaviourally, individuals display a strong risk-aversion bias in their subsequent decision-making (McDougle et al., 2016). On a neural level, an attenuation effect of the signal associated with negative reward prediction errors following execution failures has also been identified (McDougle et al., 2019). In a later study that have a more specific focus on feedback learning (Frömer et al., 2021), the authors use a sensorimotor task to investigate how the estimation of motor performance, together with subjective confidence about the estimation of motor performance affects the neural signature of feedback processing in the EEG signal, and participants learning performance. They found that the FRN amplitude, as an

index of feedback-related PE, did not scale with the degree of error that was indicated in the feedback, but with the degree of execution error indicated by the feedback that was not already accounted for by participants' estimation of their own motor performance. Moreover, it was also found that individuals with better confidence calibration (stronger correlation between the subjective confidence and the objective motor performance) displayed a faster learning rate. Altogether, these findings indicate that the sensorimotor error signal play a significant role in learning. It modulated neural response toward the predication error and its effect on value-based decision making, at the meantime, also influence feedback processing and support adaptive learning.

In the present work, we examined the FRN and P3 components in response to feedback to investigate the influence of performance and feedback evaluation on the decision of persevering or not in a given task. Participants performed a time-estimation task in which they were prompted to reproduce the estimated total duration of a visual stimulus intermittently presented on a computer screen. Feedback, either positive or negative, was given after every response. Then participants had to decide whether to reattempt the same trial or move on to a new one. Participants were rewarded when successfully reattempted the same trial, and penalised when failed at this second attempt, while moving on to the next trial had no further implications. This approach allowed us to create situations where performance and feedback were decorrelated, enabling us to study their independent effects as well as the potential interactive effects they may have on the decision to reattempt.

Given the sensitivity of the FRN to feedback's valence, we expected the FRN to show a larger amplitude in response to negative than to positive feedback. Moreover, in light of the relationship between the FRN and feedback evaluation, and more specifically to the PE elicited by the mismatch between the expected and the received feedback, the FRN should be modulated by participants' performance, which is assumed to be the main basis for participants' expectations about the feedback. Accordingly, we expected feedback and performance to have an interactive effect on the FRN, so that it should show larger amplitudes when the feedback received (positive/negative) does not match performance (good/bad). We expected this interaction to show the role of those factors in participants' decision to retry the current

trial or to move on to the next one. Additionally, the experiment should allow us to dissociate the differential contribution of these factors to the P3 component.

### 3.2.2. Methods

#### *Participants*

Participants gave written informed consent and experimental procedures were undertaken in accordance with the Declaration of Helsinki and with the approval by the Comité de Protection des Personnes Ile de France II. Participants received monetary compensation for their participation. All participants reported normal hearing and normal or corrected-to-normal vision, none of the them reported any history of neurological conditions. Data from one participant were excluded due to the withdrawal of experiment. In addition, data from five participants were excluded from analyses due to insufficient trials for reliable quantification of the ERPs in one of the experimental conditions. Data from one participant were excluded due to excessive number of noisy trials. The final sample consisted of 23 participants (14 females, 9 males; 3 left-handed; mean age = 27.3, age range = 18 – 40). Previous research on feedback processing and behavioural adaptation have found significant statistical effects of the feedback-related potentials in EEG – the feedback-related negativity (FRN), P3a and P3b with similar if not smaller sample sizes (Chase et al., 2011; Schiffer et al., 2017).

#### *Stimuli and task*

Participants performed a time-estimation task with the primary goal of reproducing the total duration of visual stimulus being presented on the screen every trial. Participants received positive or negative feedback after their response followed by a question of whether they want to reattempt the task. The time-estimation task is well established for ERP analyses (Holroyd & Krigolson, 2007; Luft et al., 2014; Miltner et al., 1997).

We used a black dot (visual angle: 0.44°) presented in the center of the screen as the visual stimuli. For each trial, the number of dots being presented was randomly chosen between 1 to 4. The duration for each of the dot was randomly selected from

300 ms to 1000 ms with a step size of 100 ms, and the interval between each of them (if the number of dots was more than one on that trial) was randomly selected from 300 ms to 1200 ms with a step size of 300 ms. Participants were instructed to estimate the total duration of all the dots combined and to ignore the interval between each of the dot in their estimation. A black check mark and a black X were used as symbols for positive and negative feedback respectively (visual angle: 1.2°) in the first attempt trial. Positive and negative feedback was given randomly apart from situation where participants displayed an extremely good (the mismatch between the estimated duration and the total duration of presented stimuli was less than 5 % of the total duration of the presented stimuli) or bad performance (worse than 95 % of the time of participants' previous performance based on preceding trials). Thus, establishing feedback validity.

For the second attempt trials (if participants decided to repeated task), feedback was given based on participants' performance. The feedback was represented using a scale from -5 to 5 points, excluding zero. A positive/negative number indicated positive and negative feedback respectively. If the mismatch between the estimation and total duration of the stimuli was no larger than 25% of the total duration of the presented stimuli, positive feedback is given. The point increased from 1 to 5 with the mismatch decreased from 25% to 5% with a step of 5%. Likewise, negative feedback was given when the mismatch is larger than 25% of the total duration of the presented stimuli, and the point went from -1 to -5 as the mismatch get larger from 25% to 50%, with a step of 5%. The point will be translated into monetary reward at the end of the experiment (1 point = 5 cents).

All stimuli were presented at the centre of the screen against a grey background on a 27-inch, 60 Hz LCD display with a distance of 100 cm from the participants. The experiment comprised eight blocks of 60 trials each, with self-paced rests between blocks. The task was delivered with Psychotoolbox-3 (Kleiner et al., 2007) running on MATLAB. Prior to the experiment, participants received both written and verbal instructions that explained the procedure of the experiment and preformed 20 practice trials.

***Procedure***

The procedure of a signal trial is illustrated in Figure 12. Each trial began with the word “Ready” presented at the centre of the screen for 200 ms, followed by a fixation cross presented for 300 ms. Then the visual stimuli (either a single black dot or a series of black dot with random interval in between) was presented on the screen. After the stimulus presentation, the word “Response” was presented on the screen to signal participants to give their response using the space bar on the keyboard. The response is given by holding down the space bar for the duration that participants estimated to be the total duration of the stimuli. A black dot was appeared on the screen during the key press to help participants visualise their estimation. Feedback is given 600 ms after the release of the key and stay on the screen for 1000 ms. Afterward, a question appeared to ask participants if they want to repeat the task, and the current number of points they are holding was also shown on the screen. Participants then used their right or left index finger to press the “S” or “L” key to indicate a “Yes” or “No” decision (the mapping between Yes/No and left/right key was counterbalanced across subjects). The question remained on the screen until participants made a response. If the decision was “Yes”, then trial repeated as described above. The only difference was that the feedback on the repeated trial was represented as number of points with a range of -5 to 5 (excluding zero). A negative number indicated negative feedback, and a positive number indicated positive feedback. If the decision was “No”, then a new trial began. The intertrial interval was 1000ms. Participants was encouraged to maximize the number of points during the experiment.

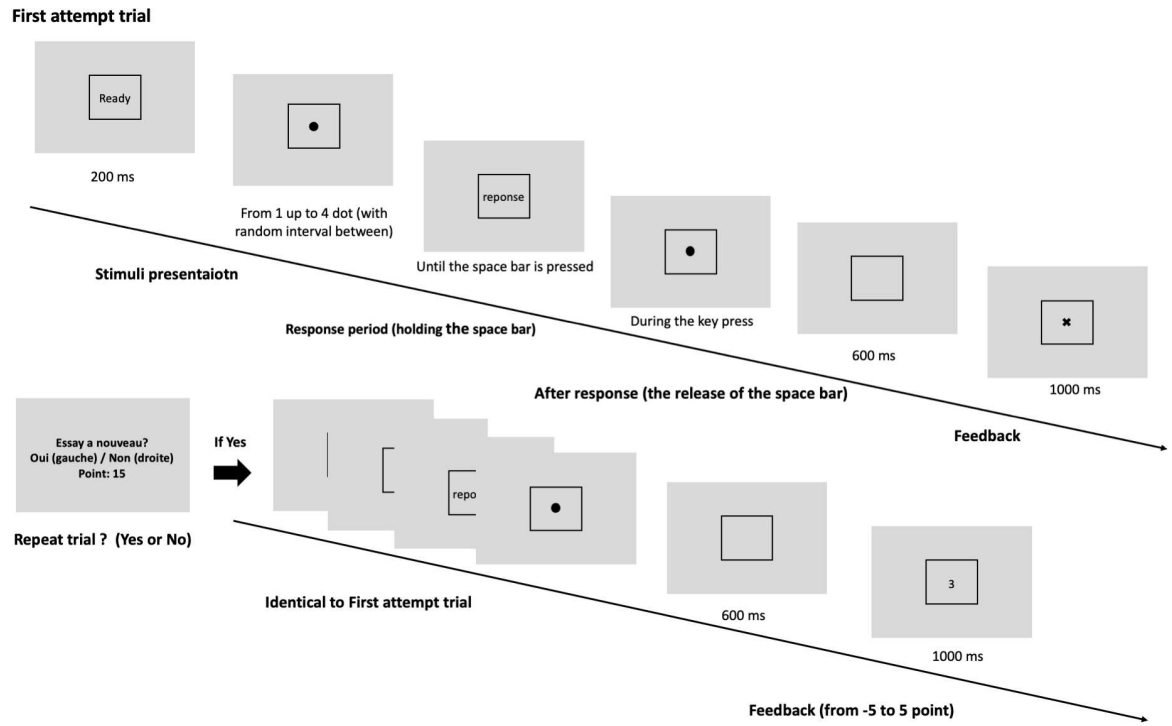



Figure 12. Schematic representation of a trial structure.
Participants learned to reproduce the duration of the presented stimuli for each trial by holding the space bar on the keyboard. Following each response, they received a positive or negative feedback. Then participants have to decide whether to give a second attempt for the same trial. The feedback on the second attempt was represented using a scale from -5 to 5 points, excluding zero. A positive/negative number indicated positive and negative feedback respectively.

### *EEG data acquisition and preprocessing*

We recorded the EEG using the PyCorder system and actiCHamp amplifiers (BrainProducts GmbH, Gilching, Germany) in DC recording mode with a sampling rate of 2000 Hz. Continuous EEG data were collected from 60 actiCAP EEG electrodes (BrainProducts GmbH) mounted on an elastic cap and referenced to right mastoid. EEG electrodes were arranged following the extended 10-10 position system (Acharya et al., 2016). Additional electrodes were placed on the right/left mastoid and on the outer canthi of both eyes. Custom-built Matlab scripts with EEGLAB (Delorme & Makeig, 2004) functions was used for the pre-processing of the EEG data. The EEG data was filtered offline (high pass: 0.1 Hz, low pass: 40 Hz), and re-referenced to linked mastoids. Bad channels were identified by visual inspection of the EEG raw data and the channels' power spectra. Independent component analysis (ICA) was performed to identify and remove components that were associated with blinks and eye movements. Subsequently, we removed all trials in which activity exceeded ± 100

μV to account for noise and large muscle artefacts, resulting in the average exclusion rate of 0.88% (SD = 1.77). Bad channels were reintroduced by interpolating data between neighbouring electrodes using spherical spline interpolation (Perrin et al., 1987). Epochs were extracted from −200 ms to +1000 ms relative to feedback onset and baselines were corrected to the 200 ms pre-stimulus interval.

The FRN was quantified in single-trial waveforms as the peak-to-peak amplitude at electrode FCz. Specifically, we identified the minimum voltage in the 200 to 300 ms window, then subtract it from the preceding positive maximum in the window from -100 to 0 ms relative to the detected negative peak. For the P3, the amplitude was measured at a cluster of fronto-central region electrodes comprising FCz, FC1, FC2, Cz, C1, C2 in a time window from 318-418 ms post-feedback onset.

***Data analysis***

We included only the first attempt trials with random feedback in the analyses. Performance on each trial was measured as the absolute error in milliseconds between the total duration and the estimated duration of the stimuli. Trials with a magnitude of absolute error below the individual mean of each participant were labelled as good performance trials, and the rest were labelled as bad performance trials.

We first used linear mixed model analyses to examine the effect of feedback and performance on FRN and P3 amplitudes separately. Then, in order to evaluate the joint effect of EEG activity, feedback and performance on the decision of whether to repeat or not a trial, we analysed decision using generalised linear mixed model with Feedback, Performance, and the amplitudes of FRN and P3 as predictors. The mixed effect model analysis has the advantages that it allows for parametric analyses of single-trial measures and robust to unequally distributed numbers of observations across participants. Furthermore, it allow us to take into account of the individual variance regarding experimental effects by including participants as random effect and the predictors as random slopes (Frömer et al., 2018). Variable that explained zero variance were excluded from the random effects structure to prevent overparameterization (Bates et al., 2015; Matuschek et al., 2017). We applied sliding difference contrasts for all the categorical predictors – feedback and performance.

Single-trial FRN, P3 amplitude in the generalised linear mixed model were mean-centred and divided by 10 as the rescaling of the variables support model identifiability (Bates et al., 2015). The models were reduced stepwise by excluding non-significant interaction terms until the respectively smaller model explained the data significantly worse than the larger model. We reported the AIC (Akaike Information Criterion) and BIC (Bayesian Information Criterion), fit indices that are smaller for better fitting models. Statistical analysis were performed using R (R Core Team, 2022) with the lme4 package (Bates et al., 2015), p-values were computed with the lmerTest package, using Satterthwaite approximation for degrees of freedom. Graphs were made using ggplot2 (Wickham, 2016) and effects package (Fox & Weisberg, 2019).

### 3.2.3. Results

#### *Behavioural data*

To test whether participants successfully performed the task, we checked the correlation between the stimuli actual duration and duration estimates in each participant. We found a significant positive correlation ($p$ <.001) in all participants, indicating that they understood and were able to perform the task adequately (Figure 13).

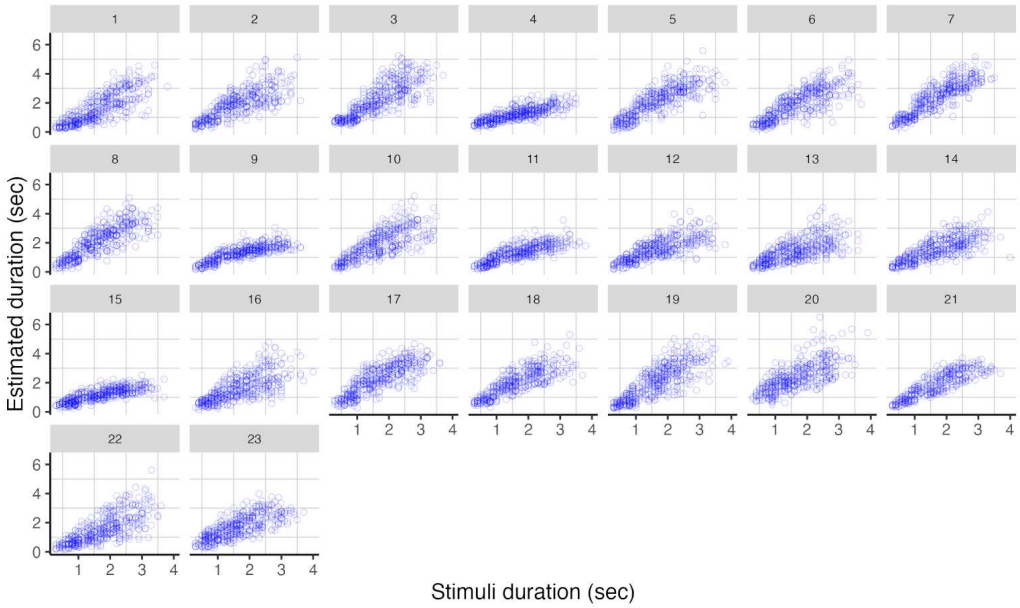



Figure 13.Correlation between stimuli duration and duration estimates for each participant.

