UNIVERSITE DE LILLE
Ecole doctorale Sciences de l’Homme et de la Société
U.F.R. de Psychologie
Laboratoire de Sciences Cognitives et Affectives (SCALab)–UMR CNRS 9193

THESE

EN VUE DE L’OBTENTION DU GRADE DE DOCTEUR

DISCIPLINE : PSYCHOLOGIE

DYNAMICS OF PROCESSING OF VISUAL KINEMATICS AND GOAL-RELATED INFORMATION DURING THE RECOGNITION OF OBJECT DIRECTED ACTIONS: BEHAVIOURAL AND NEUROPHYSIOLOGICAL EVIDENCE

DYNAMIQUE DE TRAITEMENT DES INFORMATIONS VISUELLES DE GESTE ET DE BUT LORS DE LA RECONNAISSANCE D'ACTIONS DIRIGEES VERS LES OBJETS : ARGUMENTS COMPORTEMENTAUX ET NEUROPHYSIOLOGIQUES

Présentée et soutenue publiquement
Par

Jérémy DECROIX

Sous la direction du Dr. Solène KALENINE
Le mercredi 5 Février 2020

Composition du jury :

Pr. Michael ANDRES, Université Catholique de Louvain-la-Neuve (Examinateur)
Pr. Caroline CATMUR, King’s College London (Examinatrice)
Dr. Julie GREZES, Ecole Normale Supérieure (Présidente du jury)
Dr. Solène KALENINE, Université de Lille (Directrice)
Pr. Yves ROSSETTI, Université Lyon 1 (Rapporteur)
Pr. Cosimo URGESI, Università degli studi di Udine (Examinateur)

UNIVERSITE DE LILLE
Ecole doctorale Sciences de l'Homme et de la Société
U.F.R. de Psychologie
Laboratoire de Sciences Cognitives et Affectives (SCALab)–UMR CNRS 9193

THESE

EN VUE DE L'OBTENTION DU GRADE DE DOCTEUR

DISCIPLINE : PSYCHOLOGIE

DYNAMICS OF PROCESSING OF VISUAL KINEMATICS AND GOAL-RELATED INFORMATION DURING THE RECOGNITION OF OBJECT DIRECTED ACTIONS: BEHAVIOURAL AND NEUROPHYSIOLOGICAL EVIDENCE

DYNAMIQUE DE TRAITEMENT DES INFORMATIONS VISUELLES DE GESTE ET DE BUT LORS DE LA RECONNAISSANCE D'ACTIONS DIRIGEES VERS LES OBJETS : ARGUMENTS COMPORTEMENTAUX ET NEUROPHYSIOLOGIQUES

Présentée et soutenue publiquement

Par

Jérémy DECROIX

Sous la direction du Dr. Solène KALENINE

Le mercredi 5 Février 2020

Composition du jury :

Pr. Michael ANDRES, Université Catholique de Louvain-la-Neuve (Examinateur)

Pr. Caroline CATMUR, King's College London (Examinatrice)

Dr. Julie GREZES, Ecole Normale Supérieure (Présidente du jury)

Dr. Solène KALENINE, Université de Lille (Directrice)

Pr. Yves ROSSETTI, Université Lyon 1 (Rapporteur)

Pr. Cosimo URGESI, Università degli studi di Udine (Examinateur)

Je dédie cette thèse à ma mère et ma sœur. Merci pour votre soutien indéfectible, pour votre immense patience. Nous avons traversé beaucoup d'épreuves et les choses n'ont pas toujours été simples. Merci d'avoir tenu bon et d'avoir été aussi fortes. Merci d'avoir cru en moi. Merci de m'avoir porté toutes ces années et de m'avoir toujours poussé à aller plus loin.

Je dédie également cette thèse à mon père. Merci de n'avoir jamais cessé de me soutenir malgré tes doutes. J'espère être devenu un adulte dont tu serais fier. Je porte en moi ton souvenir et tente de garder intacts les principes et valeurs que tu m'as enseignés.

Foreword and acknowledgements

Lorsque l'on interroge un docteur sur son expérience de thèse, presque sans exception, il évoquera, en premier lieu, la connaissance que cela lui a apporté sur lui-même. Au-delà des connaissances techniques et théoriques, c'est la connaissance et la découverte de soi que l'on acquiert au travers de la thèse. Je crois pouvoir dire que j'ai bien vécu ma thèse. Il y a eu des hauts (très haut) et des bas (très bas), mais j'ai eu une bonne thèse. Je crois aussi que je suis devenu une toute autre personne que celle que j'étais il y a trois ans. Il serait prétentieux de prétendre que j'ai acquis de la maturité (même si j'espère en avoir acquis) mais je pense pouvoir dire que j'ai acquis de l'assurance, de l'acceptation et de la tolérance envers moi-même, et envers les autres. On apprend aussi beaucoup sur le monde académique et sur la recherche en elle-même. Sur la recherche, on apprend la réalité et toute sa complexité. La science n'est pas le meilleur moyen d'acquérir des connaissances, c'est, je crois, le moins mauvais. La distinction est importante ; on aurait tort de croire que nous disposons de toutes les clés, et que notre système est infaillible. Du monde académique, on apprend les dessous, sa dimension politique et économique ; on apprend les relations complexes qui relient les individus qui le composent. On apprend les batailles que chacun décide de mener, les concessions que chacun décide de faire. Au sein de cet écosystème bien particulier, on s'adapte, on s'ajuste.

Il est presque impossible d'arriver dans ce milieu en ayant une connaissance réaliste de ce qui nous attend. N'est-il pas surprenant que, dans une discipline scientifique, nous devons *croire* en nos données, nous devons *défendre* nos résultats ? On découvre que la réalité est un objet complexe, et que saisir sa complexité est une tâche presque inachevable, mais à laquelle nous nous attelons néanmoins sans relâche. Nous avançons en acceptant la part d'incertitude omniprésente dans nos données, dans nos choix. Peut-être de par son objet, la part d'incertitude en Psychologie est immense. Pour accéder à des bribes de réalité, la Psychologie doit se frayer un chemin à travers de multiples obstacles, et peu ont conscience de ces obstacles. Nous n'accédons pas à la réalité de son objet d'étude mais seulement à des reflets. Nous tentons d'entrapercevoir cette réalité à travers des indicateurs, et des modèles. Nous confions nos conclusions aux statistiques, dont nous percevons bien peu les biais. Si les Neurosciences ouvrent une voie nouvelle sur la réalité à laquelle s'intéresse la Psychologie, c'est aussi une voie pleine de pièges dans lesquels il est aisé de tomber. Le fait est que nous ne savons pas, fondamentalement, ce qu'indiquent les variations que nous observons de l'activité cérébrale (Cohen, 2017; Logothetis, 2008), et que nous manquons de rigueur dans l'établissement de ce que nous érigeons en marqueur des mécanismes psychologiques auxquelles nous nous intéressons (e.g., Luck & Gaspelin, 2017; Naish, Houston-Price, Bremner, & Holmes, 2014). Ces critiques sont applicables aux mesures comportementales auxquelles la Psychologie est coutumière, mais elles sont décuplées dans le cadre des Neurosciences tant la complexité des mesures que l'on y récolte est décuplée. Aux travers de ces débats, le doctorant en Psychologie doit faire le deuil d'espérer trouver la bonne solution, la bonne manière de faire ; il doit accepter d'être faillible, il doit accepter de faire des choix.