#### *ERP analysis: FRN and P3*

Figure 14 illustrates the waveforms and topographies of the FRN and P3 ERP components as a function of feedback and performance. We applied linear mixed model analyses to examine the effect of feedback and performance on FRN and P3 amplitude separately. The model estimates are summarised in Table 1. For the FRN (left part of Table 1), we observed a main effect of Feedback with larger FRN amplitude for negative feedback compared to positive feedback trials. We did not observe a significant effect of Performance nor significant Feedback by Performance interaction. The non-significant interaction term was excluded from the model. The reduction of the model did not result in a significant drop in the model fit ($\Delta X^2(1) = 0.28$, $p = .600$), moreover, the fit indices were smaller for the reduced model (AIC $_{\text{reduced-full}} = -2$, BIC $_{\text{reduced-full}} = -9$), indicating a better fit.

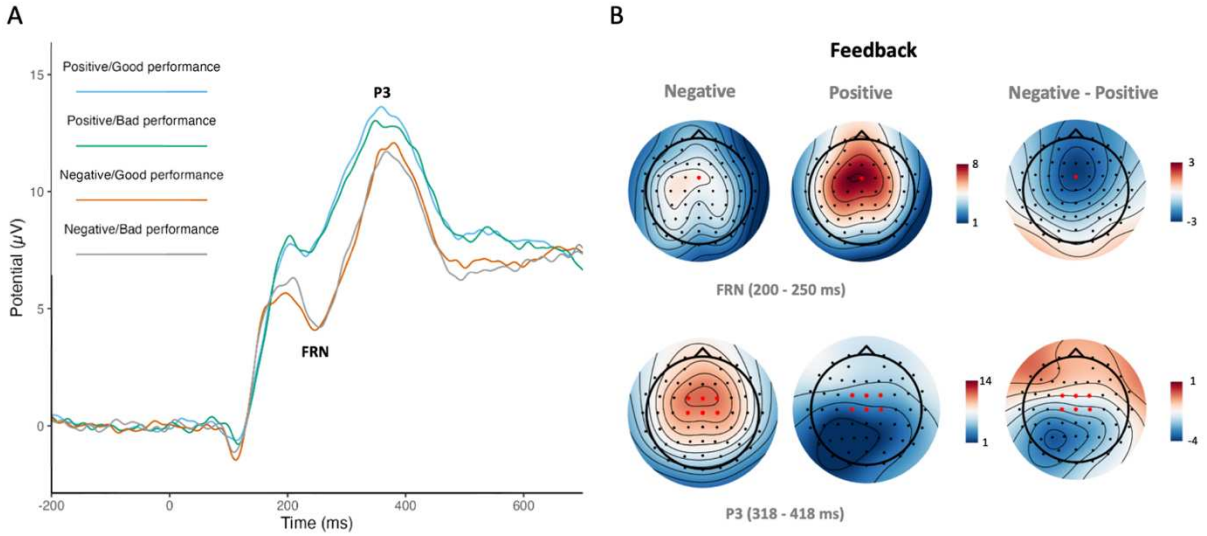



Figure 14. ERP waveforms and topographies.
**(A)** ERP waveforms corresponding to the frontocentral electrode cluster (FCz, FC1, FC2, Cz, C1, C2), showing FRN (200 – 250 ms) and P3 (310 – 418 ms). **(B)** Topographic maps of average amplitude in FRN time window and P3 time window for Negative feedback (left panel), Positive feedback (middle panel) and the difference between Negative and Positive feedback actions (right panel). The electrodes marked on the topographical maps represents the ones included in the FRN and P3 analysis respectively.

For the P3 (right part of Table 1), we observed a main effect of Feedback with larger P3 amplitude for positive feedback compared to negative feedback trials. No significant effect of Performance nor significant Feedback by Performance interaction were observed. The final model was reduced by excluding the non-significant interaction term. The exclusion of the interaction terms did not significantly decrease model fit ($\Delta X^2(1) = 0.29$, $p = .570$) and the fit indices were smaller for the reduced model (AIC $_{\text{reduced-full}} = -2$, BIC $_{\text{reduced-full}} = -9$).

**Table 1.** Feedback and performance effects on ERPs

| | FRN | | | | P3 | | | |
|---|---|---|---|---|---|---|---|---|
| *Predictors* | *Estimates* | *SE* | *t* | *p* | *Estimates* | *SE* | *t* | *p* |
| (Intercept) | -19.13 | 1.00 | -19.17 | **1.21e-15** | 11.72 | 1.03 | 11.32 | **6.99e-11** |
| Feedback | 2.30 | 0.37 | 6.29 | **1.96e-06** | 1.90 | 0.48 | 3.99 | **5.90e-04** |
| Performance | 0.01 | 0.17 | 0.07 | 0.949 | 0.30 | 0.31 | 0.98 | 0.335 |
| Random Effects | | SD | | | | SD | | |
| Residuals | | 8.02 | | | | 9.15 | | |
| Intercept | | 4.77 | | | | 4.94 | | |
| Feedback | | 1.55 | | | | 2.09 | | |
| Performance | | | | | | 1.12 | | |
| Model Parameters | | | | | | | | |
| N | | 23 | | | | 23 | | |
| Observations | | 8830 | | | | 8830 | | |
| Deviance | | 61961.3 | | | | 62329 | | |
| log-Likelihood | | -30980.7 | | | | -32163 | | |

*Formula: FRN ~Feedback + Performance + (Feedback |participant);*
*P3 ~Feedback + Performance + (Feedback + Performance |participant)*

### *Brain-behaviour relationship*

We used a generalised mixed model to estimate the effect of Feedback, Performance, FRN amplitude, and P3 amplitude on participants' decisions. Table 2 displays the fixed-effects estimates, standard errors, z-values, as well as estimates of the square root of the variance components (SD) and goodness of fit parameters for the mixed effect model. The final model excluded all the non-significant interactions and the exclusion of those interaction terms did not significantly decrease model fit ($\Delta X^2(7) = 3.83$, $p = .799$). The fit indices were smaller for the reduced model (AIC $_{\text{reduced-full}} = -10$, BIC $_{\text{reduced-full}} = -60$), which indicated a better fit. The results showed that feedback valence predicted the probability of decision, so that the probability of choosing to reattempt the ongoing trial was significantly higher after receiving positive feedback. A significant effect of Performance was also observed, indicating that the probability of reattempting the trial was significantly higher after good performance.

**Table 2.** Effects of feedback, performance, FRN and P3 amplitude on decision.

| *Predictors* | *Estimates* | *SE* | *z-value* | *p-value* |
|---|---|---|---|---|
| (Intercept) | 0.32 | 0.23 | 1.41 | 0.158141 |
| Feedback | 1.43 | 0.40 | 3.59 | **0.000336** *** |
| Performance | 0.32 | 0.10 | 3.05 | **0.002285** ** |
| FRN | 1.02e-05 | 0.03 | 0 | 0.999762 |
| P3 | -0.05 | 0.03 | -1.66 | 0.097237 |
| Performance: FRN | -0.18 | 0.06 | -2.82 | **0.004784** ** |
| Performance: P3 | -0.07 | 0.06 | -1.17 | 0.242919 |
| FRN: P3 | -0.07 | 0.03 | -2.44 | **0.014586** * |
| Performance: FRN: P3 | 0.13 | 0.05 | 2.36 | **0.018082** * |

| *Random Effects* | *SD* | *Model Parameters* | |
|---|---|---|---|
| Intercept | 1.07 | N | 23 |
| Feedback | 1.88 | Observations | 8830 |
| Performance | 0.42 | log-Likelihood | -4504.1 |
| | | Deviance | 9008.2 |

*Formula: Decision ~ Feedback + Performance*(FRN*P3) + (feedback + performance | participant)*

In addition, analysis also revealed significant performance by FRN interaction, FRN by P3 interaction and a three-way interaction between performance, FRN and P3. For the performance by FRN interaction, we observed that larger FRN amplitude was related to a higher probability of repeating the current trial in good performance trials, while this relationship reversed in bad performance trials, in which a larger FRN amplitude was associated with lower probability of retrying the trial (Figure 15). In other words, when the FRN amplitude was relatively large, the model predicted a large effect of Performance on decision, i.e., a decision coherent with performance, and this effect was predicted to get smaller along with smaller FRN amplitudes.

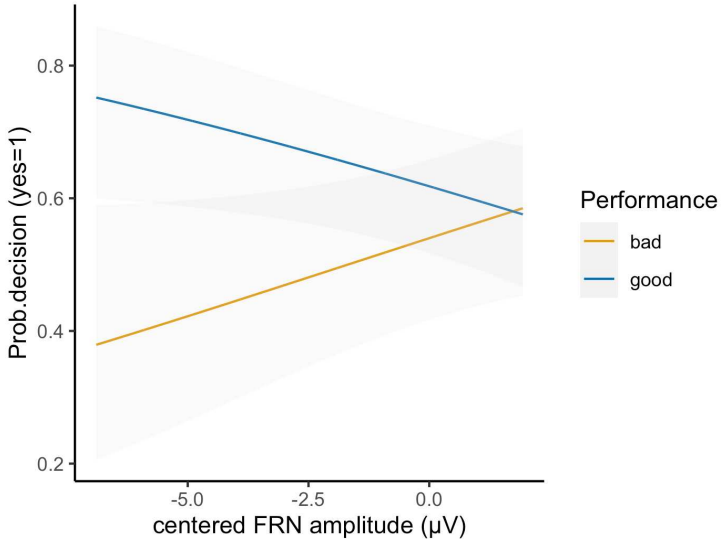


Figure 15. Model estimation of the Performance by FRN interaction.
Performance by FRN interaction on the probability of repeating the current trial. The shaded regions represent 95% confidence intervals.

Furthermore, we observed a FRN by P3 interaction indicating a positive correlation between FRN amplitude and the probability of repeating task when the P3 amplitude was small. This correlation became progressively more negative with increasing P3 amplitude (Fig 17A). The three-way interaction of performance, FRN and P3 further indicated that the aforementioned FRN by P3 interaction only took place in bad performance trials (Fig 17B).

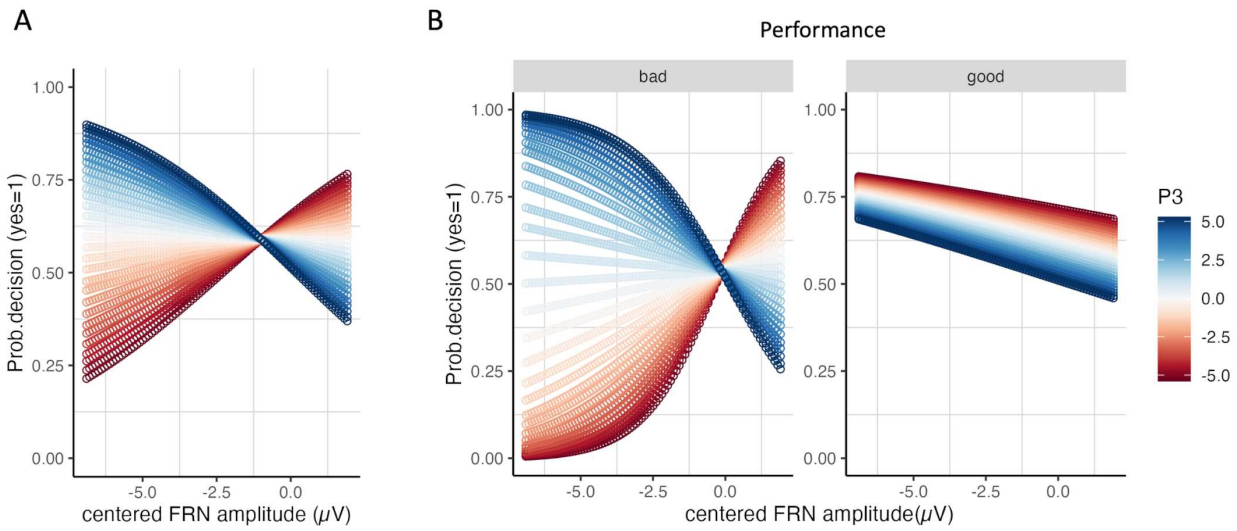



*Figure 16*. Model estimation of interaction between Performance, FRN and P3. **(A)** Interaction of FRN and P3 amplitude on the probability of repeating trial. (**B)** Interaction of FRN and P3 amplitude on the probability of repeating trial by Performance.

## 3.2.4. Discussion

In the present study, we tested the influence of feedback and performance on the decision to retry or not the ongoing trial in a time estimation task. In the experiment, participants were asked to reproduce the total duration of an intermittently presented visual stimulus. Feedback was given after every response, and participants had to decide after the feedback whether to retry the same trial and try to earn reward points, or to move on to the next trial. Results showed that both performance and feedback influenced participants' decision, since the probability of retrying the current trial was significantly higher after receiving positive feedback and after a good performance in the first try. We focused our analyses on the amplitudes of two feedback- and decision-related ERP components, the FRN and the P3. The main results found concerned the

FRN. Previous works have shown the sensitivity of the FRN amplitude to feedback's valence. In line with those studies, in the present experiment the FRN showed larger amplitude in response to negative than to positive feedback. Previous works have also related the amplitude of the FRN to the size of PE. Our results are also in agreement with this relationship, and provide further insight in how PE size influences participants' decisions. Specifically, we found that the larger this PE is, the more likely participants are to make their decision, be this to reattempt the current trial or to move on to the next, in accordance to their performance and regardless of the feedback received. In the opposite direction, the smaller PE is the more likely participants will base their decision on the feedback received.

The results obtained in the analyses showed that participants' decision to retry the ongoing trials in order to earn a reward was influenced both by how good their performance was in their first attempt and by the valence of the feedback they received. Specifically, two independent main effects of Performance and Feedback were found, indicating that subjects were more likely to retry the current trial when performance was good and when feedback was positive. No interaction between these factors was observed. The influence of feedback on decision suggests that participants were not aware of the fact that positive and negative feedback were randomly given regardless of actual performance. The influence of performance on the likelihood to retry a trial indicates that participants had an internal estimate of their execution that they used to make their decisions, at least in some trials. The models including the ERP data provided more information about the circumstances under which performance information might have been used, as we will see in the next paragraphs.

The first linear mixed model analysis, performed to investigate the influence of performance and feedback on FRN and P3 amplitudes, revealed that only the latter had any effect on both components, while performance had no significant effect on them and neither interacted in a significant way with feedback. Results revealed, as expected, that negative feedback was associated with larger FRN amplitude compared to positive feedback, in agreement with previous studies showing that FRN is in general associated with the degree of negative PE (Holroyd et al., 2004; Holroyd & Coles, 2002; Luft et al., 2014; Nieuwenhuis et al., 2004, but see Holroyd et al., 2008). Regarding P3, a larger P3 response was observed in response to positive compared

to negative feedback. This result is congruent with studies showing larger P3 amplitude following positive feedback in motor learning tasks, in which positive feedback has been shown to be more relevant than negative feedback for performance improvement (Chiviacowsky & Wulf, 2007). It is also coherent with findings showing larger P3 response toward motivational significant stimuli (Nieuwenhuis et al., 2005), as positive feedback is likely to contain higher motivational significance in this task since it would signify a higher probability for future reward.

A subsequent linear model analysis was performed to address the influence of performance, feedback, FRN amplitude, and P3 amplitude, on participants' decisions. This analysis showed that the decision of whether to reattempt or not the same trial was heavily driven by feedback valence, as the probability of reattempting the current trial was significantly higher after positive than negative feedback. Importantly, this result also indicates that participants were not aware that positive and negative feedback were randomly given in our experimental design. Moreover, results revealed a significant effect of Performance, suggesting that participants were able to access and use information from their internal monitoring system to help in their decision making. This effect depended on the amplitude of the FRN, as reflected by the significant FRN by Performance interaction. Specifically, we observed that only when the FRN amplitude was relatively large, the model predicted a large effect of Performance on decision, and that, conversely, the effect was predicted to get smaller along with smaller FRN amplitudes. Given the established relationship between FRN amplitude and PE size, a larger FRN amplitude would indicate here that participants had an expectation about the valence of the upcoming feedback, resulting from a more or less accurate estimate of their performance. This expectation would be compared with the actual feedback received, and the degree of mismatch between expected and received feedback would be reflected on the FRN amplitude. Hence, the larger the FRN, the more likely that participants prefer to rely on their own performance estimate to make their decision instead of the feedback, making them more likely to retry the current trial when their performance was good, or to move on to the next one when performance was bad. According to this interpretation, a relatively smaller FRN could result from two possible circumstances. One possibility is that participants were able to correctly monitor their performance and the feedback received was in agreement with their estimate. As a result, PE would be small and consequently the amplitude of

the FRN would also be reduced. In this case, participants should be more likely to retry a given trial when performance is good and feedback is positive, and more likely to move on to the next trial when performance is bad and feedback is negative. Our results do not support this interpretation, given the main finding that the larger the FRN, the more likely the decision is coherent with performance. Alternatively, a smaller FRN could result from participants being unable to monitor their performance in those trials and therefore not generating expectations about the feedback, so that PE in response to the feedback would be relatively smaller. Hence, decisions would not be based on performance in those trials, but rather on the feedback received, as indicated by the main effect of Feedback. This explanation would be in agreement with the notion that the absence of clear expectations diminishes effect of surprise (Hayden et al., 2011), and with previous studies in which a smaller FRN was described when participants cannot form a direct relation between the outcome and their behaviour (Holroyd et al., 2009; Holroyd & Coles, 2002). It would also be in agreement with the notion that PE resulting from unpredicted stimuli is smaller than that resulting from mispredicted, and even predicted, stimuli (Hsu et al., 2015). Specifically, according to this notion, the presence of an expectation, be this accurate or not, would generate a larger PE (and thus larger FRN) compared to trials in which no expectation is present. Therefore, a relatively large FRN amplitude would in general be observed when participants have an internal estimate of their performance in a given trial.

Finally, an interaction between FRN and P3, together with a three-way interaction between Performance, FRN, and P3 completed the results about the relationship between FRN amplitude and Performance by showing an exception to the pattern described above. Specifically, these interactions revealed that while a large FRN was more likely related to the participants' decision to move on to the next trial when performance was bad, as they are more likely to make their decision according to their estimate of their own performance, a large FRN did not predict a decision according to performance when followed by an also large P3. Rather, in this case participants were more likely to retry the ongoing trial. It is not clear to us why a larger P3 amplitude would increase participants' likelihood of retrying a given trial despite presumably being aware that their performance was bad. It has been previously reported that P3 amplitude varies with subjective confidence, being larger for higher confidence or certainty in a given response or estimate (Butterfield & Mangels, 2003;

Hillyard et al., 1971; Ye et al., 2019). It is therefore possible that there is a number of trials in which participants had high confidence in their ability to perform the task correctly regardless of their execution in their first attempt, and decided to give it a second try. Unfortunately, a limitation of our data is that they don't allow us to test this hypothesis. Given the well-established role of confidence in decision-making (Rouault et al., 2019; Yeung & Summerfield, 2012b), future research should incorporate this variable to the experimental design to test its contribution to participants' decision-making.

To summarise, our results show that participants used information from both external (feedback) and internal (performance monitoring) sources to make their decisions in the proposed experimental task. We found that both factors play a role in the decision process, as participants were in general more likely to repeat the current trial when performance was good, regardless of the feedback received, and when feedback was positive, regardless of their performance. The analyses of the P3, and particularly of the FRN component, allowed us to shed some light on how these factors influenced participants' decisions. In line with previous studies pointing out the sensitivity of the FRN amplitude to feedback's valence, the FRN showed larger amplitude in response to negative than to positive feedback. More importantly, in relation to previous works relating the amplitude of the FRN to the size of PE, our results revealed that the amplitude of the FRN, regardless of the valence of the feedback received, seems to indicate whether participants have a clear estimate of their performance and, consequently, will more likely rely on that estimate for their decision or, in the absence of such an estimate, will rather rely on the feedback received.

It is important to note that an obvious limitation of our experimental design is that we did not explicitly ask participants to estimate their performance on every trial, since we wanted to avoid questions that could interfere with their decisions. Therefore, we do not have direct information about participants' internal performance estimates on a trial basis, neither about how accurate these are nor about how confident participants felt about their estimates. Those factors may have an impact on the decision of whether to go for a further attempt in the task (Frömer et al., 2021) but

were only indirectly inferred in the present work. Future studies should address this issue for a better understanding of the results presented here.

## 3.3. Experiment 3 – The impacts of confidence and feedback reliability in learning adjustment

This chapter is based on behavioural data collect in a pilot experiment designed by **Chung, W. Y.,** Darriba, Á., Yeung, N., & Waszak, F. to investigate how the impact of feedback in behavioural adjustment depend on people's confidence in their learning progress, as well as the reliability of the feedback itself.

### 3.3.1. Introduction

In Experiment 2, our investigation centered on the influence of external and internal motor feedback on the decision of reattempting task. We found results which suggested that both type of feedback play a significant role in the decision and the amplitude of the feedback related potential – FRN predicted the effect of internal motor feedback on the decision by reflecting whether participants was able to form a clear prediction about their motor performance (Hayden et al., 2011; Holroyd et al., 2009; Hsu et al., 2015). While these results have provided valuable insights into the decision-making process of reattempting in learning, certain limitations became apparent. Firstly, in the experiment design, participants at most were only given two attempt to perform the same task and we changed the task parameters on every trial to ensure that it would be clear for the participants that when they decided to move on to a new trial, it will be a new task. Under this experimental design, we are unable to observe the process of learning or to estimate any learning rate of the task with only data from two trials with the same task parameters. Secondly, despite our suggestion of the amplitude of the FRN could reflect an internal awareness of motor performance, we lack direct evidence to support it as we never asked participant to explicitly state any prediction they have on their performance after motor execution. More importantly, even if there was always a prediction of their motor performance due to internal motor model (Wolpert et al., 1995, 2011), it would still be curial to know how confidence participants felt about those predictions as this would have a direct impact on the

processing of feedback, action outcome, also the possible learning improvement (Frömer et al., 2021).

We addressed the aforementioned limitations in the following experiment. In experiment 3, we employed a learning task where participants need to rely on the feedback, they received on every trial to progressively enhance their performance and learn the underlying distribution for each experimental block. Specifically, participants were told to imagine that they are playing a shooting game in a very windy forest, Therefore, in order to hit the target that is always present in the middle, they need to learn the optimal aiming location based on the wind strength at the time. The wind strength for each experimental block is control by a Gaussian distribution with a deviation of 5 and the mean value only change between block. Hence, participants would be able to gradually learned the mean value of the distribution of each block with trial and error.

One critical factor of this experiment is that while the learning depends on feedback in this task, we induced uncertainty about the feedback itself by explicitly controlling for the feedback reliability between blocks, as to investigate the impact of external uncertainty on learning adjustment. Previous studies have reported that the degree to which learning is updated based on the prediction error on each observed outcome is modulated by the environmental uncertainty (Behrens et al., 2007; McGuire et al., 2014; Nassar et al., 2010). While in most studies, uncertainty is induced in the environmental context (e.g. abrupt change in the mean value of the underlying distribution). Since we have a specific focus on the effect of feedback in learning, we manipulated the uncertainty in the feedback itself instead of in the environment. We expected that learning would be in general improved with more reliable feedback compared to less reliable feedback, also the learning adjustment between trials based on received feedback would be more evident in high feedback reliability block, while participants would be more reserved and updating their behaviour in a slower rate in situation where they consider feedback to be less reliable.

Furthermore, we also examined the impact of internal uncertainty on learning by asking participants to rate their confidence on their learning progress for the block (i.e. how well you think you have learn the wind strength for the current block?) at the beginning of every trial. Different from experiment 2 where the internal uncertainty was

regarding motor execution error, here with a non-motor task, the internal uncertainty is streamed from the state of knowledge (Kahneman & Tversky, 1982). Confidence about the state of knowledge is expected to improve with the number of observations and human learners usually demonstrated a rational measure of confidence where the confidence is linearly correlated with the accuracy and the precision of their estimation (Meyniel, Schlunegger, et al., 2015; Meyniel, Sigman, et al., 2015). Importantly, confidence has been repeatedly found to modulate the impact of observed outcomes on both the neural and behavioural levels. At the neural level, higher confidence has been shown to suppress the response to unexpected outcomes, while at the behavioural level, for an equivalent degree of surprise, the update of stimuli or action values is expected to be smaller with high confidence and larger with low confidence (Meyniel, 2020; Meyniel & Dehaene, 2017). This confidence-weighted learning phenomenon has been observed across various contexts and with different types of stimuli, including probabilistic learning with visual or auditory stimuli (Meyniel, 2020; Meyniel, Schlunegger, et al., 2015), motor learning (Frömer et al., 2021) and value-based decision making (E. Payzan-LeNestour et al., 2013), suggesting that the use of confidence to modulate the effect of observed outcomes in updating learning is a general, modality-free phenomenon. In our experiment, we expect confidence to calibrate with performance. Furthermore, we anticipate that the effect of feedback on learning adjustment will be modulated by the level of confidence. Specifically, when confidence is high, the impact of feedback on the degree of response adjustment is lessened.

Finally, we continue our investigation of the role of feedback in decision-making in this experiment. In experiment 2, we have shown that feedback is a significant factor for participants to decide whether or not they would give a second attempt in a task. Here in experiment 3, we investigated the role of feedback in the decision of exploration or exploitation (Schulz & Gershman, 2019; Wilson et al., 2021). After receiving feedback about their chosen aiming location in our task, participants can freely decide whether to fire a bullet at the selected location, despite the limited number of bullets available in each block. If the bullet hits the target, participants are rewarded. This decision reflects a scenario where participants opt to exploit the chosen option rather than continue exploring other possibilities. While rewards that are less than expected have consistently been shown to drive people to explore other options,

findings on whether uncertainty increases or decreases the propensity for exploration have been inconsistent (Cockburn et al., 2021; Frank et al., 2009; Gershman, 2018; É. Payzan-LeNestour & Bossaerts, 2012). In our study, we investigate to what extent the decision to explore or to stick with the chosen option depends on the received feedback and whether uncertainty about the feedback itself and about participants' learning performance also affects this decision.

### 3.3.2. Methods

#### *Participants*

Data were collected from 6 participants (3 females, 3 males; 5 left-handed; mean age = 25.8, age range = 23 – 29). All participants reported normal hearing and normal or corrected-to-normal vision, and none reported any history of neurological conditions. Written informed consent was obtained and experimental procedures were undertaken in accordance with the Declaration of Helsinki and with the approval by the Comité de Protection des Personnes Ile de France II. Participants received monetary compensation for their participation (10€ per hour, plus reward depend on task performance).

#### *Stimuli and task*

Participants performed a shooting task where they needed to learn the average wind strength in each experimental block to increase their success rate in finding the aiming location that would result in hitting the target. During the introduction of experiment, we asked participants to imagine that they are playing a shooting game in a windy forest. As a result, they cannot simply aim directly at the target but had to adjust their aiming location depending on their estimation of the wind's strength. Participants were told that the wind strength would remain constant within each block and change dramatically between blocks. Therefore, they should try to learn the specific range of wind strength in each block (for example, if the wind strength could be from 0 to 100, the range in a block might be set between 80 and 90). To learn the wind strength, participants have the opportunity to fire a tracer (test shot) on every trial. Afterwards, they would receive feedback indicating whether firing at the selected

location would result in a hit or a miss of the target. Importantly, they were informed that the feedback from the tracer was not always accurate, its accuracy depended on the tracer's quality. If it was a high-quality tracer, the accuracy would be around 80%; if it was a low-quality one, the accuracy would be around 70%. After receiving feedback from the tracer, participants then can use this information to decide whether to fire a real bullet at the selected location, while the number of bullets per block was limited to 25. Only hitting the target with a bullet would result in a reward (two coins). The goal of the task is to earn as much reward as possible with the limited number of bullets, and the reward will be translated into a monetary reward at the end of the experiment.

At the beginning of each block, participants are informed about whether they will have a high- or low-quality tracer for the block. At the beginning of every trial, they have to rate their confidence level on how sure they feel that they have already learn the wind strength for the current block. This rating is done by sliding a small square (visual angle: 0.57°) along a scale bar and the response is confirmed by pressing the “C” key on the keyboard. Afterward, the target (a large square with a fixation cross inside, visual angle: 2.01°) is shown in the center of screen on a scale bar. Participant select the aiming location by sliding a small square (visual angle: 0.57°) on the scale bar. A number is displayed below the target square to indicates the value of the expected wind strength for the chosen location: the number is 0 when the small square is directly below the target, increasing from 0 to -100 as it moves to the left side of the target and from 0 to 100 as it moves to the right side of the target. After deciding on the aiming location, participants confirm their choice by pressing the space bar on the keyboard. For tracer feedback, the outcome is displayed above the selected aiming location, as “Hit!” or “Miss!”. If participants take longer than 10 seconds to select an aiming location, a “Too late!” feedback appears, and the trial end. For the feedback of the actual bullet shot, a successful hit of the target is indicated by an image of two coins, while a miss is marked by a black cross.

The wind strengths in each block are controlled by a Gaussian distribution with varying mean value (Figure 18). The standard deviation of the Gaussian distribution is set at 5. In each trial, the wind strength is randomly drawn from this distribution, resulting in 68% of the wind strength values falling within -5 to +5 of the mean value, and 95% within -10 to +10 of the mean value. The mean value for the wind distribution

in each block was selected from [45, 65, 85], with the restriction that two consecutive blocks could not have the same mean value. Each mean value was used once for a high and once for a low feedback reliability block. This setup allows participants to gradually learn the mean of each block through trial and error. Additionally, we implemented a step-case procedure to accommodate a certain degree of mismatch between the response value and the actual wind strength value for each trial. The maximum mismatch can be up to 10, decreasing by one step after a correct response and increasing by one step after an incorrect response.

All stimuli were presented at the center of the screen against a grey background on a 27-inch, 60 Hz LCD display, positioned 100 cm away from the participants. The experiment consisted of six blocks, with self-paced rests allowed between each block. A block concluded when participants exhausted their number of bullets. The task was ran using Psychotoolbox-3 (Kleiner et al., 2007) on MATLAB. Prior to the experiment, participants received both written and verbal instructions explaining the procedure and completed two practice blocks (one high feedback reliability block and one low feedback reliability block) to familiarize themselves with the task.

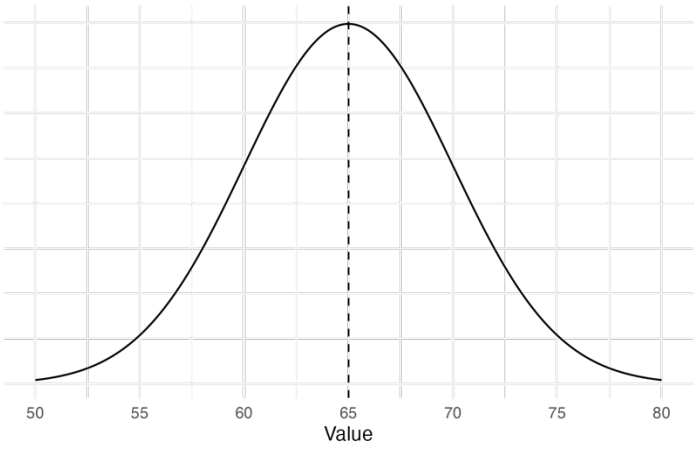



Figure 17. An example of the wind strength distribution.
The mean value of this wind strength distribution is 65 and approximately 68% of the time the wind strength is between 60 to 70, 95% of the time between 55 to 75.

***Procedure***

The procedure of a single trial is illustrated in Figure 19. Each trial begins with a confidence rating scale, where participants rate their confidence level and use the 'C' key to confirm their response. This is followed by a fixation cross presented for 200 ms, after which the target appears in the middle of the screen, on top of a scale bar. Participants then see a moving wind image coming from either the right or the left side of the screen. The purpose of the moving wind image is to indicate the wind direction and to suggest which side of the target the aiming location should be (e.g., if the wind is coming from the left side of the screen towards the right, then the aiming location should be placed on the left side of the target). After the wind image disappears, participants have 10 seconds to select their desired aiming location for the tracer shot and confirm their response with the 'Space bar.' The tracer's feedback appears above the selected aiming location 500 ms after the response is given, with the words 'Hit!' and 'Miss!' indicating whether the tracer hit anything. After 1000 ms, a question 'Fire or not?' appears on the screen, accompanied by the options 'Yes' and 'No.' The remaining number of bullets and the currently earned reward for the block are also displayed. Participants can then decide whether to fire a bullet by clicking 'Yes' or 'No' on the screen. If 'Yes' is selected and the bullet hits the target, an image of two coins is displayed for 1000 ms; if the bullet misses, a black cross is shown instead. If 'No' is selected, a new trial begins. The intertrial interval is 500 ms. Participants are encouraged to maximize their rewards, which will be converted into a monetary reward at the end of the experiment (1 coin = 5 cents).