L'aspect le plus inexplicable de la thèse réside peut-être dans l'implication émotionnelle que les doctorants mettent dans leur thèse. Pourquoi avons-nous tant de difficultés à prendre de la distance ? Pourquoi ressentons-nous avec tant d'intensité les réussites et les échecs de nos projets ? Peut-être parce que le poids de chaque réussite et de chaque échec est proportionnel à l'énergie investie, et aux possibles conséquences que cela implique. Peut-être aussi parce qu'au travers de la thèse, nous apprenons le poids des responsabilités, et la nécessité d'endosser des choix que nous n'avons parfois pas fait nous-mêmes. Peut-être aussi parce que nous apprenons à gérer des relations dans lesquelles il peut être difficile d'établir des compromis. La recherche, c'est aussi chercher un équilibre difficile entre ses valeurs, celles des autres, les contraintes politiques et économiques. Je crois qu'il est impossible de traverser cette période sans consentir à un rééquilibrage. Ce rééquilibrage est d'autant plus ardu quand on prend conscience que les contraintes politiques et académiques vont parfois à l'encontre des contraintes imposées par le milieu scientifique. Il est encore plus ardu quand on prend conscience que les fondations mêmes des disciplines scientifiques reposent sur les valeurs de ceux qui les composent, et que ces valeurs sont parfois rudement éprouvées. J'ai pu entendre dire par le passé que la recherche et le monde académique possédaient la double caractéristique d'être un milieu où l'on est à la fois très seul et très entouré. Je pense maintenant percevoir la nature de cette double caractéristique. Nous sommes constamment en interaction, et inextricablement emmêlés dans un réseau relationnel complexe, mais en définitif, nous sommes seuls face à nos choix.

Je ne suis ni triste, ni désabusé de ces considérations, je les considère pragmatiques, et ce sont ces réflexions et ces choix, ces nécessités et ces contraintes qui m'ont accompagné lors de la réalisation de cette thèse. C'est l'association de ces différentes contraintes et de ces réflexions qui a contribué à rendre mon expérience de thèse riche et intéressante. C'est au travers de cet intérêt, non seulement pour ma thématique de recherche, mais aussi pour le contexte dans lequel la recherche se déroule que j'ai pu trouver mon confort dans ma pratique de la recherche.

Pour toutes ces raisons, je dois infiniment à ma directrice de thèse, Solène Kalénine. Solène, merci pour votre confiance, votre patience et votre énergie inébranlable. Merci de m'avoir écouté et de m'avoir laissé du temps lorsque j'en avais besoin. Merci de m'avoir appris à tendre vers l'excellence tout en tenant compte des réalités propres à la recherche. Merci de m'avoir accompagné dans mes périodes de doute. Merci de m'avoir appris à ne jamais cesser d'avancer. Enfin, merci d'avoir suscité chez moi le sentiment d'avoir, plus que la place d'étudiant, la place de collaborateur. Je crois que j'aurais du mal à retrouver une relation professionnelle comme celle que j'ai pu avoir avec vous, mais je sais que vous m'avez donné tous les outils nécessaires pour m'épanouir dans ma vie professionnelle future.

J'aimerais également remercier mes collaborateurs et collègues, Yannick Wamain, Clémence Roger, Nicolas Morgado, Cynthia Collette et Marc Godard. Merci pour votre disponibilité et la qualité de nos échanges. Merci de me permettre de garder intact ma curiosité et mon goût pour le débat. Merci pour la place que vous me laissez dans nos échanges.

J'aimerais adresser un remerciement particulier à Angela Bartolo, qui a accepté de porter ma candidature devant l'école doctorale. Merci pour votre confiance, votre dynamisme et votre bonne humeur.

Je remercie mon laboratoire, avec une pensée particulière pour Emmanuelle Fournier, Sabine Pierzchala, Laurent Ott et Marion Vincent pour toute l'aide technique et administrative, mais également pour le soutien moral. Merci d'avoir accueilli mes visites et demandes inopinées. Merci également aux membres du laboratoire de m'avoir soutenu dans l'établissement de mes collaborations internationales. Merci à l'Ecole Doctorale, et en particulier à Catherine Maignant et Sabrina Abed pour leur disponibilité et leur réactivité face à mes incessantes sollicitations.

Merci à Alessio Avenanti de m'avoir accueilli dans son laboratoire. Merci pour votre gentillesse et votre vivacité d'esprit, votre capacité à saisir l'essentiel et la portée de vos réflexions ne cesseront jamais d'être une source d'étonnement. Merci à Sara Borgomaneri pour sa pédagogie et son enthousiasme. Merci de m'avoir apporté des réflexions sur la recherche et sur l'orientation de ma carrière.

Un immense merci à Laura Caton et Fanny Grisetto, merci d'être des co-bureaux géniales. Merci pour votre confiance, votre disponibilité et votre soutien indéfectible. Merci pour vos relectures et pour la mise en page. Cette thèse vous appartient en partie tant vous avez contribué. Je sais combien la thèse n'est pas toujours facile, et que malgré les difficultés que vous pouvez rencontrer, vous trouvez toujours du temps à consacrer à autrui. Merci pour vos petites attentions et votre bonne humeur malgré les épreuves. J'attends avec impatience le jour où j'assisterai à vos soutenances. #teamprincesse#paillette

Merci à Ségolène Guérin pour son éternel et inépuisable (genre, vraiment) enthousiasme. Merci pour tes relectures et tes conseils. #queenAPA6

Merci à Clément Beugnet, Alicia Rassel et Sébastien Vanstavel, il semblerait que je sois le premier à soutenir de notre promo, mais je sais que je ne serai pas le dernier, et malgré vos doutes, j'ai confiance en vous.

Un remerciement particulier à Audrey Lavallée. Merci pour ta confiance et ta tolérance. J'aimerais que tu puisses croire en toi autant que moi. Merci pour la qualité de tes réflexions, merci pour tes qualités humaines, merci pour ton positionnement clinique. Tu ne t'en rends peut-être pas compte, mais merci d'être, pour moi, un exemple d'humanité. Merci pour ta modestie et ta patience.