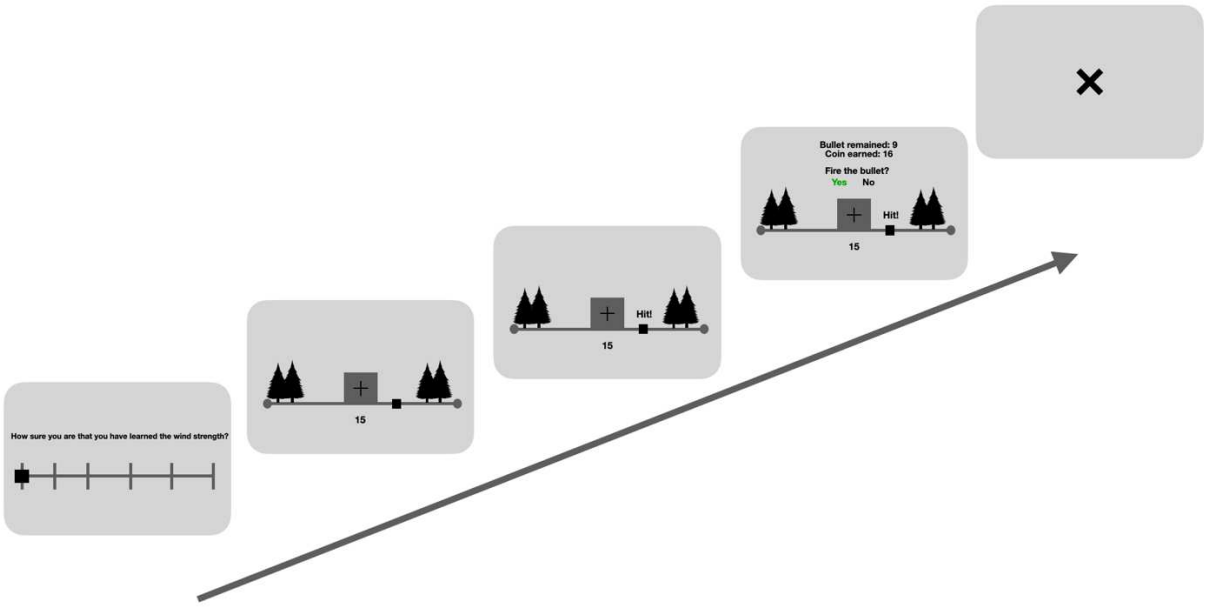


Figure 18. An example of a single trial structure.
In this example trial, positive feedback was received from the tracer and the participant made the decision to fire the bullet but the bullet did not result in a hit of the target.

### *Data analysis*

Trials that exceeded response time limit and aiming location did not follow the indicated wind direction are excluded from the analysis (1.24% of total trials). We labelled trials as good performance trial if the response value is within one standard deviation of the mean value. Successful learning performance is accessed by entering the number of good performance trials into a repeated-measures ANOVA with the factors Time (first / second half of the block) and Block type (high / low feedback reliability). If the number of trials with good performance is significantly larger in the second half of the block compared to the first half, this would indicate successful learning over time. Moreover, we examined possible impact of feedback reliability on learning performance with the factor – Block type.

We conducted linear mixed model analyses to examine the effects of confidence, previous trial feedback, previous trial performance, and block type on the degree of response adjustment (the difference in response value from one trial to the next). Additionally, we explored the impact of previous trial feedback, previous trial performance, and block type on confidence. Lastly, we analysed the decision of exploration (whether to fire a bullet) using a generalized mixed model, with feedback, performance, confidence, and block type as predictors.

The mixed effect model analysis has the advantages that it allows for parametric analyses of single-trial measures and robust to unequally distributed numbers of observations across participants, which is suitable for the current experimental design since the number of trials for each block is not fixed but depended on when does the bullets ran out. Plus, this analysis allows us to use both category and continuous variables as predictors. Furthermore, it take into account of the individual variance regarding experimental effects by including participants as random effect and the predictors as random slopes (Frömer et al., 2018). Variable that explained zero variance were excluded from the random effects structure to prevent overparameterization (Bates et al., 2015; Matuschek et al., 2017). We applied sliding difference contrasts for all the categorical predictors – feedback, performance and

block type. Confidence rating is transformed using max-min scaling (0 to 1) based on each participant data. The models were reduced stepwise by excluding non-significant interaction terms until the respectively smaller model explained the data significantly worse than the larger model. We reported the AIC (Akaike Information Criterion) and BIC (Bayesian Information Criterion), fit indices that are smaller for better fitting models. Statistical analysis were performed using R (R Core Team, 2022) with the ez package and lme4 package (Bates et al., 2015). For the mixed effect models, p-values were computed with the lmerTest package, using Satterthwaite approximation for degrees of freedom. Graphs were made using ggplot2 (Wickham, 2016) and effects package (Fox & Weisberg, 2019).

### 3.3.3. Results

***Learning performance***

Figure 20 displays the changes in error value (the mismatch between the response value and the mean value) averaged across all blocks. To determine whether participants successfully learned the wind strength in each block and if learning performance was impacted by the feedback's reliability for the block, we conducted a repeated measures ANOVA with factors Time (first/second half of the block) and Block Type (high/low feedback reliability) on the number of good performance trials (responses that are within one standard deviation of the mean value). We observed a significant effect for Time, where the number of good performance trials was significantly larger in the second half of the blocks compared to the first half, $F(1,5) = 10.95$, $p = .021$, $\eta 2 = .69$. No significant effect was observed for Block Type, nor was there a significant interaction. Participants' learning performance significantly improved in the second half of the block compared to the first half, suggesting that they successfully learned the wind strength over time, regardless of whether it was in the high or low feedback reliability block.

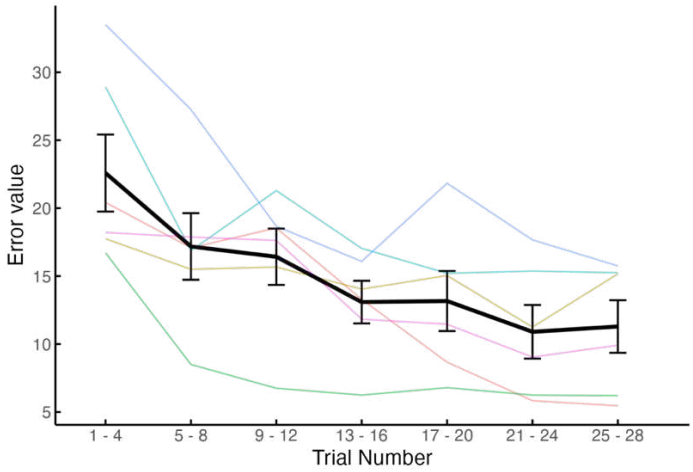



Figure 19. Changes in average error value over trials.
Coloured lines show error value of individual participants averaged over all blocks. Black line shows average error value of all participants, with error bars showing standard error.

### *Response adjustment*

The model estimates for the degree of response adjustment are summarised in Table 3. The final model excluded all the non-significant interactions and the exclusion of those interaction terms did not significantly decrease model fit ($\Delta X^2(6) = 4.98$, $p = .546$). The fit indices were smaller for the reduced model (AIC $_{\text{reduced-full}} = -7$, BIC $_{\text{reduced-full}} = -40$), which indicated a better fit. The significant Confidence x Feedback interaction suggested that the degree of response adjustment is modulated by confidence: participants shifted their response to a larger degree when confidence was low, whereas when confidence was high, the change in response was expected to be smaller. The interaction between confidence and feedback on the degree of response adjustment resulted from feedback having a significant effect on response adjustment only when confidence was relatively low, with negative feedback inducing a larger degree of response adjustment compared to positive feedback (Figure 21).

For the significant interaction between Performance and Block type, and the significant three-way interaction among Performance, Block type, and Confidence on the degree of response adjustment, the results showed that responses are adjusted according to performance. Changes in response were smaller after a good performance trial and larger after a poor performance trial. This effect was only present in blocks with high feedback reliability (Figure 22), suggesting improved learning in blocks with more reliable feedback as participants exhibited greater performance

awareness. Additionally, when performance showed an effect on response adjustment, this effect was also modulated by confidence. Similar to how confidence modulated the effect of feedback, the influence of performance on response adjustment was also only observed when confidence was relatively low, leading to the three-way interaction (Figure 23).

**Table 3.** Effects of confidence, previous trial feedback, previous trial performance, feedback reliability on the degree of response shifting.

| *Predictors* | *Estimates* | *SE* | *t-value* | *p-value* |
|---|---|---|---|---|
| (Intercept) | 7.23 | 1.20 | 6.04 | 0.000586 *** |
| Confidence | -5.77 | 1.72 | -3.35 | **0.012222** * |
| Performance | -1.77 | 0.94 | -1.88 | 0.060659 |
| Feedback | -5.32 | 0.72 | -7.36 | **2.81e-13** *** |
| Block type | -1.66 | 0.88 | -1.88 | 0.060216 |
| Confidence:Performance | 0.11 | 1.47 | 0.07 | 0.941403 |
| Confidence:Feedback | 6.26 | 1.21 | 5.19 | **2.37e-07** *** |
| Confidence:Block type | 2.17 | 1.40 | 1.56 | 0.119579 |
| Performance:Block type | -8.20 | 1.80 | -4.55 | **5.76e-06** *** |
| Confidence:Performance:Block type | 9.27 | 2.81 | 3.30 | **0.001000** *** |
| *Random Effects* | *SD* | | | |
| Intercept | 2.70 | | | |
| Confidence | 3.81 | | | |
| Model Parameters | | | | |
| N | 6 | | | |
| Observations | 1881 | | | |
| Deviance | 13119.4 | | | |
| Log-Likelihood | -6559.7 | | | |

*Formula: Response shifting ~ Feedback* Confidence + Block type *( Performance*Confidence) + (Confidence | participant)*

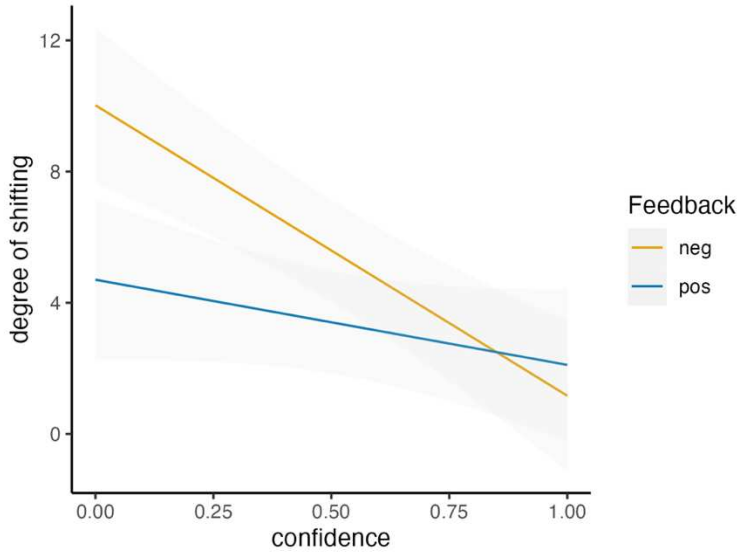



Figure 20. Model estimation of the Feedback by Confidence interaction on the degree of response adjustment. The shaded regions represent 95% confidence intervals.

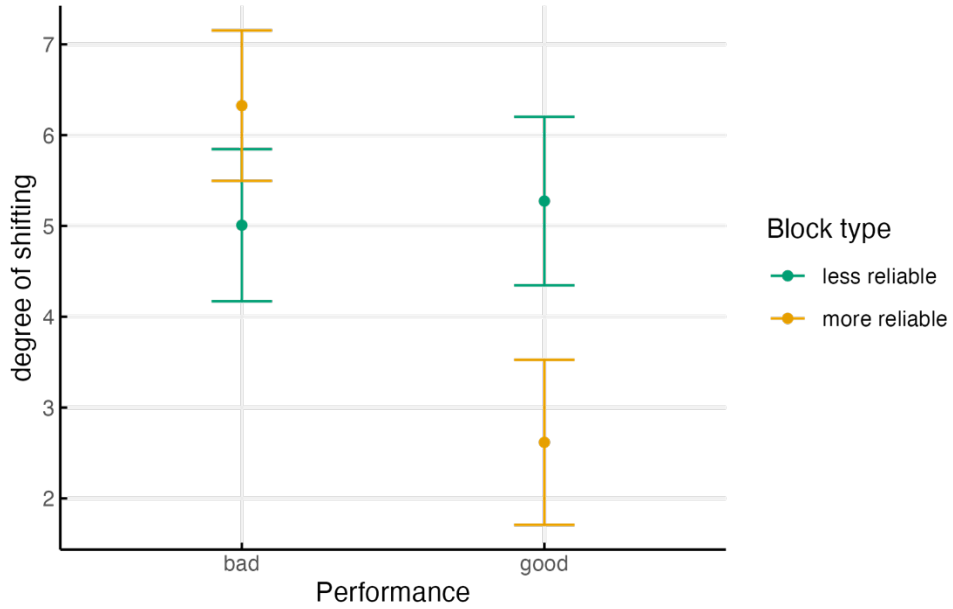



Figure 21. Model estimation of the Performance by Block type interaction on the degree of response shifting. The degree of response shifting between trial is predicted to be larger after bad performance but only low feedback reliability block.

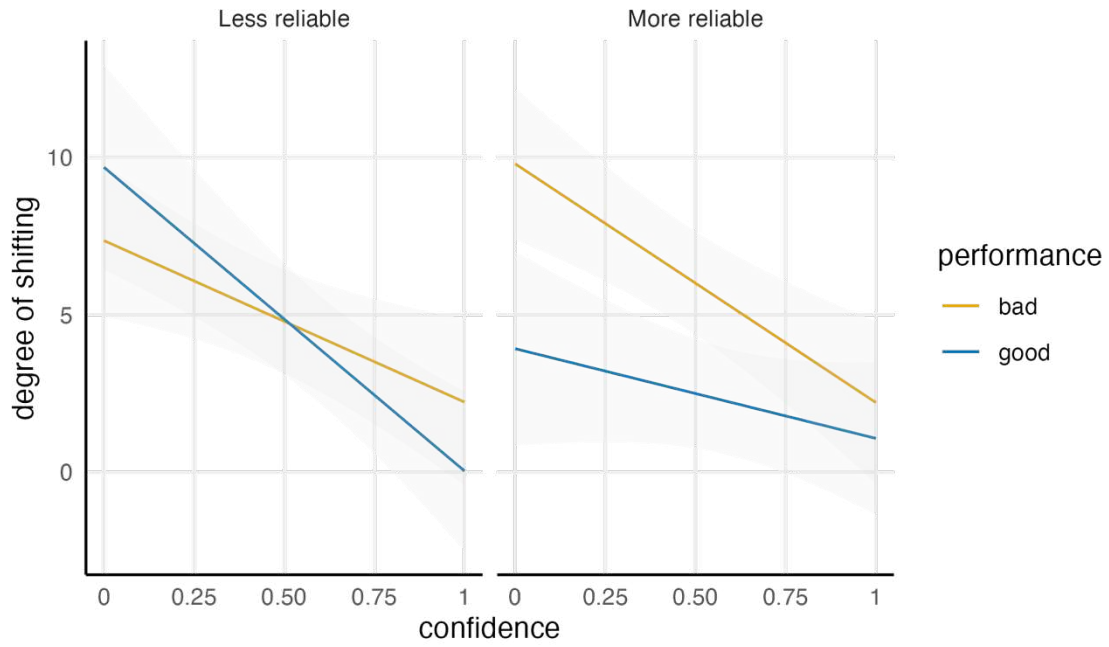



Figure 22. Model estimation of interaction between Performance, Confidence and Block type on the degree of response shifting.

### *Confidence*

The model estimates for confidence ratings are summarized in Table 4. Significant main effects of Performance, Feedback, and Block type were observed, along with a significant Performance x Feedback interaction and a significant three-way interaction among Performance, Feedback, and Block type. Interpreting the results in conjunction with the interaction terms revealed that participants generally exhibited good calibration between subjective confidence ratings and performance, with confidence predicted to be higher following a good performance trial and lower after a poor performance. Feedback also impacted confidence, with negative feedback leading to lower confidence and positive feedback resulting in higher confidence. However, the effect of feedback on confidence was only evident when performance was poor, not when it was good—in which case, feedback did not affect confidence (Figure 24). Furthermore, when feedback did affect confidence, the effect was stronger in blocks with high feedback reliability (Figure 25).