Merci à Alice Cartaud et Vincent Lenglin. Merci pour votre soutien, nos réflexions, nos débats, notre vision du monde de la recherche. Merci pour votre regard critique et honnête sur le monde. Merci de faire de moi un meilleur chercheur et de toujours me pousser à me remettre en question. #keurkeurlove #lapvaluecestnul #onfiniraparchangerlemonde

Merci à Fabrizia Gallo, Clémentine Garric et Justine Blampain. Merci pour votre bonne humeur, vos attentions et votre soutien. Merci pour votre enthousiasme. Ne changez rien les filles !

Merci à Apolline Delobeau et Lucas Ronat pour leur soutien. On se connait maintenant depuis un petit moment, et j'ai plaisir à voir nos parcours évoluer, et à regarder le chemin que nous avons parcouru. Merci à Christopher Anciaux, merci pour ta stabilité et ta loyauté, pour ta patience et ton soutien. Merci de m'ouvrir à un monde en dehors de la recherche, merci d'avoir étendu mes perspectives.

Pour finir, j'adresse mes remerciements aux membres de mon jury, Pr. Michael Andres, Pr. Caroline Catmur, Dr. Julie Grèzes, Pr. Yves Rossetti et Pr. Cosimo Urgesi. Merci d'avoir accepté d'évaluer mon travail. C'est avec un peu d'appréhension et d'excitation que je me soumets à votre évaluation. J'espère que ce travail sera à la hauteur de vos attentes et que vous en apprécierez le contenu.

Jérémy Decroix,
Le 12 Novembre 2019, Lille.

Funding

This thesis was funded by the French National Research Agency (ANR-16-CE28-0003) and the region Hauts-de-France.

A short 3-month visiting internship in the University of Bologna benefited from a grant of the Jean Walter Zellidja foundation (French Academy).

International communications benefited from international mobility grants of the Ecole doctorale Sciences de l'Homme et de la Société (University of Lille).

Résumé

Les actions sont des mouvements complexes dirigés vers un but. En dépit du fait que ces buts ne soient pas directement visibles, nous parvenons, en tant qu'observateur, à les identifier et à les anticiper avec succès. Dans cette thèse, nous avons identifié deux approches pour expliquer comment, en tant qu'observateur, nous parvenons à reconnaitre les actions d'autrui. Les approches sensorimotrices considèrent la reconnaissance des actions d'autrui comme une propagation ascendante de la perception des gestes à la reconnaissance des buts. Les gestes sont vus ici comme la principale source d'information à partir de laquelle le but de l'acteur peut être extrait. A l'opposé, les approches prédictives considèrent que l'observateur ne peut pas comprendre les gestes de l'autre sans préalablement prédire son but probable. L'observateur doit donc extraire le but probable de l'action à partir d'autres sources d'information pour le guider dans son traitement des gestes observés. Les connaissances sur le traitement dynamique des gestes observés d'une part et des informations sur le but d'autre part sont critiques pour dissocier ces deux approches et pour permettre une meilleure compréhension des mécanismes qui sous-tendent la reconnaissance des actions d'autrui. Néanmoins, les données empiriques dans cette direction manquent cruellement. Afin de combler cette lacune, nous nous sommes intéressés à la priorité donnée aux gestes observés relativement à celle donnée aux informations sur le but lors de la reconnaissance des actions d'autrui. La contribution des informations sur le geste observé et sur le but a été évaluée indépendamment grâce à l'introduction de violation de geste et/ou de but dans des photographies d'actions dirigées vers des objets. Grâce à des méthodes comportementales (amorçage et tâche de recherche visuelle), nous avons constaté que les informations liées au but sont priorisées durant les premiers stades du traitement visuel des actions, alors que les informations liées aux gestes observés sont priorisées durant les derniers stades du traitement visuel des actions. Grâce à des méthodes neurophysiologiques (potentiels évoqués et stimulation magnétique transcrânienne), nous avons observé que si les deux types d'information sont décodées dès les stades perceptifs du traitement des actions, c'est le traitement informations sur le but (et non sur le geste) qui guide les étapes sémantiques du traitement des actions. Nous apportons par ailleurs des arguments justifiant de l'implication critique du réseau fronto-pariétal dans l'intégration des deux sources d'information. Enfin, nous montrons que la tendance à donner plus d'importance aux informations sur le but relativement à celles sur les gestes observés dépend de caractéristiques individuelles. Dans l'ensemble, les données rapportées ici sont en accord avec les approches prédictives de la reconnaissance des actions. Ces résultats sont discutés à la lumière d'un faisceau d'arguments qui suggèrent que les gestes observés sont utilisés pour mettre à jours des prédictions sur le but de l'acteur, elles-mêmes préalablement dérivées d'autres sources d'informations. Nos données nous amènent finalement à envisager une approche pluraliste de la reconnaissance des actions observées, avec un ensemble de stratégies dont l'usage varie en fonction des situations et des individus.

Mots-clés : Compréhension d'action – Geste – But d'action – Utilisation d'outil – Amorçage – Oculométrie – EEG – TMS

Abstract

Actions are complex, goal-directed, movements, and, despite being hidden in the actor's mind, observers successfully identify and anticipate actor's goal. In this thesis, we identified two main approaches to explain how observers recognise others' actions. Sensorimotor approaches consider action recognition as bottom-up propagation from the perception of visual kinematics to the recognition of action goals. Visual kinematics are viewed here as the primary source of visual information from which goal-related information is extracted. In contrast, predictive approaches assume that observers cannot make sense of visual kinematics without a prediction about the actor's goal. Observers would extract goal-related information from non-motor sources of information to guide the processing of the visual kinematics. Information about the temporal dynamics of activation of visual kinematics and goal-related information during action visual processing is critical to disentangle the two approaches and to provide a better understanding of the mechanisms underlying action recognition, but empirical data in this direction are clearly lacking. In order to fill this gap, we investigated the relative priority given to visual kinematics versus non-motor goal-related information during the recognition of others' actions. The contribution of visual kinematics and non-motor goal-related information was independently evaluated by introducing violations of grip and/or visual goal in photographs of object-directed actions. Using behavioural methods (priming and visual-search paradigms), we demonstrated that non-motor goal-related information was prioritised over visual kinematics during the first steps of visual action processing, whereas visual kinematics were prioritised over goal-related information later during visual action processing. Using neurophysiological methods (event-related potential and transcranial magnetic stimulation priming paradigms), we found that both visual kinematics and non-motor goal-related information are already processed during the perceptual stages of action processing, but that action semantic processing is guided by goal-related information rather than visual kinematics. We further provide evidence supporting the critical involvement of the frontoparietal network in the later integration of visual kinematics and non-motor goal-related information. We finally showed that the priority given to non-motor goal-related information over visual kinematics during action visual processing depends on individual social characteristics. Together, the findings reported are consistent with predictive approaches of action recognition. Results are discussed in the light of converging evidence suggesting that visual kinematics are used to update goal predictions that have been previously derived from non-motor goal-related information. Yet findings further orient towards a pluralist view of action understanding, in which the strategies used to process others' actions may vary depending on situations and individuals.