**Table 4.** Effects of previous trial feedback, previous trial performance, feedback reliability on confidence.

| *Predictors* | *Estimates* | *SE* | *t-value* | *p-value* |
|---|---|---|---|---|
| (Intercept) | 0.51 | 0.03 | 15.27 | 5.49e-06 *** |
| Performance | 0.30 | 0.05 | 5.94 | **0.000953** *** |
| Feedback | 0.07 | 0.01 | 5.22 | **1.97e-07** *** |
| Block type | 0.03 | 0.01 | 2.53 | **0.011606** * |
| Performance:Feedback | -0.17 | 0.03 | -6.64 | **4.14e-11** *** |
| Performance:Block type | -0.04 | 0.03 | -1.67 | 0.095780 |
| Feedback:Block type | 0.03 | 0.03 | 1.11 | 0.265952 |
| Performance:Feedback:Block type | -0.17 | 0.05 | -3.38 | **0.000752** *** |
| *Random Effects* | *SD* | | | |
| Intercept | 0.08 | | | |
| Performance | 0.12 | | | |
| Model Parameters | | | | |
| N | 6 | | | |
| Observations | 1881 | | | |
| Deviance | 350.9 | | | |
| Log-Likelihood | -175.4 | | | |

*Formula: Confidence ~ Performance * Feedback * Block type + (Performance | participant)*

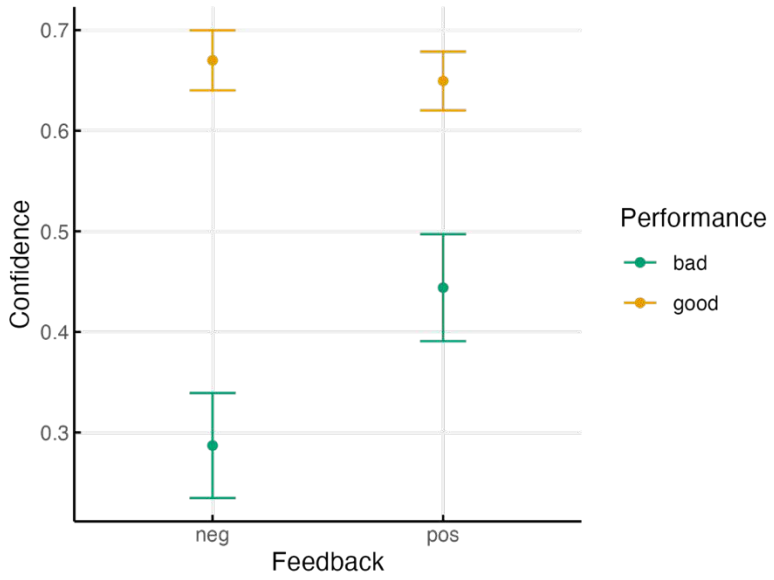



Figure 23. Model estimation of the Feedback by Performance interaction.
The effect of the previous trial feedback and performance on confidence rating. The error bars represent standard error.

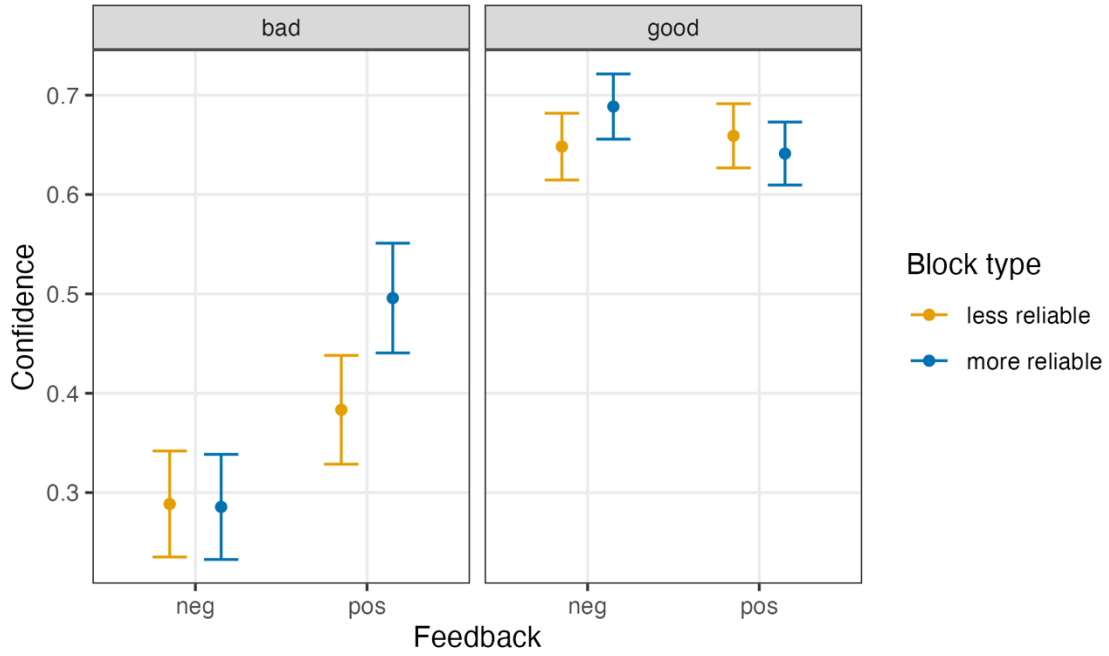



Figure 24. Model estimation of interaction between Performance, Feedback and Block type.
Feedback and Block type only shown significant interaction after bad performance trials. The error bars represent standard error.

### *Decision of exploration*

We used a generalised mixed model to estimate the effect of Confidence, Feedback, Performance and Feedback reliability on participants' decisions. Results are summarised in Table 5. The final model excluded all the non-significant interactions and the exclusion of those interaction terms did not significantly decrease model fit ($\Delta X^2(16) = 22.40$, $p = .131$). The fit indices were smaller for the reduced model (AIC $_{\text{reduced-full}} = -9$, BIC $_{\text{reduced-full}} = -98$). The model's results indicated that participants were more likely to stop exploring and stick with the chosen option for a possible reward after receiving positive feedback and when confidence was high, as well as in blocks with lower feedback reliability. The tendency to explore less in blocks with lower feedback reliability may be attributed to the lower quality of evidence obtained during the sampling process.

**Table 5.** Effects of confidence, feedback, performance and feedback reliability on decision.

| *Predictors* | *Estimates* | *SE* | *z-value* | *p-value* |
|---|---|---|---|---|
| (Intercept) | -0.86 | 1.39 | -0.62 | 0.535587 |
| Feedback | 5.89 | 1.54 | 3.82 | **0.000136** *** |
| Performance | -0.54 | 0.35 | -1.55 | 0.120436 |
| Confidence | 3.10 | 1.22 | 2.54 | **0.011211** * |
| Block type | -0.47 | 0.15 | -3.18 | **0.001489** ** |
| *Random Effects* | *SD* | | | |
| Intercept | 3.29 | | | |
| Feedback | 3.49 | | | |
| Performance | 0.65 | | | |
| Confidence | 2.87 | | | |
| Model Parameters | | | | |
| N | 6 | | | |
| Observations | 1917 | | | |
| Deviance | 1340.1 | | | |
| Log-Likelihood | -670.1 | | | |

*Formula: Decision ~ Feedback + Performance + Confidence + Block type + (Feedback + Performance + Confidence| participant)*

### 3.3.4. Discussion

In this study, we investigated the effect of feedback on learning adjustments under conditions of uncertainty. One source of uncertainty was derived from the feedback itself, which we manipulated by varying the reliability of the feedback. Another source of uncertainty arose from the participants' internal estimations of learning performance. We carefully studied the impact of these two sources of uncertainty on the feedback effect in learning adjustments using a task where participants were required to learn the mean value of wind strength distributions using feedback. We observed how the response value was adjusted after each piece of feedback and in relation to the reported confidence about learning progress in the trial. Moreover, we examined the decision to explore by creating a situation where, in every trial, participants could freely decide whether to fire for a possible reward, with a limited number of bullets available. This created a scenario where participants could choose to continue exploring other options or decide to stick with the previously chosen option.

Firstly, our results demonstrated that participants displayed successful learning over time, with the response value getting closer to the mean value of the distribution

in the second half compared with the first half of the block. Moreover, feedback reliability did not significantly influence learning performance. This absence of any discernible impact of feedback reliability on learning performance may seems surprising, given that previous research has shown that lower feedback reliability lead to reduce learning performance (Di Gregorio et al., 2019; Schiffer et al., 2017; Wurm et al., 2022). This absence of an effect for feedback reliability on learning performance in our results is likely due to our analysis approach, which may have been too insensitive to detect any subtle effects. Specifically, we only ran a straightforward comparison of the number of good performance trials between the first and second half of the blocks, which can simply give us information about whether participants have successfully learned the task. Given that feedback reliability in our study was maintained at approximately 70% even in the low feedback reliability blocks. We expected that participants would be able to learn the task in both the high and low feedback reliability conditions, and any difference between the two condition is probably subtle and require a finer separation of trials into discrete time bins. To uncover the influences of feedback reliability on learning, future analyses will benefit from adopting more sensitive analysis approaches, such as Bayesian learning rate estimation, computational modelling, and to expand the number of trials to better observe learning dynamics across discrete intervals.

In terms of learning adjustment, we found that confidence is a significant factor in predicting the extent to which individuals would change their response from one trial to the next. Low confidence induced a greater change between trials, while under high confidence, responses were more stable and changed to a lesser degree from one trial to the next. More importantly, while we expected that feedback would be a significant factor in determining the degree of response adjustment, where negative feedback would lead to a greater change in response value, this effect of feedback was only observed when confidence was low. However, when confidence was high, response did not adjust based on feedback. This finding is consistent with previous studies which have shown that the effect of outcomes on updating learning or behaviour is modulated by confidence, and a smaller update is expected even for unexpected outcomes when confidence is high (Meyniel, 2020; Meyniel & Dehaene, 2017). This result is also significant in demonstrating that the influence of feedback on learning is not merely a function of prediction error. Furthermore, we discovered that

participants also adjusted their response according to their actual performance. Specifically, if the response value was already within one standard deviation (SD) of the mean of the distribution, then there was little further adjustment in the response value. However, this pattern was only observed in blocks with high feedback reliability. We reasoned that this pattern of results could reflect improved learning under more reliable feedback compared to less reliable feedback, with participants showing greater awareness of their performance and displaying better knowledge in their estimation of the mean value of the distribution. Likewise, the effect of performance on response adjustment also interacts with the level of confidence, where adjustments based on performance occur only when confidence is low, and not when confidence is high. This suggests that confidence generally weights the updating of learning, regardless of the specific input causing the adjustment.

In our analysis of confidence ratings, consistent with previous studies, we found that confidence was calibrated to performance, with higher confidence displayed for better performance and lower confidence after poor performance(Boldt & Yeung, 2015; Frömer et al., 2021; Yeung & Summerfield, 2012). We also observed an effect of feedback on confidence ratings, but this occurred only when performance was poor, which we believe indicated a diminished effect of feedback once the task is learned. Previous studies have observed a reduction in FRN amplitudes in response to feedback in the latter parts of learning tasks, where subjects demonstrated successful learning behaviourally. This indicated that feedback elicited a lesser degree of neural response once it was no longer needed for updating learning (Bellebaum and Daum 2008; Eppinger et al., 2008; Hajcak et al., 2007; Krigolson et al., 2009; Pietschmann et al., 2008). Similarly, in our study, feedback no longer had the same effect on participants' confidence levels once the task was more or less learned, and feedback did not provide as much information compared to the beginning of the task, where learning depended on it. Moreover, when feedback influenced confidence about learning progress at the beginning of the task, the effect of feedback on confidence was more pronounced if the feedback was more reliable. This suggests that participants considered the reliability of the feedback and used this information to appropriately adjust their internal estimation of their performance.

Finally, regarding the decision to explore, we found that participants were more likely to stay with the chosen option when the received feedback was positive, when the confidence rating was high, and with less reliable feedback. The result about the feedback is aligns with previous findings that outcomes less than expected tend to encourage more exploratory behaviour (Schulz & Gershman, 2019; Wilson et al., 2021). In terms of whether uncertainties drive exploration, it is interesting that we found uncertainty about learning performance encouraged participants to continue exploring other options, while uncertainty about feedback discouraged it. We reasoned that increasing the number of observations would help increase confidence about task knowledge, but at the same time, if the feedback reliability is low, then the informational value of each observation decreases. This realization ultimately reduces the tendency to sample more evidence.

Altogether, the results from our study demonstrated that the effect of feedback on learning adjustment is contingent on confidence, with learning more likely to be updated based on feedback only when confidence is low. We also reaffirmed that individuals are capable of developing a subjective confidence measure that closely aligns with their objective performance. Furthermore, feedback has an effect on confidence, but this effect is confined to the beginning of the task when it is not yet well learned. We observed an effect for feedback reliability in the way people adjust their confidence based on feedback, indicating that the information is used to adjust the impact of feedback information in terms of how much to update the internal estimation of performance based on the external feedback. These findings help us to demonstrated that the effect of feedback in learning is not limited perdition error and confidence regulated learning by controlling the degree of response adjustment based on each received outcome.

# 4. GENERAL DISCUSSION

At the beginning of this thesis, we have discussed how the reinforcement learning model, being a powerful tool in discerning the learning process towards optimal behaviours in both animal and human learning. It has significantly enhanced our understanding of the neural mechanisms involved in learning and decision-making by

providing explicit parameters for model-based analysis in complex neural data. However, when it comes to applying this model to feedback learning in humans, certain limitations become evident. The model employs a rather simplistic approach to characterizing feedback – positive if the reward is greater than expected, negative when the reward is less than expected. The influence of feedback on learning is thus examined exclusively from this binary perspective. While this method may be suitable for examining basic learning processes in highly controlled laboratory environments, it may not fully capture the nuances of human feedback learning in real life.

In our daily life, we receive feedback in various forms. Sometimes, it simply contains the outcome as either positive or negative. Other times, it may offer insights into our performance and how it can be improved, if at all. Additionally, we receive internal feedback from our actions, which plays a crucial role in error detection and in our learning performance. Therefore, information from feedback can serves multiple purposes. It can be used to learn options value as suggested in the reinforcement learning model, to improve performance when it contains detailed information about the behaviour itself, or even to guide decisions that are relevant for learning. For example, we might sometime be wondering if we should persist with a task, seek external advice, or explore alternative strategies, and we based those decision on the feedbacks we received. The way human use feedback in learning is flexible and dynamic, varying according to the specific content of the feedback and the individual's objectives.

More importantly, learning in daily lives needs to incorporate the fact that we live in a world that is consistently changing and full of unpredictability. There is a significant amount of evidence showing that the sense of uncertainty shapes the way we learn (Boldt et al., 2019; Gallistel et al., 2014; Meyniel, Schlunegger, et al., 2015; E. Payzan-LeNestour et al., 2013). Humans are naturally rather good at perceiving abrupt changes in the probability of their environment (Gallistel et al., 2014; Meyniel et al., 2015), and more critically, it has been repeatedly reported that perceived uncertainty in the environmental context supports adaptive learning. In this process, we adjust the learning rate that controls the updating of the value of each event/action/stimulus in proportion to the prediction error (Behrens et al., 2007; McGuire et al., 2014; Nassar et al., 2010). This adaptability is essential for optimal learning, as it is beneficial to

adopt a steeper learning rate in an unstable environment in order to keep pace with constant changes in the world. Meanwhile, it is necessary to avoid a high learning rate in a stable environment, since learning the true underlying probability would be slow if we adjust our behaviour to every random fluctuation present in the environment. Moreover, as we adapt our learning based on the uncertainty perceived in the external context, due to the metacognitive abilities of human learners, we also possess a second level of uncertainty related to the accuracy of our estimations at the first level. This second level of uncertainty regarding our internal estimations is commonly expressed as confidence. The sense of confidence has been defined as "a belief about the validity of our own thoughts, knowledge, or performance that relies on a subjective feeling" (Grimaldi et al., 2015). This general feeling of confidence is present in any kind of internal estimation we might have and is not limited to a specific type of task or learning. For example, confidence can be about the estimation of reward probability in a decision-making task or about how different two stimuli are in a visual perception task. Importantly, it has been demonstrated that confidence is a rational measure closely correlated with the accuracy of a prediction, as well as the precision of that prediction. This correlation is then utilized to modulate learning in such a manner that, for a given discrepancy between observation and prediction, the update of the prediction based on the observation is smaller when the confidence associated with the prediction is higher (Meyniel & Dehaene, 2017).