Key-words: Action understanding – Visual kinematics – Action goals – Tool use – Priming – Eyetracking – EEG – TMS

Publications

Publications related to the PhD thesis

Decroix, J., Borgomaneri, S., Kalénine, S. & Avenanti, A. (in prep). Frontal and parietal integration of visual kinematics and functional goals during object-directed action recognition: Evidence from TMS-priming?

Decroix, J., Morgado, N. & Kalénine, S. (in prep). Preference for visual goal over grip explained by individual characteristics during the recognition of object-directed actions.

Decroix J., Roger, C. & Kalénine S. (under review). Electrophysiological correlates of grip and goal decoding during the visual processing of others' object-directed actions.

Decroix J. & Kalénine S. (2019). What first drives visual attention during the recognition of object-directed action? The role of kinematics and goal information. *Attention, Perception & Psychophysics, 1-10*. doi: 10.3758/s13414-019-01784-7

Decroix J. & Kalénine S. (2018). Timing of grip and goal activation during action perception: A priming study. *Experimental Brain Research, 1-16*. doi: 10.1007/s00221-018-5309-0

Other publications

Wamain Y., Sahaï A., **Decroix J.**, Coello Y. & Kalénine S. (2018). Conflict between gesture representations reduces μ rhythm desynchronization during manipulable object perception: an EEG study. *Biological Psychology, 132, 202-211*. doi: 10.1016/j.biopsycho.2017.12.004

Kalénine, S., Wamain, Y., **Decroix, J.**, & Coello, Y. (2016). Conflict between object structural and functional affordances in peripersonal space. *Cognition, 155, 1-7*. doi: 10.1016/j.cognition.2016.06.006

Communications

Communications related to the PhD thesis

Decroix, J., Morgado, N. & Kalénine, S. (25-28 September 2019). Individual dispositions predict goal priority during the recognition of others' actions. *European Society of Cognitive Psychology*, Tenerife, Spain. **Talk**.

Decroix, J., Borgomaneri, S., Kalénine, S. & Avenanti, A. (7-9 March 2019). The inferior parietal cortex integrates kinematics and functional goal information during the recognition of object-directed actions: A TMS-Priming experiment. *International Convention of Psychological Science*, Paris, France. **Poster**.

Decroix, J., Roger, C. & Kalénine, S. (2-7 October 2018). An electrophysiological study of grip and goal decoding during action observation. *Society for Psychophysiological Research*, Quebec City, Canada. **Poster**.

Decroix, J. & Kalénine, S. (27-31 August 2017). Gesture and outcome processing during the recognition of actions among distractors: Evidence from eyetracking. *European Congress on Visual Perception*, Berlin, Germany. **Poster**.

Decroix, J. & Kalénine, S. (29-30 June 2017). Gesture and outcome activation during the decoding of other's action: which one is processed first? *Workshop TRACE,* Nanterre, France. **Talk**.

Decroix, J. & Kalénine, S. (4-7 May 2017). Timing of gesture and outcome processing during action decoding. *CAOs Workshop on Concepts, Actions and Objects: Functional and Neural Perspectives*, Rovereto, Italy. **Poster**.

Decroix, J. & Kalénine, S. (12-13 December 2016). Gesture and outcome decoding during action perception: When are the two representational levels activated and interrelated? *International Symposium on Apraxia and Action,* Angers, France. **Talk & Poster**.

Other communications

Wamain, Y., Sahaï, A., **Decroix, J.**, Coello, Y. & Kalénine, S. (3-7 September 2017). When visual objects evoke multiple gestures: Impact of conflict between affordances on µ rhythm desynchronization. *European Society for Cognitive Psychology*, Postdam, Germany. **Poster**.

Wamain, Y., Sahaï, A., **Decroix, J.**, Coello, Y. & Kalénine, S. (27-31 August 2017). Neurophysiological correlates of conflict between gesture representations during object perception. *European Congress on Visual Perception,* Berlin, Germany. **Poster**.

Decroix, J., Wamain, Y. & Kalénine, S. (1 December 2016). Conflictualité dans l'espace: Influence de l'espace sur l'activation des représentations de geste d'utilisation et de déplacement d'objet. *XIe Journée Scientifique des Jeunes Chercheurs en Psychologie*, Lille, France. **Talk**.

Sahaï, A., **Decroix, J.**, Wamain, Y., Coello, Y. & Kalénine, S. (2015, September). Space influences evocation of structural and functional object-related gestures. *6th International Conference on Spatial Cognition: "Space and Situated Cognition"*, Rome, Italy. **Poster**.

Contents

FOREWORD AND ACKNOWLEDGEMENTS I
FUNDING VII
RÉSUMÉ VIII
ABSTRACT IX
PUBLICATIONS X
COMMUNICATIONS XI
CONTENTS XIII
FIGURES XVI
TABLES XXI
ABBREVIATIONS XXII
PREFACE XXIII
INTRODUCTION .... - 1 -

Chapter 1: The complexity of action .... - 3 -

1. Representations, components and dimensions in action processing .... - 3 -
2. Components of action .... - 7 -
   2.1. Aery intentions .... - 10 -
   2.2. The concept of goal .... - 12 -
   2.3. The representation of the motor act .... - 17 -
   2.4. The hierarchical organisation of actions .... - 21 -
3. Summary .... - 23 -

Chapter 2 The place of motor acts and goals when understanding others' actions: Mirror neurons, sensorimotor and predictive approaches of action recognition .... - 27 -

1. Discovery of mirror neurons and first interpretations .... - 28 -
   1.1. In the monkey brain .... - 28 -
   1.2. In the human brain .... - 30 -
   1.3. Proof of existence does not mean proof of function .... - 32 -
2. Sensorimotor approaches: Motor simulation guides action recognition .... - 33 -
   2.1. Initial sensorimotor accounts .... - 33 -
   2.2. Mirror neurons represent associations not goals .... - 34 -
   2.3. Simulation and mirror neurons: an insight from motor control .... - 36 -
   2.4. Synthesis .... - 38 -
3. Predictive approaches: Motor simulation does not guide action recognition .... - 39 -
   3.1. The emulation account .... - 40 -
   3.2. The predictive coding framework .... - 41 -
   3.3. The affordance-matching hypothesis .... - 43 -
   3.4. Synthesis .... - 44 -
4. General summary .... - 45 -

Chapter 3: What can we learn from behavioural and neurophysiological studies on action recognition? Considerations and current directions .... - 47 -