At the conception of this PhD project, we realized that the framework of feedback learning nowadays is still predominantly dominated by the perspective of reinforcement learning models, where feedback simply drives learning by the degree of prediction error it elicits. Hence, one of our goals in this project is to investigate the impact of uncertainty in feedback learning. Moreover, apart from uncertainty, we also recognized that the mechanisms of reinforcement learning are still poorly understood in the context of motor skill acquisition. Motor learning differs from non-motor learning tasks in the way that we receive feedback about our motor execution error and already have a prediction of our own action-effect from the internal model of motor executions (Blakemore et al., 2000; Hommel et al., 2001; Prinz, 1997; Wolpert et al., 2011). Previous studies have found that humans are able to discount their motor execution error when updating the value of their decision based on the received outcome. Specifically, the value of the decision does not decrease after an undesired negative

outcome if it can be attributed to a motor error. Furthermore, the neural response for prediction error is also found to be suppressed if the outcome can be associated with a failure in motor execution (McDougle et al., 2016, 2019). While external feedback is helpful and commonly used to improve performance in motor learning tasks, we are still in need of a more comprehensive framework that informs us of the impact of internal motor feedback on the processing of external feedback and how it may affect the way we apply external feedback to improve motor skills. Additionally, we recognize that beyond its role in learning adjustment, feedback plays a crucial part in subsequent decision-making. This includes decisions on whether to continue or give up on learning, as well as exploring other potential options. Therefore, in this project, we also explored the role of feedback in decision-making, taking into account how these decisions could be influenced not only by feedback but also by confidence and internal motor feedback.

In our first experiment, we quantified people's ability to generate predictions of the sensory outcomes caused by their own actions and demonstrated that the generation of action-effect predictions is evident across different types of actions. Individuals were able to predict the effects of their own actions equally well, whether the effect was associated with the selection of action or the timing of performing an action. Following this, we conducted a second experiment to investigate the impact of internal motor feedback and external feedback on the decision to attempt a sensorimotor task a second time. The results showed that external feedback significantly influenced the decision on whether to continue learning. Internal motor feedback also had a significant effect on this decision-making process, although its impact was not as consistent as that of external feedback. We observed that the influence of performance on the decision seemed to depend on the precision of the internal prediction of motor performance. Specifically, an impact of motor performance on the decision was observed only when there was an increased amplitude of the feedback-related negativity (FRN) following the presentation of feedback. The increased amplitude of FRN in this context is likely indicative of a more precise prediction of performance. A more precise prediction typically led to a heightened prediction error response toward the observed outcome (Press et al., 2020). Additionally, previous findings have also reported a diminished effect of surprise, as reflected by a smaller FRN amplitude, when there was an absence of clear expectations (Hayden et al., 2011). While we did not ask participants to reported their

estimation of performance and their confidence about their estimation. We believe that the result where only a precise performance prediction impacted the neural signal of feedback processing and decision-making behaviour serves as supporting evidence for the confidence weighting principle in learning regulation. According to this principle, for a given surprise, the update is smaller when the confidence about the prediction is higher (Meyniel, 2020). Confidence should reflect the precision of an estimate, characterized by whether the distribution is spread (indicating low confidence) or concentrated (indicating high confidence) around the estimate. In practice, confidence is often formalized as the precision of the probability distribution of a prediction (its inverse variance), a formulation that aligns well with individuals' self-reported confidence (Friston, 2009; Meyniel, Schlunegger, et al., 2015; Meyniel & Dehaene, 2017), Our result of observing an impact when the prediction of motor performance is more precise could suggests that a higher confidence level is assigned to this prediction, even though we did not explicitly inquire about it in the task. This influenced the effect of feedback, where participants decided to base their decision on their internal estimation of performance instead of external feedback. On the other way around, when the precision of the prediction about performance is low, individuals tend to rely more on external feedback to make their decision.

In experiment 3, we gained a clearer understanding of how confidence is used to regulate learning from feedback. In this experiment, we explored the impact of subjective confidence about performance and feedback on learning adjustment, as well as the decision to explore. To address the limitation in experiment 2, where we did not ask participants to report their confidence about their performance, in this task, participants were required to report their confidence level about their learning progress for the block at the beginning of every trial. Furthermore, we addressed another limitation from experiment 2, where we were only able to study the effect of feedback and performance on decision-making but not on learning rate or response adjustment, due to participants being limited to repeating the same task at most twice. To overcome this, we employed a learning task where participants needed to learn the mean value of an underlying distribution using the feedback received on every trial. Specifically, participants were instructed to imagine they were playing a shooting game in a forest with variable wind conditions, where the wind strength in each block was relatively consistent, governed by a normal distribution. In this scenario, if participants

learned to aim at a location where the distance between the aiming location and the target closely matched the mean value of the wind strength distribution, then their shots would have a higher probability of hitting the target. In this task, we observed that the effect of feedback on learning adjustment is modulated by subjective confidence about learning performance, where negative feedback only induced greater response adjustment when confidence was low but had no effect when confidence was high. Moreover, in this task, we investigated not only the impact of internal uncertainty about task performance but also external uncertainty about the feedback itself by manipulating feedback reliability. While we did not observe a clear effect of feedback reliability on learning adjustment, feedback reliability had an impact on how strongly feedback influenced participants' confidence about their learning performance at the beginning of the task. More reliable feedback influenced confidence more compared to less reliable feedback. However, feedback stopped having any impact on confidence once the task was more thoroughly learned. For the decision of exploration, we discovered that participants were more inclined to cease exploring and stick with the chosen option when they received positive feedback, when their confidence rating was high, and notably, when feedback reliability was low. This observation suggests that negative feedback and high uncertainty regarding learning performance prompt exploration. Furthermore, with high feedback reliability, the informational value of each observation increased compared to scenarios with low feedback reliability, this potentially heightened the propensity to use exploration as a means to sample the reward probability of other possible options.

Our studies demonstrated significant impact of feedback on both learning adjustments and subsequent decision-making. This is particularly evident in decisions about whether to persist with a task or abandon it, and when to continue exploring or to stop. A key objective of this PhD project was to explore how uncertainty affects the processing of feedback and its subsequent impact on learning adjustments. Our findings indicate that an individual's confidence in their internal assessment of task performance heavily influences the effect of feedback on learning adjustments. Feedback plays a significant role in modifying learning adjustments primarily when confidence levels are low. Conversely, at high levels of confidence, the impact of feedback on learning is markedly reduced. This underscores the pivotal role of confidence in bridging feedback and learning adjustments, highlighting that an

individual's perception of their performance is crucial in determining how feedback influences their learning process. Importantly, this illuminates that feedback's role in learning extends beyond mere prediction error, with confidence acting as a key regulator by controlling the extent of learning adjustment following feedback.

While confidence plays a significant role in regulating learning, it's important to note that the term "confidence" is used broadly in the literature, yet it represents slightly different concepts depending on the nature of the internal estimation being evaluated. For instance, in our second experiment, where the internal estimation relates to motor execution, confidence does not linearly correlate with performance. In this context, one can be highly confident in making a motor error, as well as in executing a movement perfectly. Here, the relationship between confidence and performance is more U-shaped. Conversely, in our third experiment, confidence is tied to the internal estimation of state knowledge. In this scenario, confidence linearly correlates with the number of observations (or the amount of received feedback in our case), meaning higher confidence always suggests better performance. This nuanced understanding of what confidence represents in different task contexts is crucial for dissecting the specific ways feedback effects are modulated. Furthermore, a key distinction between an internal estimation of motor performance and the estimation of the state of knowledge is the reliance on external feedback. In tasks involving motor performance, such as in Experiment 2 where feedback was predominantly random, we observed participants making decisions based on their performance rather than on feedback. This suggests that confidence assignment in motor tasks does not necessarily depend on external feedback. Conversely, for estimations regarding the state of task knowledge, external feedback, or observed outcomes, become the primary source for constructing our performance and confidence estimates. Therefore, we might expect internal estimations regarding motor performance to be more pronounced against environmental uncertainty and uncertainty related to the feedback itself. For non-motor tasks, feedback is expected to interact more bidirectionally with confidence. This interaction was observed in the results of Experiment 3, where feedback significantly influenced the level of confidence at the beginning of the task. Only when the performance estimation began to stabilize due to an increased number of observations did feedback start to impact confidence less. Simultaneously, the impact of feedback was strengthened when participants knew the feedback was more reliable. This

dynamic showcases how feedback and confidence interact differently, depending on whether the task involves motor performance or the estimation of state knowledge.

## 4.1. Conclusion

While the reinforcement learning framework offers a robust basis for grasping how feedback facilitates learning through prediction errors, there's a pressing need to expand upon it. We should strive to develop a more nuanced framework that integrates factors ubiquitous in our daily lives and crucial for information processing and learning. Such an enriched framework would significantly enhance our ability to translate laboratory findings into practical real-world applications. By acknowledging and incorporating these everyday factors, we can gain a deeper and more accurate understanding of learning processes as they occur in natural settings, paving the way for more effective learning strategies and interventions.

# 5. BIBLIOGRAPHY


Acharya, J. N., Hani, A. J., Cheek, J., Thirumala, P., & Tsuchida, T. N. (2016). American Clinical Neurophysiology Society Guideline 2: Guidelines for Standard Electrode Position Nomenclature. *The Neurodiagnostic Journal*, *56*(4), 245–252. https://doi.org/10.1080/21646821.2016.1245558

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

Arias-Carrión, O., Stamelou, M., Murillo-Rodríguez, E., Menéndez-González, M., & Pöppel, E. (2010). Dopaminergic reward system: A short integrative review. *International Archives of Medicine*, *3*(1), 24. https://doi.org/10.1186/1755-7682-3-24

Baayen, R. H. (2008). *Analyzing Linguistic Data: A Practical Introduction to Statistics using R*. Cambridge University Press. https://doi.org/10.1017/CBO9780511801686

Baayen, R. H., & Milin, P. (2010). *Analyzing reaction times*. https://doi.org/10.21500/20112084.807

Baess, P., Jacobsen, T., & Schröger, E. (2008). Suppression of the auditory N1 event-related potential component with unpredictable self-initiated tones: Evidence for internal forward models with dynamic stimulation. *International Journal of Psychophysiology*, *70*(2), 137–143. https://doi.org/10.1016/j.ijpsycho.2008.06.005

Baess, P., Widmann, A., Roye, A., Schröger, E., & Jacobsen, T. (2009). Attenuated human auditory middle latency response and evoked 40-Hz response to self-

initiated sounds. *European Journal of Neuroscience*, *29*(7), 1514–1521. https://doi.org/10.1111/j.1460-9568.2009.06683.x

Baldeweg, T. (2007). ERP Repetition Effects and Mismatch Negativity Generation: A Predictive Coding Perspective. *Journal of Psychophysiology*, *21*(3–4), 204–213. https://doi.org/10.1027/0269-8803.21.34.204

Baldeweg, T., Klugman, A., Gruzelier, J., & Hirsch, S. R. (2004). Mismatch negativity potentials and cognitive impairment in schizophrenia. *Schizophrenia Research*, *69*(2–3), 203–217. https://doi.org/10.1016/j.schres.2003.09.009

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

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

Bayer, H. M., & Glimcher, P. W. (2005). Midbrain Dopamine Neurons Encode a Quantitative Reward Prediction Error Signal. *Neuron*, *47*(1), 129–141. https://doi.org/10.1016/j.neuron.2005.05.020

Bayer, H. M., Lau, B., & Glimcher, P. W. (2007). Statistics of midbrain dopamine neuron spike trains in the awake primate. *Journal of Neurophysiology*, *98*(3), 1428–1439.

Behrens, T. E. J., Woolrich, M. W., Walton, M. E., & Rushworth, M. F. S. (2007). Learning the value of information in an uncertain world. *Nature Neuroscience*, *10*(9), 1214–1221. https://doi.org/10.1038/nn1954

Bell, A. J., & Sejnowski, T. J. (1995). *An information-maximisation approach to blind separation and blind deconvolution*. 38.

Bendixen, A., SanMiguel, I., & Schröger, E. (2012). Early electrophysiological indicators for predictive processing in audition: A review. *International Journal of Psychophysiology*, *83*(2), 120–131. https://doi.org/10.1016/j.ijpsycho.2011.08.003

Berns, G. S., McClure, S. M., Pagnoni, G., & Montague, P. R. (2001). Predictability modulates human brain response to reward. *Journal of Neuroscience*, *21*(8), 2793–2798.

Björklund, A., & Dunnett, S. B. (2007). Dopamine neuron systems in the brain: An update. *Trends in Neurosciences*, *30*(5), 194–202.

Blakemore, S.-J., Wolpert, D., & Frith, C. (2000). Why can't you tickle yourself? *Neuroreport*, *11*(11), 11–16.

Boldt, A., Schiffer, A.-M., Waszak, F., & Yeung, N. (2019). Confidence Predictions Affect Performance Confidence and Neural Preparation in Perceptual Decision Making. *Scientific Reports*, *9*(1), 4031. https://doi.org/10.1038/s41598-019-40681-9

Boldt, A., & Yeung, N. (2015). Shared Neural Markers of Decision Confidence and Error Detection. *The Journal of Neuroscience*, *35*(8), 3478–3484. https://doi.org/10.1523/JNEUROSCI.0797-14.2015

Bounmy, T., Eger, E., & Meyniel, F. (2023). A characterization of the neural representation of confidence during probabilistic learning. *NeuroImage*, *268*, 119849. https://doi.org/10.1016/j.neuroimage.2022.119849

Brass, M., & Haggard, P. (2008). The What, When, Whether Model of Intentional Action. *The Neuroscientist*, *14*(4), 319–325. https://doi.org/10.1177/1073858408317417

Briggs, K. E., & Martin, F. H. (2009). Affective picture processing and motivational relevance: Arousal and valence effects on ERPs in an oddball task. *International Journal of Psychophysiology*, *72*(3), 299–306.

Bürkner, P.-C. (2017). brms: An R Package for Bayesian Multilevel Models Using Stan. *Journal of Statistical Software*, *80*(1), Article 1. https://doi.org/10.18637/jss.v080.i01

Butterfield, B., & Mangels, J. A. (2003). Neural correlates of error detection and correction in a semantic retrieval task. *Cognitive Brain Research*, *17*(3), 793–817. https://doi.org/10.1016/S0926-6410(03)00203-9

Carlebach, N., & Yeung, N. (2023). Flexible use of confidence to guide advice requests. *Cognition*, *230*, 105264. https://doi.org/10.1016/j.cognition.2022.105264

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., & Riddell, A. (2017). Stan: A Probabilistic Programming Language. *Journal of Statistical Software*, *76*(1), Article 1. https://doi.org/10.18637/jss.v076.i01

Carrillo-de-la-Peña, M. T., & Cadaveira, F. (2000). The effect of motivational instructions on P300 amplitude. *Neurophysiologie Clinique/Clinical Neurophysiology*, *30*(4), 232–239.

Cavanagh, J. F., Figueroa, C. M., Cohen, M. X., & Frank, M. J. (2012). Frontal Theta Reflects Uncertainty and Unexpectedness during Exploration and Exploitation. *Cerebral Cortex*, *22*(11), 2575–2586. https://doi.org/10.1093/cercor/bhr332

Cavanagh, J. F., Frank, M. J., Klein, T. J., & Allen, J. J. B. (2010). Frontal theta links prediction errors to behavioral adaptation in reinforcement learning.