1. The Action Observation Network .... - 47 -
2. Empirical support for sensorimotor and predictive approaches .... - 50 -
   2.1. Evidence in favour of sensorimotor approaches .... - 50 -
   2.2. Evidence in favour of predictive approaches .... - 51 -
   2.3. Interim summary .... - 53 -
3. The temporal dynamics of action recognition .... - 55 -
   3.1. Contribution of neurophysiological studies .... - 55 -
   3.2. Contributions of behavioural experiments .... - 59 -
4. General summary .... - 62 -

GENERAL METHODOLOGY .... - 65 -

1. Overview of the section .... - 67 -

2. An action violation paradigm: choice and theoretical justification - 67 -
  2.1. van Elk and colleagues (2008)'s paradigm - 68 -
  2.2. The choice of object-directed actions - 69 -
  2.3. Targeting the visual kinematic component: grip configuration - 70 -
  2.4. Targeting the non-motor goal-related dimension: visual information about the functional goal - 71 -
  2.5. Task instruction: overall correctness judgement about the action - 73 -
  2.6. Stimuli - 74 -
3. Effects of interest - 77 -
4. Statistical strategy - 78 -
  4.1. Mixed-effects model, a brief introduction - 79 -
  4.2. Identifying the random structure - 80 -
  4.3. Evaluating significance - 84 -
  4.4. Effect size measures - 87 -
5. Summary of the methodological strategy - 88 -

EXPERIMENTAL CONTRIBUTION - 89 -

Chapter 1: Timing of grip and goal activation during action recognition: behavioural experiments - 91 -

PART 1: Priming Experiments - 92 -

1. First Experiment: 66 and 300 ms prime durations - 94 -
  1.1. Methodology - 95 -
  1.2. Results - 100 -
  1.3. Interim discussion - 102 -
2. Second Experiment: 120 and 220 ms prime durations - 105 -
  2.1. Methodology - 106 -
  2.2. Results - 107 -
  2.3. Interim discussion - 111 -
3. Experiment 3: Replication at 66 ms and 220 ms prime durations - 113 -
  3.1. Methodology - 114 -
  3.2. Results - 115 -
  3.3. Interim discussion - 119 -
4. Discussion on priming study - 122 -

PART 2 Visual Search Experiment - 123 -

1. Methods - 125 -
  1.1. Participants - 125 -
  1.2. Stimuli and Design - 126 -
  1.3. Procedure - 127 -
  1.4. Fixation proportion - 127 -
  1.5. Saliency Maps - 128 -
  1.6. Data Analysis - 129 -
2. Results - 131 -
  2.1. Main analysis of fixation proportions - 131 -
3. Interim Discussion - 133 -

SYNTHESIS - 135 -

Chapter 2: Neurophysiological correlates of grip and visual goal information - 139 -

PART 1: ERP correlates of grip and visual goal information decoding during the processing of others' actions with objects - 141 -

1. Methods - 145 -
  1.1. Participants - 145 -
  1.2. Stimuli - 146 -
  1.3. Procedure - 146 -
  1.4. EEG recording and analysis - 148 -
  1.5. Data analysis - 151 -
  1.6. Statistical approach - 152 -
2. Results - 153 -
  2.1. P100, N170, and P300 on posterior site - 153 -
  2.2. N300 and N400 on anterior site - 155 -

3. Interim discussion - 155 -
PART 2: Frontal and parietal integration of visual kinematics and functional goals during object-directed action recognition: Evidence from TMS-priming? - *161* -
1. Methods - 168 -
    1.1. Participants - 168 -
    1.2. Stimuli - 168 -
    1.3. Design and procedure - 168 -
    1.4. Transcranial Magnetic Stimulation (TMS) - 170 -
    1.5. Data pre-processing - 172 -
    1.6. Data analysis using mixed-effect models - 172 -
2. Results - 174 -
3. Interim discussion - 176 -
SYNTHESIS - *178* -
Chapter 3: Individual characteristics and action recognition: Explaining the individual differences in prioritising goal over grip during the first steps of action recognition - 183 -
1. Methodology - 187 -
    1.1. Participants - 187 -
    1.2. Procedure - 187 -
    1.3. Data analysis - 192 -
2. Results - 193 -
    2.1. Cross-correlation - 193 -
    2.2. Multiple Linear Regressions - 194 -
3. Discussion - 195 -
GENERAL DISCUSSION - 199 -
1. Overall summary - 201 -
2. Visual kinematics and non-motor goal-related information during the recognition of object-directed actions - 203 -
    2.1. Object identity, an unlikely predictor - 204 -
    2.2. The late involvement of visual kinematics in action recognition - 206 -
    2.3. Synthesis - 207 -
3. Versatile action recognition: various strategies and mechanisms - 208 -
    3.1. The specificity of object-directed actions - 209 -
    3.2. A pluralist view of action recognition - 211 -
4. Concluding remarks - 214 -
REFERENCES - 217 -
APPENDICES - 239 -
Appendix 1 - 241 -
Appendix 2 - 242 -
Appendix 3 - 243 -
Appendix 4 - 244 -
Appendix 5 - 245 -
Appendix 6 - 246 -
Appendix 7 - 247 -
Appendix 8 - 248 -
Appendix 9 - 250 -
Appendix 10 - 251 -
Appendix 11 - 255 -
Appendix 12 - 256 -
FRENCH SUMMARY - 257 -

Figures

Figure 1. A simple action hierarchy. Inspired from Csibra (2008). ....- 9 -

Figure 2. Schematic representation of conceptualisation before (upper part) and after Lashley (1951)'s proposal. ....- 14 -

Figure 3. The two routes model of limb apraxia. Inspired from Rothi et al. (1991). ....- 18 -

Figure 4. Example of task in which action knowledge influences perceptual judgement. Objects are visually presented either in the reachable space or not. Participants make perceptual judgements about the object (e.g., "is it reachable?" or "is it a kitchen object?") with their feet. Objects affording competing gesture representations (i.e., "conflictual", e.g., a calculator afford both a power grasp to move it and a poking gesture to make some calculations) elicit slower response times than objects affording non-competing gesture representations (i.e., "non-conflictual"). From Godard, Wamain, Delepoulle and Kalénine (2019). ....- 20 -

Figure 5. Summary of the different concepts used in the action recognition literature and their putative organisation. ....- 24 -

Figure 6. Example of a unit registered through single-cell recording selectively discharging during monkey grasping movements and during monkey observation of grasping movements made by the experimenter. A. The experimenter grasps the food. B. The monkey grasps the food. Arrows indicate the (approximate) onset of grasping. From di Pellegrino et al. (1992). ....- 29 -

Figure 7. Mirror neuron network in humans according to the mirror neuron theorists. The human mirror neurons are thought to encompass mainly areas from the inferior frontal gyrus and the inferior parietal lobule. In red, the significant clusters of activation (FDR, p<.05) revealed by each ALE analysis superimposed on a rendered ch2better template using MRIcroN. From Molenberghs et al. (2012). .. - 31 -

Figure 8. The three steps leading to the birth of mirror neurons according to the Associative Learning account. A. Weak connection between sensory and motor neurons. B. Frequent associations between motor neurons and sensory neurons. C. Motor neurons are now also discharging during the mere perception of action. From Heyes (2010). ....- 35 -

Figure 9. During action recognition, internal models could be computed through the mirror neuron network (Inferior Frontal Gyrus, IFG; Inferior Parietal Lobule, IPL; Superior Temporal Sulcus, STS; solid line) or computed through the temporal, parietal and cerebellar (CB) complex (dashed line). Inspired from Miall (2003). ....- 38 -

Figure 10. The three steps of action processing following the predictive coding framework. 1. Contextual information is conveyed from the occipital-temporal cortex (OTC) through the medial temporal gyrus (MTG) up to the inferior frontal gyrus (IFG; dashed arrows). 2. Contextual information in the IFG allows the computation of predicted kinematics, the prediction is passed down to the inferior parietal lobule (IPL) and the superior temporal sulcus (STS; green arrows). 3. Prediction errors are sent back to each node of the mirror neuron network (IFG, IPL) to improve the prediction (red arrows). . - 42 -

Figure 11. Results from de Lange et al. (2008). A. Goal violation (extraordinary intention) induces more activity in the inferior frontal gyrus (IFG) whereas grip violations (extraordinary means) induces more activity in the occipitotemporal regions. B. These results occur irrespective of the task given to the participants (i.e., attending to intentions or attending to means). .... - 49 -

Figure 12. Example of classical tasks used in studies supporting sensorimotor approaches. Participants have to guess the goal of the actor (in blue) on the sole basis of the visual kinematics. These tasks usually involve objects, without any semantic content (e.g., a wooden stick). From Quesque & Coello (2015). .... - 51 -

Figure 13. Example of trial sequence in the paradigm developed by Hudson et al. (2016, 2018). Once the fixation cross disappeared, participants heard "I'll take it" or "I'll leave it". Then, participants watched a hand moving back and forth towards the object until it stopped. After a blank screen, a static frame of the hand and the object appeared and participants had to say whether the hand was closer or further than the last frame of the video. Typically, participants will say "closer" if the announced goal was "I'll take it" and "further" if the announced goal was "I'll leave it". From Hudson et al. (2018). .... - 53 -

Figure 14. From Avanzini et al. (2013). Example of stimuli (upper part) for the apparent motion (AM) condition where a picture of an object is immediately followed by a hand-on-object picture. Below, the microstate segmentation results in the AM condition. Each micro-state is represented with a different colour. Micro-state represents temporary period of stability in the brain activity. Above each micro-state is displayed the corresponding map topography. .... - 56 -

Figure 15. From Cavallo et al. (2013). Motor Evoked Potentials were recorded while participants watched an action video in which an actor reached and grasped a bottle. The bottle could be behind or in front of an obstacle. The actor could use a precision grip (PG) or a whole hand grasp (WHG). Two muscles of the hand were recorded (FDI: first dorsal interosseous; ADM: abductor digiti minimi). FDI is involved in both PG and WHG but ADM is much more involved in WHG. Thus, their ratio represents the relative simulation of PG versus WHG. MEP were recorded at the start of the video and when the hand reached the object. Results showed that in the PG-unconstrained, MEP first reflect the visual kinematics one could expect from the context (i.e., a WHG, indexed by more activity of the ADM), and only later reflect the actual visual kinematics (i.e., a PG, indexed by more activity of the FDI). . . - 58 -

Figure 16. Stimuli used by Kalénine et al. (2013). Participants watched two videos and had to judge whether the two videos were the same or different. In the baseline pairs, the two videos were different both in terms of action goal (i.e., outcome) and in terms of kinematics (i.e., means) used to perform the action. In the means detection pairs, participants had to detect that the kinematics used to perform the action (i.e., applying detergent) were not the same in the two videos. In the outcome detection pairs, participants had to detect that the goal of the action was different between the two videos, although the kinematics used to do it were the same. In the identical pairs, participants had to say that the two videos were the same, although the visual perspective was not the same between the two videos. Overall participants were slower to answer when one of the two action dimensions were different (i.e., means detection pairs or outcome detection pairs) than when the two actions dimensions were different (i.e., baseline pairs). .... *- 60 -*

Figure 17. Example of the stimuli used by van Elk et al. (2008)........................ *- 69 -*

Figure 18. Pattern of gaze fixation as a function of the task. For judging about the weight of the object, participants essentially used saccade whereas to predict which object is about to be grasp, participants essentially tracked the hand movement and used fixation. From Flanagan et al. (2013). .... *- 74 -*

Figure 19. Example of the stimuli of Set 1. .... *- 75 -*

Figure 20. Example of the stimuli of Set 2. .... *- 76 -*

Figure 21. Example of action photographs containing grip and/or goal violations for a given stimulus. In this example, the actress grasps the upright (full and closed) water bottle with a typical clench grip, a correct action regarding the typical function of the object (drink). In the photograph presenting a goal violation, the actress grasps the bottle with the same typical clench grip but the bottle is upside down, preventing her from drinking. In the photograph presenting a grip violation, the actress grasps the bottle with an atypical grip from below that does not prevent her from drinking. The photograph combining grip and goal violations displays the actress grasping the upside-down bottle with the atypical grip. Action photographs are then divided into eight prime-target pairs according grip similarity, goal similarity and response type. An additional neutral prime-target pair was included in Experiment 2. The FSIM index provides a measure of low-level perceptual similarity between prime and target. .... *- 97 -*

Figure 22. Individual mean priming effects in milliseconds according to grip, goal and prime duration. Prime durations of 66 and 300 ms were tested in Experiment 1 (both yes and no response here), and prime durations of 120 and 220 ms were evaluated in Experiment 2 (only yes-response displayed). Small dots represent different individuals. Black dots represent the mean priming effects. Error bars represent standard error. .... *- 101 -*

Figure 23. Individual mean priming effects in milliseconds according to grip, goal, prime duration and experiment type. For 66ms of prime duration, data from Experiment 1 (Initial) and Experiment 3 (Replication) are displayed. For 220ms of prime duration, data from Experiment 2 (Initial) and Experiment 3

(Replication) are displayed. Overall, the same pattern of data is observed. Small dots represent different individuals. Black dots represent the mean priming effects. Error bars represent standard error. .... *- 119 -*

Figure 24. Design of the visual search experiment. .... *- 126 -*

Figure 25. Mean fixation proportion and standard errors (error bars) over time as a function of image condition. The dashed-rectangle represents the window of interest for the analysis. .... *- 131 -*

*Figure 26. Model fit of the data (lines) and actual data (dots) for the Grip-distractor and Goal-distractor***.** *- 132 -*