*NeuroImage*, *49*(4), 3198–3209. https://doi.org/10.1016/j.neuroimage.2009.11.080

Cavanagh, J. F., Zambrano-Vazquez, L., & Allen, J. J. B. (2012). Theta lingua franca: A common mid-frontal substrate for action monitoring processes. *Psychophysiology*, *49*(2), 220–238. https://doi.org/10.1111/j.1469-8986.2011.01293.x

Chase, H. W., Swainson, R., Durham, L., Benham, L., & Cools, R. (2011). Feedback-related Negativity Codes Prediction Error but Not Behavioral Adjustment during Probabilistic Reversal Learning. *Journal of Cognitive Neuroscience*, *23*(4), 936–946. https://doi.org/10.1162/jocn.2010.21456

Chaumon, M., Bishop, D. V. M., & Busch, N. A. (2015). A practical guide to the selection of independent components of the electroencephalogram for artifact correction. *Journal of Neuroscience Methods*, *250*, 47–63. https://doi.org/10.1016/j.jneumeth.2015.02.025

Chiviacowsky, S., & Wulf, G. (2007). Feedback After Good Trials Enhances Learning. *Research Quarterly for Exercise and Sport*, *78*(2), 40–47. https://doi.org/10.1080/02701367.2007.10599402

Cockburn, J., Man, V., Cunningham, W., & O'Doherty, J. P. (2021). *Novelty and uncertainty interact to regulate the balance between exploration and exploitation in the human brain* [Preprint]. Neuroscience. https://doi.org/10.1101/2021.10.13.464279

Cohen, M. X. (2011). Error-related medial frontal theta activity predicts cingulate-related structural connectivity. *NeuroImage*, *55*(3), 1373–1383. https://doi.org/10.1016/j.neuroimage.2010.12.072

Cohen, M. X., & Ranganath, C. (2007). Reinforcement Learning Signals Predict Future Decisions. *The Journal of Neuroscience*, *27*(2), 371–378. https://doi.org/10.1523/JNEUROSCI.4421-06.2007

Cohen, M. X., Wilmes, K. A., & Van De Vijver, I. (2011). Cortical electrophysiological network dynamics of feedback learning. *Trends in Cognitive Sciences*, *15*(12), 558–566. https://doi.org/10.1016/j.tics.2011.10.004

Costa-Faidella, J., Baldeweg, T., Grimm, S., & Escera, C. (2011). Interactions between 'What' and 'When' in the Auditory System: Temporal Predictability Enhances Repetition Suppression. *Journal of Neuroscience*, *31*(50), 18590–18597. https://doi.org/10.1523/JNEUROSCI.2599-11.2011

Darriba, Á., & Waszak, F. (2018). Predictions through evidence accumulation over time. *Scientific Reports*, *8*(1), 494. https://doi.org/10.1038/s41598-017-18802-z

Daw, N. D., & Doya, K. (2006). The computational neurobiology of learning and reward. *Current Opinion in Neurobiology*, *16*(2), 199–204. https://doi.org/10.1016/j.conb.2006.03.006

Daw, N. D., Gershman, S. J., Seymour, B., Dayan, P., & Dolan, R. J. (2011). Model-Based Influences on Humans' Choices and Striatal Prediction Errors. *Neuron*, *69*(6), 1204–1215. https://doi.org/10.1016/j.neuron.2011.02.027

Dayan, P., & Sejnowski, T. J. (1996). Exploration bonuses and dual control. *Machine Learning*, *25*, 5–22.

Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*(1), 9–21. https://doi.org/10.1016/j.jneumeth.2003.10.009

Desender, K., Murphy, P., Boldt, A., Verguts, T., & Yeung, N. (2019). A Postdecisional Neural Marker of Confidence Predicts Information-Seeking in Decision-Making. *The Journal of Neuroscience*, *39*(17), 3309–3319. https://doi.org/10.1523/JNEUROSCI.2620-18.2019

Deumens, R., Blokland, A., & Prickaerts, J. (2002). Modeling Parkinson's disease in rats: An evaluation of 6-OHDA lesions of the nigrostriatal pathway. *Experimental Neurology*, *175*(2), 303–317.

Di Gregorio, F., Ernst, B., & Steinhauser, M. (2019). Differential effects of instructed and objective feedback reliability on feedback-related brain activity. *Psychophysiology*, *56*(9), e13399. https://doi.org/10.1111/psyp.13399

Donaldson, K. R., Oumeziane, B. A., Hélie, S., & Foti, D. (2016). The temporal dynamics of reversal learning: P3 amplitude predicts valence-specific behavioral adjustment. *Physiology & Behavior*, *161*, 24–32.

Elsner, B., & Hommel, B. (2001). Effect anticipation and action control. *Journal of Experimental Psychology: Human Perception and Performance*, *27*, 229–240.

Fiorillo, C. D., Tobler, P. N., & Schultz, W. (2003). Discrete coding of reward probability and uncertainty by dopamine neurons. *Science*, *299*(5614), 1898–1902.

Fox, J., & Weisberg, S. (2019). *An R Companion to Applied Regression* (3rd ed.). Sage, Thousand Oaks CA.

Frank, M. J., Doll, B. B., Oas-Terpstra, J., & Moreno, F. (2009). Prefrontal and striatal dopaminergic genes predict individual differences in exploration and exploitation. *Nature Neuroscience*, *12*(8), 1062–1068. https://doi.org/10.1038/nn.2342

Frank, M. J., Woroch, B. S., & Curran, T. (2005). Error-Related Negativity Predicts Reinforcement Learning and Conflict Biases. *Neuron*, *47*(4), 495–501. https://doi.org/10.1016/j.neuron.2005.06.020

Franken, I. H., Van Strien, J. W., Bocanegra, B. R., & Huijding, J. (2011). The P3 event-related potential as an index of motivational relevance. *Journal of Psychophysiology*.

Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *360*(1456), 815–836. https://doi.org/10.1098/rstb.2005.1622

Friston, K. (2009). The free-energy principle: A rough guide to the brain? *Trends in Cognitive Sciences*, *13*(7), 293–301. https://doi.org/10.1016/j.tics.2009.04.005

Frömer, R., Maier, M., & Abdel Rahman, R. (2018). Group-Level EEG-Processing Pipeline for Flexible Single Trial-Based Analyses Including Linear Mixed Models. *Frontiers in Neuroscience*, *12*, 48. https://doi.org/10.3389/fnins.2018.00048

Frömer, R., Nassar, M. R., Bruckner, R., Stürmer, B., Sommer, W., & Yeung, N. (2021). Response-based outcome predictions and confidence regulate feedback processing and learning. *eLife*, *10*, e62825. https://doi.org/10.7554/eLife.62825

Gallistel, C. R., Krishan, M., Liu, Y., Miller, R., & Latham, P. E. (2014). The Perception of Probability. *Psychological Review*, *121*, 99–123.

Garrido, M. I., Kilner, J. M., Stephan, K. E., & Friston, K. J. (2009). The mismatch negativity: A review of underlying mechanisms. *Clinical Neurophysiology*, *120*(3), 453–463. https://doi.org/10.1016/j.clinph.2008.11.029

Gershman, S. J. (2018). Deconstructing the human algorithms for exploration. *Cognition*, *173*, 34–42. https://doi.org/10.1016/j.cognition.2017.12.014

Greenwald, A. G. (1970). Sensory feedback mechanisms in performance control: With special reference to the ideo-motor mechanism. *Psychological Review*, *77*(2), 73–99. https://doi.org/10.1037/h0028689

Haenschel, C. (2005). Event-Related Brain Potential Correlates of Human Auditory Sensory Memory-Trace Formation. *Journal of Neuroscience*, *25*(45), 10494–10501. https://doi.org/10.1523/JNEUROSCI.1227-05.2005

Hayden, B. Y., Heilbronner, S. R., Pearson, J. M., & Platt, M. L. (2011). Surprise Signals in Anterior Cingulate Cortex: Neuronal Encoding of Unsigned Reward Prediction Errors Driving Adjustment in Behavior. *The Journal of Neuroscience*, *31*(11), 4178–4187. https://doi.org/10.1523/JNEUROSCI.4652-10.2011

Hillyard, S. A., Hink, R. F., Schwent, V. L., & Picton, T. W. (1973). Electrical Signs of Selective Attention in the Human Brain. *Science*, *182*(4108), 177–180. https://doi.org/10.1126/science.182.4108.177

Hillyard, S. A., Squires, K. C., Bauer, J. W., & Lindsay, P. H. (1971). Evoked Potential Correlates of Auditory Signal Detection. *Science*, *172*(3990), 1357–1360. https://doi.org/10.1126/science.172.3990.1357

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

Holroyd, C. B., & Krigolson, O. E. (2007). Reward prediction error signals associated with a modified time estimation task. *Psychophysiology*, *44*(6), 913–917. https://doi.org/10.1111/j.1469-8986.2007.00561.x

Holroyd, C. B., Krigolson, O. E., Baker, R., Lee, S., & Gibson, J. (2009). When is an error not a prediction error? An electrophysiological investigation. *Cognitive, Affective, & Behavioral Neuroscience*, *9*(1), 59–70. https://doi.org/10.3758/CABN.9.1.59

Holroyd, C. B., Larsen, J. T., & Cohen, J. D. (2004). Context dependence of the event-related brain potential associated with reward and punishment. *Psychophysiology*, *41*(2), 245–253. https://doi.org/10.1111/j.1469-8986.2004.00152.x

Holroyd, C. B., Nieuwenhuis, S., Yeung, N., & Cohen, J. D. (2003). Errors in reward prediction are reflected in the event-related brain potential. *Neuroreport*, *14*(18), 2481–2484.

Holroyd, C. B., Pakzad-Vaezi, K. L., & Krigolson, O. E. (2008). The feedback correct-related positivity: Sensitivity of the event-related brain potential to unexpected positive feedback. *Psychophysiology*, *45*(5), 688–697. https://doi.org/10.1111/j.1469-8986.2008.00668.x

Holroyd, C. B., & Yeung, N. (2012). Motivation of extended behaviors by anterior cingulate cortex. *Trends in Cognitive Sciences*, *16*(2), 122–128. https://doi.org/10.1016/j.tics.2011.12.008

Hommel, B., Müsseler, J., Aschersleben, G., & Prinz, W. (2001). The Theory of Event Coding (TEC): A framework for perception and action planning. *Behavioral and Brain Sciences*, *24*(5), 849–878. https://doi.org/10.1017/S0140525X01000103

Horváth, J. (2015). Action-related auditory ERP attenuation: Paradigms and hypotheses. *Brain Research*, *1626*, 54–65. https://doi.org/10.1016/j.brainres.2015.03.038

Hsu, Y.-F., Bars, S. L., Hämäläinen, J. A., & Waszak, F. (2015). Distinctive Representation of Mispredicted and Unpredicted Prediction Errors in Human Electroencephalography. *Journal of Neuroscience*, *35*(43), 14653–14660. https://doi.org/10.1523/JNEUROSCI.2204-15.2015

Hughes, G., Desantis, A., & Waszak, F. (2013). Attenuation of auditory N1 results from identity-specific action-effect prediction. *European Journal of Neuroscience*, *37*(7), 1152–1158. https://doi.org/10.1111/ejn.12120

Hughes, G., & Waszak, F. (2011). ERP correlates of action effect prediction and visual sensory attenuation in voluntary action. *NeuroImage*, *56*(3), 1632–1640. https://doi.org/10.1016/j.neuroimage.2011.02.057

Jacobsen, T., & Schröger, E. (2001). Is there pre-attentive memory-based comparison of pitch? *Psychophysiology*, *38*(4), 723–727. https://doi.org/10.1111/1469-8986.3840723

Javitt, D., Steinschneider, M., Schroeder, C., & Arezzo, J. (1996). Role of cortical N-methyl-D-aspartate receptors in auditory sensory memory and mismatch negativity generation: Implications for schizophrenia. *Proceedings of the National Academy of Sciences of the USA*, *93*, 11962–11967.

Joel, D., Niv, Y., & Ruppin, E. (2002). Actor–critic models of the basal ganglia: New anatomical and computational perspectives. *Neural Networks*, *15*(4–6), 535–547.

Kahneman, D., & Tversky, A. (1982). Variants of uncertainty. *Cognition*, *11*, 143–157. https://doi.org/10.1016/0010- 0277(82)90023-3

Karayanidis, F., Robaey, P., Bourassa, M., de Koning, D., Geoffroy, G., & Pelletier, G. (2000). ERP differences in visual attention processing between attention-deficit hyperactivity disorder and control boys in the absence of performance differences. *Psychophysiology*, *37*(3), 319–333.

Kleiner, M., Brainard, D., & Pelli, D. (2007). What's new in Psychtoolbox-3? *Perception*, *36*, 1–16.

Knutson, B., Fong, G. W., Adams, C. M., Varner, J. L., & Hommer, D. (2001). Dissociation of reward anticipation and outcome with event-related fMRI. *Neuroreport*, *12*(17), 3683–3687.

Knutson, B., Westdorp, A., Kaiser, E., & Hommer, D. (2000). FMRI visualization of brain activity during a monetary incentive delay task. *Neuroimage*, *12*(1), 20–27.

Kononowicz, T. W., Roger, C., & Van Wassenhove, V. (2019). Temporal Metacognition as the Decoding of Self-Generated Brain Dynamics. *Cerebral Cortex*, *29*(10), 4366–4380. https://doi.org/10.1093/cercor/bhy318

Kononowicz, T. W., & Van Wassenhove, V. (2019). Evaluation of Self-generated Behavior: Untangling Metacognitive Readout and Error Detection. *Journal of Cognitive Neuroscience*, *31*(11), 1641–1657. https://doi.org/10.1162/jocn_a_01442

Korka, B., Schröger, E., & Widmann, A. (2019). Action Intention-based and Stimulus Regularity-based Predictions: Same or Different? *Journal of Cognitive Neuroscience*, *31*(12), 1917–1932. https://doi.org/10.1162/jocn_a_01456

Korka, B., Schröger, E., & Widmann, A. (2021). The encoding of stochastic regularities is facilitated by action-effect predictions. *Scientific Reports*, *11*(1), 6790. https://doi.org/10.1038/s41598-021-86095-4

Korka, B., Widmann, A., Waszak, F., Darriba, Á., & Schröger, E. (2021). The auditory brain in action: Intention determines predictive processing in the auditory system—A review of current paradigms and findings. *Psychonomic Bulletin & Review*. https://doi.org/10.3758/s13423-021-01992-z

Krieghoff, V., Brass, M., Prinz, W., & Waszak, F. (2009). Dissociating what and when of intentional actions. *Frontiers in Human Neuroscience*, *3*. https://doi.org/10.3389/neuro.09.003.2009

Kühn, S., & Brass, M. (2010). Planning not to do something: Does intending not to do something activate associated sensory consequences? *Cognitive, Affective, & Behavioral Neuroscience*, *10*(4), 454–459. https://doi.org/10.3758/CABN.10.4.454

Kühn, S., Gevers, W., & Brass, M. (2009). The Neural Correlates of Intending Not to Do Something. *Journal of Neurophysiology*, *101*(4), 1913–1920. https://doi.org/10.1152/jn.90994.2008

Kühn, S., Seurinck, R., Fias, W., & Waszak, F. (2010). The internal anticipation of sensory action effects: When action induces FFA and PPA activity. *Frontiers in Human Neuroscience*. https://doi.org/10.3389/fnhum.2010.00054

Le Bars, S., Darriba, Á., & Waszak, F. (2019). Event-related brain potentials to self-triggered tones: Impact of action type and impulsivity traits. *Neuropsychologia*, *125*, 14–22. https://doi.org/10.1016/j.neuropsychologia.2019.01.012

Luck, S. J., & Kappenman, E. S. (2012). ERP components and selective attention. *The Oxford Handbook of Event-Related Potential Components*, 295–327.

Luft, C. D. B., Takase, E., & Bhattacharya, J. (2014). Processing Graded Feedback: Electrophysiological Correlates of Learning from Small and Large Errors.

*Journal of Cognitive Neuroscience*, *26*(5), 1180–1193. https://doi.org/10.1162/jocn_a_00543

Maier, M. E., & Steinhauser, M. (2013). Updating expected action outcome in the medial frontal cortex involves an evaluation of error type. *Journal of Neuroscience*, *33*(40), 15705–15709.