Figure 27. Design and procedure of the experiment. ***A.*** *Stimuli were divided in four experimental conditions by manipulating the typicality of the grip and the visual goal of the action. A fifth neutral condition was added as a control of object information.* ***B.*** *Procedure on a given trial. Responses were prompted for only 12% of the trials.* .... *- 148 -*

Figure 28. ***A.*** *Scalp map distribution corresponding to the two identified time-regions. Bold circles indicate the electrodes that have been averaged to obtain the mean amplitude of the P100, N170 and P300 respectively.* ***B.*** *Grand average ERP at the posterior site.* .... *- 151 -*

Figure 29. ***A.*** *Scalp map distribution corresponding to the two identified time-regions. Bold circles indicate the electrodes that have been averaged to obtain the mean amplitude of the N300 and N400 respectively.* ***B.*** *Grand average ERP at the anterior site.* .... *- 152 -*

Figure 30. ***A.*** *ERP as a function of grip typicality and goal typicality at the posterior site. Ribbons represent standard errors.* ***B.*** *Individual mean estimates of the main effect of grip typicality (blue dots), main effect of goal typicality (green dots) for the P100, N170 and P300 components. Black dots represent overall mean and error bars represent standard errors.* .... *- 154 -*

Figure 31. ***A.*** *ERP as a function of grip typicality and goal typicality at the anterior site. The green font represents the time-window of the N300 component. The orange font represents the time-window of the N400 component. Ribbons represent standard errors.* ***B.*** *Individual mean estimates of the main effect of grip typicality (blue dots), main effect of goal typicality (green dots) for the N300 and N400 components. Black dots represent overall mean and error bars represent standard errors.* .... *- 156 -*

Figure 32. Schematic representation of a TMS-priming paradigm. Without TMS, related prime-target pairs (in green) will be associated with faster response times on target judgement (e.g., same or different as the prime?) than unrelated prime-target pairs (in black). When TMS is applied at the target onset, related prime-target pairs will be associated with slower response times than unrelated prime-target pairs. TMS disturb the information preactivated by the prime. .... *- 167 -*

Figure 33. Trial procedure of the TMS-priming paradigm .... *- 170 -*

Figure 34. Brain areas stimulated. IFG: Inferior Frontal Gyrus; IPL: Inferior Parietal Lobule .... *- 172 -*

Figure 35. Individual mean RT in milliseconds as a function of TMS-condition (IPL, IFG or Sham), Response-type (yes or no), Grip-similarity ("Grip-similar" or "Grip-dissimilar") and Goal-similarity ("Goal-similar" or "Goal-dissimilar"). Error bars represent standard errors. Each dot represents an individual mean. Black dots represent the mean averaged over the participants. ....- 173 -

Figure 36. Individual mean difference in terms of partial repetition cost (i.e., difference between situations in which only one dimension is shared between the prime and the target and situations in which both grip and goal, or neither of them, are shared between the prime and the target) in milliseconds between IFG and sham on the right, and between IPL and sham on the left according to response type when both Grip and Goal were similar between the prime and the target versus when only one of the two dimension. Each dot represents an individual mean. Black dots represent the mean averaged over the participants. Error bars represent standard error. ....- 175 -

Figure 37. Impact of the sense of social power on goal priority via dominance. Bêta represents the estimate of the model for social power (path c) and dominance (path b) predicting goal priority and the estimate for social power predicting dominance (path a). ....- 195 -

Figure 38. Correlation between the goal-priority index and the score of dominance. The higher the score of dominance, the lower the goal priority index. ....- 196 -

Figure 39. Correlation between the goal priority index and the score of social power. The higher the score of social power, the higher the goal priority index. ....- 197 -

Tables

Table 1. Results of the principal component analysis. .... *- 83 -*

Table 2. Demographic data for Experiment 1, 2 and 3. .... *- 114 -*

Abbreviations

General

e.g.	Exempli gratia
i.e.	Id est
vs.	versus
cf.	Confer

Statistics

ANOVA	Analysis of Variance
M	mean
SD	Standard Deviation
SE	Standard Error
RT	Response Time

Neuroscientific methods

*f*MRI	Functional magnetic resonance imagery
EEG	Electroencephalography
ERP	Event related potential
TMS	Transcranial magnetic stimulation
MEP	Motor evoked potential

Brain regions

AON	Action observation network
IFG	Inferior frontal gyrus
IPL	Inferior parietal lobule
LOTC	Lateral occipitotemporal cortex

Preface

“La plupart des gens, quand ils bougent, eh bien ils bougent en fonction de ce qu’il y a autour d’eux. Juste en ce moment, quand j’écris, il y a Constitution qui passe avec le ventre qui traîne par terre. Cette chatte n’a aucun projet construit dans la vie mais elle se dirige vers quelque chose, probablement un fauteuil. Et ça se voit dans sa façon de bouger : elle va vers. Maman vient de passer en direction de la porte d’entrée, elle sort faire des courses et en fait, elle est déjà dehors, son mouvement s’anticipe lui-même.”[1]

L’Elégance du Hérisson, Muriel Barbery

Most of our life has a social purpose. Most of our actions are pointless if there is no one to notice them, to evaluate them, to judge them. A Buddhist saying wonders *“Does the falling tree make any noise if there is no one to hear it?”*. At the same time, we also excel at evaluating others’ actions. We are not doing it solely on purpose, but spontaneously. It is almost impossible for us to not understand what others are doing and why they are doing it. To some extent, this is what this thesis is about. We investigated some of the bases of our abilities to make sense of others’ actions. Before any subjective and value judgements, there is the recognition of the action.

Yet, actions are goal-directed movements, and goals are hidden in the actor’s mind. The last decades have intensively contributed to demonstrate that goals are visible in the actor’s movements well before its end, but also that we foresee the actors’ upcoming actions through the information we have about their environment. In this thesis, we first review the concepts retrieved from the action planning and motor control literature on which action recognition is based. In the subsequent sections, we discuss the theories and data that have been gathered

---

[1] *“Most people, when they move, well they move depending on whatever's around them. At this very moment, as I am writing, Constitution the cat is going by with her tummy dragging close to the floor. This cat has absolutely nothing constructive to do in life and still she is heading toward something, probably an armchair. And this is visible in the way she moves, she goes towards. Mummy just heads towards the front door, she’s about to go shopping and, in fact, she’s already outside, her movements anticipate itself.” Personal translation.*

around action recognition and their limitations. We argue that improving our understanding of action recognition requires to sneak into the temporal dynamic of the process. After designing and describing an experimental strategy to do so, we provide empirical evidence in this direction. Through behavioural and neurophysiological methods, we explore the cognitive and neuronal foundation of action recognition. We tried to understand how the observers and their brain make use of the visual information they have access to. In a last part, we considered how the individual characteristics of the observers affect the way they recognise others' actions. Our aim was to tackle the issue of action recognition from every critical side: the cognitive mechanisms, their neuronal implementation and the impact of the inherent variability of the human subject.