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, *164*(1), 177–190. https://doi.org/10.1016/j.jneumeth.2007.03.024

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing Type I error and power in linear mixed models. *Journal of Memory and Language*, *94*, 305–315. https://doi.org/10.1016/j.jml.2017.01.001

McDougle, S. D., Boggess, M. J., Crossley, M. J., Parvin, D., Ivry, R. B., & Taylor, J. A. (2016). Credit assignment in movement-dependent reinforcement learning. *Proceedings of the National Academy of Sciences*, *113*(24), 6797–6802. https://doi.org/10.1073/pnas.1523669113

McDougle, S. D., Butcher, P. A., Parvin, D. E., Mushtaq, F., Niv, Y., Ivry, R. B., & Taylor, J. A. (2019). Neural Signatures of Prediction Errors in a Decision-Making Task Are Modulated by Action Execution Failures. *Current Biology*, *29*(10), 1606-1613.e5. https://doi.org/10.1016/j.cub.2019.04.011

McGuire, J. T., Nassar, M. R., Gold, J. I., & Kable, J. W. (2014). Functionally Dissociable Influences on Learning Rate in a Dynamic Environment. *Neuron*, *84*(4), 870–881. https://doi.org/10.1016/j.neuron.2014.10.013

Meyniel, F. (2020). Brain dynamics for confidence-weighted learning. *PLOS Computational Biology*, *16*(6), e1007935. https://doi.org/10.1371/journal.pcbi.1007935

Meyniel, F., & Dehaene, S. (2017). Brain networks for confidence weighting and hierarchical inference during probabilistic learning. *Proceedings of the National Academy of Sciences*, *114*(19). https://doi.org/10.1073/pnas.1615773114

Meyniel, F., Schlunegger, D., & Dehaene, S. (2015). The Sense of Confidence during Probabilistic Learning: A Normative Account. *PLOS Computational Biology*, *11*(6), e1004305. https://doi.org/10.1371/journal.pcbi.1004305

Meyniel, F., Sigman, M., & Mainen, Z. F. (2015). Confidence as Bayesian Probability: From Neural Origins to Behavior. *Neuron*, *88*(1), 78–92. https://doi.org/10.1016/j.neuron.2015.09.039

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

Mognon, A., Jovicich, J., Bruzzone, L., & Buiatti, M. (2011). ADJUST: An automatic EEG artifact detector based on the joint use of spatial and temporal features: Automatic spatio-temporal EEG artifact detection. *Psychophysiology*, *48*(2), 229–240. https://doi.org/10.1111/j.1469-8986.2010.01061.x

Mueller, V. A., Brass, M., Waszak, F., & Prinz, W. (2007). The role of the preSMA and the rostral cingulate zone in internally selected actions. *NeuroImage*, *37*(4), 1354–1361. https://doi.org/10.1016/j.neuroimage.2007.06.018

Näätänen, R. (1982). Processing Negativity: An Evoked-Potential Reflection of Selective Attention. *Psychological Bulletin*, *92*(3), 605.

Näätänen, R. (1990). The role of attention in auditory information processing as revealed by event-related potentials and other brain measures of cognitive function. *Behavioral Brain Sciences*, *13*, 201–288.

Näätänen, R., Kujala, T., & Winkler, I. (2011). Auditory processing that leads to conscious perception: A unique window to central auditory processing opened by the mismatch negativity and related responses: Auditory processing that leads to conscious perception. *Psychophysiology*, *48*(1), 4–22. https://doi.org/10.1111/j.1469-8986.2010.01114.x

Näätänen, R., Paavilainen, P., & Reinikainen, K. (1989). Do event-related potentials to infrequent decrements in duration of auditory stimuli demonstrate a memory trace in man? *Neuroscience Letters*, *107*(1–3), 347–352. https://doi.org/10.1016/0304-3940(89)90844-6

Näätänen, R., Paavilainen, P., Rinne, T., & Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: A review. *Clinical Neurophysiology*, *118*(12), 2544–2590. https://doi.org/10.1016/j.clinph.2007.04.026

Nassar, M. R., Bruckner, R., & Frank, M. J. (2019). Statistical context dictates the relationship between feedback-related EEG signals and learning. *eLife*, *8*, e46975. https://doi.org/10.7554/eLife.46975

Nassar, M. R., Wilson, R. C., Heasly, B., & Gold, J. I. (2010). An Approximately Bayesian Delta-Rule Model Explains the Dynamics of Belief Updating in a Changing Environment. *The Journal of Neuroscience*, *30*(37), 12366–12378. https://doi.org/10.1523/JNEUROSCI.0822-10.2010

Nieuwenhuis, S., Aston-Jones, G., & Cohen, J. D. (2005). Decision making, the P3, and the locus coeruleus—Norepinephrine system. *Psychological Bulletin*, *131*(4), 510–532. https://doi.org/10.1037/0033-2909.131.4.510

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

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2011). FieldTrip: Open Source Software for Advanced Analysis of MEG, EEG, and Invasive Electrophysiological Data. *Computational Intelligence and Neuroscience*, *2011*, 1–9. https://doi.org/10.1155/2011/156869

Pagnoni, G., Zink, C. F., Montague, P. R., & Berns, G. S. (2002). Activity in human ventral striatum locked to errors of reward prediction. *Nature Neuroscience*, *5*(2), 97–98.

Pavlov, I. P. (1927). *Conditioned reflexes.* Oxford University Press.

Payzan-LeNestour, É., & Bossaerts, P. (2012). Do not Bet on the Unknown Versus Try to Find Out More: Estimation Uncertainty and “Unexpected Uncertainty” Both Modulate Exploration. *Frontiers in Neuroscience*, *6*. https://doi.org/10.3389/fnins.2012.00150

Payzan-LeNestour, E., Dunne, S., Bossaerts, P., & O’Doherty, J. P. (2013). The Neural Representation of Unexpected Uncertainty during Value-Based Decision Making. *Neuron*, *79*(1), 191–201. https://doi.org/10.1016/j.neuron.2013.04.037

Perrin, F., Pernier, J., Bertnard, O., Giard, M. H., & Echallier, J. F. (1987). Mapping of scalp potentials by surface spline interpolation. *Electroencephalography*

*and Clinical Neurophysiology*, *66*(1), 75–81. https://doi.org/10.1016/0013-4694(87)90141-6

Pescetelli, N., Hauperich, A.-K., & Yeung, N. (2021). Confidence, advice seeking and changes of mind in decision making. *Cognition*, *215*, 104810. https://doi.org/10.1016/j.cognition.2021.104810

Pessiglione, M., Seymour, B., Flandin, G., Dolan, R. J., & Frith, C. D. (2006). Dopamine-dependent prediction errors underpin reward-seeking behaviour in humans. *Nature*, *442*(7106), 1042–1045. https://doi.org/10.1038/nature05051

Polich, J. (2007). Updating P300: An integrative theory of P3a and P3b. *Clinical Neurophysiology*, *118*(10), 2128–2148. https://doi.org/10.1016/j.clinph.2007.04.019

Pouget, A., Drugowitsch, J., & Kepecs, A. (2016). Confidence and certainty: Distinct probabilistic quantities for different goals. *Nature Neuroscience*, *19*(3), 366–374. https://doi.org/10.1038/nn.4240

Prinz, W. (1990). A common coding approach to perception and action. In *In Relationships between perception and action* (pp. 167–201). Springer.

Prinz, W. (1997). Perception and Action Planning. *European Journal of Cognitive Psychology*, *9*(2), 129–154. https://doi.org/10.1080/713752551

R Core Team. (2022). *R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.* https://www.R-project.org/

RCore, T. (2016). *R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria*.

Riesel, A., Weinberg, A., Endrass, T., Meyer, A., & Hajcak, G. (2013). The ERN is the ERN is the ERN? Convergent validity of error-related brain activity across different tasks. *Biological Psychology*, *93*(3), 377–385.

Rodríguez, M., Abdala, P., & Obeso, J. A. (2000). Excitatory responses in the 'direct'striatonigral pathway: Effect of nigrostriatal lesion. *Movement Disorders*, *15*(5), 795–803.

Roesch, M. R., Calu, D. J., & Schoenbaum, G. (2007). Dopamine neurons encode the better option in rats deciding between differently delayed or sized rewards. *Nature Neuroscience*, *10*(12), 1615–1624. https://doi.org/10.1038/nn2013

Rouault, M., Dayan, P., & Fleming, S. M. (2019). Forming global estimates of self-performance from local confidence. *Nature Communications*, *10*(1), Article 1. https://doi.org/10.1038/s41467-019-09075-3

Sallet, J., Camille, N., & Procyk, E. (2013). Modulation of feedback-related negativity during trial-and-error exploration and encoding of behavioral shifts. *Frontiers in Neuroscience*, *7*. https://doi.org/10.3389/fnins.2013.00209

Sanmiguel, I., Todd, J., & Schröger, E. (2013). Sensory suppression effects to self-initiated sounds reflect the attenuation of the unspecific N1 component of the auditory ERP: Auditory N1 suppression: N1 components. *Psychophysiology*, *50*(4), 334–343. https://doi.org/10.1111/psyp.12024

Sassenhagen, J., & Draschkow, D. (2019). Cluster-based permutation tests of MEG/EEG data do not establish significance of effect latency or location. *Psychophysiology*, *56*(6), e13335. https://doi.org/10.1111/psyp.13335

Saupe, K., Widmann, A., Trujillo-Barreto, N. J., & Schröger, E. (2013). Sensorial suppression of self-generated sounds and its dependence on attention.

*International Journal of Psychophysiology*, *90*(3), 300–310. https://doi.org/10.1016/j.ijpsycho.2013.09.006

Schiffer, A.-M., Siletti, K., Waszak, F., & Yeung, N. (2017). Adaptive behaviour and feedback processing integrate experience and instruction in reinforcement learning. *NeuroImage*, *146*, 626–641. https://doi.org/10.1016/j.neuroimage.2016.08.057

Schönberg, T., Daw, N. D., Joel, D., & O'Doherty, J. P. (2007). Reinforcement Learning Signals in the Human Striatum Distinguish Learners from Nonlearners during Reward-Based Decision Making. *The Journal of Neuroscience*, *27*(47), 12860–12867. https://doi.org/10.1523/JNEUROSCI.2496-07.2007

Schultz, W., Dayan, P., & Montague, P. R. (1997). A Neural Substrate of Prediction and Reward. *Science*, *275*(5306), 1593–1599.

Schulz, E., & Gershman, S. J. (2019). The algorithmic architecture of exploration in the human brain. *Current Opinion in Neurobiology*, *55*, 7–14. https://doi.org/10.1016/j.conb.2018.11.003

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning: An introduction*. MIT Press.

Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction*. MIT press.

Thorndike, E. L. (1905). *The elements of psychology*. New York, NY: Seiler.

Thorndike, E. L. (1911). *Animal intelligence: Experimental studies.* New York: Macmillan.

Thorndike, E. L. (1913). Ideo-motor action. *Psychological Review*, *20*, 91–106. https://doi.org/doi:10.1037/h0072027

Timm, J., SanMiguel, I., Keil, J., Schröger, E., & Schönwiesner, M. (2014). Motor Intention Determines Sensory Attenuation of Brain Responses to Self-initiated Sounds. *Journal of Cognitive Neuroscience*, *26*(7), 1481–1489. https://doi.org/10.1162/jocn_a_00552

Timm, J., SanMiguel, I., Saupe, K., & Schröger, E. (2013). The N1-suppression effect for self-initiated sounds is independent of attention. *BMC Neuroscience*, *14*(1), 2. https://doi.org/10.1186/1471-2202-14-2

Tobler, P. N., Dickinson, A., & Schultz, W. (2003). Coding of predicted reward omission by dopamine neurons in a conditioned inhibition paradigm. *Journal of Neuroscience*, *23*(32), 10402–10410.

van de Vijver, I., Ridderinkhof, K. R., & Cohen, M. X. (2011). Frontal Oscillatory Dynamics Predict Feedback Learning and Action Adjustment. *Journal of Cognitive Neuroscience*, *23*(12), 4106–4121. https://doi.org/10.1162/jocn_a_00110

Van Der Helden, J., Boksem, M. A. S., & Blom, J. H. G. (2010). The Importance of Failure: Feedback-Related Negativity Predicts Motor Learning Efficiency. *Cerebral Cortex*, *20*(7), 1596–1603. https://doi.org/10.1093/cercor/bhp224

Vroomen, J., & Stekelenburg, J. J. (2010). Visual Anticipatory Information Modulates Multisensory Interactions of Artificial Audiovisual Stimuli. *Journal of Cognitive Neuroscience*, *22*(7), 1583–1596. https://doi.org/10.1162/jocn.2009.21308

Wagner, A. R., & Rescorla, R. A. (1972). Inhibition in Pavlovian conditioning: Application of a theory. *Inhibition and Learning*, 301–336.

Walentowska, W., Moors, A., Paul, K., & Pourtois, G. (2016). Goal relevance influences performance monitoring at the level of the FRN and P3 components. *Psychophysiology*, *53*(7), 1020–1033.

Waszak, F., Cardoso-Leite, P., & Hughes, G. (2012). Action effect anticipation: Neurophysiological basis and functional consequences. *Neuroscience & Biobehavioral Reviews*, *36*(2), 943–959. https://doi.org/10.1016/j.neubiorev.2011.11.004

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer.

Williams, C. C., Ferguson, T. D., Hassall, C. D., Abimbola, W., & Krigolson, O. E. (2021). The ERP, frequency, and time–frequency correlates of feedback processing: Insights from a large sample study. *Psychophysiology*, *58*(2), e13722. https://doi.org/10.1111/psyp.13722

Wilson, R. C., Bonawitz, E., Costa, V. D., & Ebitz, R. B. (2021). Balancing exploration and exploitation with information and randomization. *Current Opinion in Behavioral Sciences*, *38*, 49–56. https://doi.org/10.1016/j.cobeha.2020.10.001

Winkler, I., Karmos, G., & Näätänen, R. (1996). Adaptive modeling of the unattended acoustic environment reflected in the mismatch negativity event-related potential. *Brain Research*, *742*(1–2), 239–252. https://doi.org/10.1016/S0006-8993(96)01008-6

Wise, R. A. (2004). Dopamine, learning and motivation. *Nature Reviews Neuroscience*, *5*(6), 483–494.

Wise, R. A., Spindler, J., & Legault, L. (1978). Major attenuation of food reward with performance-sparing doses of pimozide in the rat. *Canadian Journal of Psychology/Revue Canadienne de Psychologie*, *32*(2), 77.

Wolpert, D. M., Diedrichsen, J., & Flanagan, J. R. (2011). Principles of sensorimotor learning. *Nature Reviews Neuroscience*, *12*(12), 739–751. https://doi.org/10.1038/nrn3112

Wolpert, D. M., & Flanagan, J. R. (2016). Computations underlying sensorimotor learning. *Current Opinion in Neurobiology*, *37*, 7–11. https://doi.org/10.1016/j.conb.2015.12.003

Wolpert, D. M., Ghahramani, Z., & Jordan, M. I. (1995). An Internal Model for Sensorimotor Integration. *Science*, *269*(5232), 1880–1882. https://doi.org/10.1126/science.7569931

Wurm, F., Walentowska, W., Ernst, B., Severo, M. C., Pourtois, G., & Steinhauser, M. (2022). Task Learnability Modulates Surprise but Not Valence Processing for Reinforcement Learning in Probabilistic Choice Tasks. *Journal of Cognitive Neuroscience*, *34*(1), 34–53. https://doi.org/10.1162/jocn_a_01777

Yasuda, A., Sato, A., Miyawaki, K., Kumano, H., & Kuboki, T. (2004). Error-related negativity re£ects detection of negative reward prediction error. *Neuroreport*, *15*(16), 2561–2565.

Ye, M., Lyu, Y., Sclodnick, B., & Sun, H.-J. (2019). The P3 Reflects Awareness and Can Be Modulated by Confidence. *Frontiers in Neuroscience*, *13*, 510. https://doi.org/10.3389/fnins.2019.00510

Yeung, N., & Sanfey, A. G. (2004). Independent Coding of Reward Magnitude and Valence in the Human Brain. *The Journal of Neuroscience*, *24*(28), 6258–6264. https://doi.org/10.1523/JNEUROSCI.4537-03.2004

Yeung, N., & Summerfield, C. (2012a). 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

Yeung, N., & Summerfield, C. (2012b). 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

Yordanova, J., Falkenstein, M., Hohnsbein, J., & Kolev, V. (2004). Parallel systems of error processing in the brain. *NeuroImage*, *22*(2), 590–602. https://doi.org/10.1016/j.neuroimage.2004.01.040

Zapparoli, L., Seghezzi, S., & Paulesu, E. (2017). The What, the When, and the Whether of Intentional Action in the Brain: A Meta-Analytical Review. *Frontiers in Human Neuroscience*, *11*, 238. https://doi.org/10.3389/fnhum.2017.00238