Investigating the bases of action recognition is, I believe, an important requirement to further access to a broader understanding of human social cognition. Although language is an important part of social cognition, aggregating evidence demonstrate that many important parts of social cognition bypass the linguistic communications. Every part of the society, or at least each time several humans are involved, is concerned by social cognition. I believe that most of the challenge society is facing involve, at some point, social cognition. As the future is concerned with how humans are living together, improving our understanding of social cognition will have important implications on the way we will tackle these challenges.

Introduction

CHAPTER 1: THE COMPLEXITY OF ACTION

Actions are complex phenomena. In action processing (be it performed, perceived or imagined), researchers have investigated the nature of the motor act, the origin of its control, the influence of the environment or the influence of memory about the world and how to act upon it. They also have investigated the action plans on which isolated motor acts are implemented, the goals, intentions or desired outcomes that motivate or organise these action plans. Finally, the processes underlying these components, whether these are based on stored-knowledge or online computations, or the nature of the structure underlying the organisation of such components and the role of attention or consciousness in such organisation have also all been investigated. This short enumeration aimed at illustrating the various processes and components that are mentioned in various theoretical considerations in the literature of action processing. Most of these concepts have received, at least partially, some empirical and theoretical supports. The organisation of actions has been considered from very different point of view, but action concepts and their structuration have often been borrowed from the literature on action planning and motor control. In this first chapter, we will attempt to define what an action is. Our aim here is not to draw a general theory of action organisation, but rather to provide a definition that could be suitable for investigating the recognition of others' actions. We will first review some of the concepts used in the theoretical accounts of action recognition, then we will discuss each of the key concepts in more details.

1. Representations, components and dimensions in action processing

Throughout the 20th century, it appears clear that actions are not only the fact of peripheral activities but arose from central activity, that is, from the brain. Although it seems natural nowadays, the idea took a long time to arise and to be accepted. Some authors have extensively examined how such idea came out of history of science and philosophy (see for example Changeux, 1983; Jeannerod, 2009), and we will just accept it as an inherited and well-supported assumption. Jeannerod (1994) advocated that "*any study dealing with motor behaviour must take into account the fact that the overt component (the movement) is only part of the phenomenon as a whole. The hidden component (the representation) exists in its own right*" (p. 201). Our primary interest will precisely be these hidden components Jeannerod (1994) was referring to, and such hidden components of actions are actually the ones that action recognition theorists are mostly interested in, although the overt components of actions–not only movements, but also visual and social context or objects, among others, are here considered as being part of the overt components–are often the ones through which the hidden components of action are investigated[2] (but see Wong, Haith, & Krakauer, 2015 for a critique). In motor control and action planning, researchers try to understand how one perform a given action, whereas in action recognition, we try to understand how observers extract and process the components they have access to. Such processes necessarily involve the processing of overt components of the actions performed by the actor, but also the memory observers have about both the world and the actor (Bach & Schenke, 2017) along with some individual biases or individual preferences they have on the way they process an action.

When Jeannerod (1994) introduced the idea that actions were mainly the problem of the brain, he also advocated that action processing relied on representations. The term representation is almost never defined, but can be understood as some "mental" or "internalised" equivalent for coding in the brain (or in the mind) some aspects of the external world–the body being considered as part of the "external" world (Vilarroya, 2017). Formally

---

[2] In this respect, observers and researchers are somewhat equal, as observers often only have overt components of others actions to make sense of others' actions.

speaking, mental representations have been defined either as some internal symbols representing the external reality or as the processes using such symbols (Morgan, 2014; Pezzulo, 2011; Piccinini, 2018). In the 90s, representations were mainly considered as components that can be "*readily accessed consciously; images of visual scenes, faces, and words* [that] *can be generated and then described verbally*" (p. 187 Jeannerod, 1994)–the semantic representations in Jeannerod's terminology. In contrast, Jeannerod (1994) proposed that actions involve a particular form of representations: the pragmatic representations. Pragmatic representations are meant to be "*normally rapidly transformed into movements*" (p. 187 Jeannerod, 1994). Most debates on representations discussed their formats–e.g., semantic versus pragmatic as in Jeannerod (1994)'s theoretical considerations–and contents–e.g., does the brain represent knowledge about the mechanical properties of the world, or does it represent knowledge about the motor program to act upon it (Buxbaum, 2017; Osiurak & Badets, 2016). Representations are nowadays freely considered through very different levels, formats and contents as reviewed by Vilarroya (2017).

Another topic of discussion is the distinction between cognitive (i.e., "in the mind") and neural (i.e., "in the brain") representations. In cognitive neurosciences, this distinction is quite loose. Cognitive representations have often been merely superposed on the brain activity recorded through different methodologies. Such shortcuts are indeed questionable and some authors argued that the conceptual tools used to interpret neuroscientific data should be different than the one developed by cognitive psychologists (e.g., Kiverstein & Miller, 2015). In our understanding, cognitive representations are dissociated from neural ones in that the former may be agnostic regarding their neural implementations. In other words, if the idea that the brain is at the origin of cognitive processes is widely admitted, the components the brain is working with (i.e., the "code" or "internal symbols") may be different from the ones identified by cognitive psychologists[3] (Logothetis, 2008). Although cognitive neuroscience is aware of this issue, the distinction is rarely discussed. One of the reasons may be because researchers rarely employ methodologies that directly address those questions. This is the case of this thesis as well: the empirical strategy that will be used here will not allow drawing any clear-cut conclusions about the format and content of representations.

Adding complexity to the debate, representations can either be seen as "static" components or as "dynamic" components. Following Vilarroya (2017)'s analysis, neural representations are somewhat static. Representational processes are for example dissociated from situational processes (i.e., processes integrating past, present and future event; Vilarroya, 2017). In contrast, Pezzulo (2011) argued that "*representations are dynamical processes produced by the re-enactment of sensorimotor brain structures*" (p. 92). The idea that the brain used simulation mechanisms to represent events of the world has been already proposed by Jeannerod (1994, 1999). Thus, Pezzulo (2011) argued that representations may also be transient and produced only when necessary. Consequently, representations may be static stored-representations or transient, reconstructed on the go, representations.

Eventually, one of the most and only consensual property of representations is that they should be detachable from the elements they represent (Pezzulo, 2011; Vilarroya, 2017). In other words, representations stand for elements of the world regardless of the presence of these elements. As such, they exist independently from the elements they represent (Pezzulo, 2011; Vilarroya, 2017). This property points out towards the necessary distinction between the elements of world and their representation in the brain (or in the mind). Then, we go back to Jeannerod's distinction between hidden and overt components of actions. Jeannerod (1994,

---

[3] Vilarroya (2017) used the example of "colour". Colours may be an important representational component, yet colours are not directly perceived, but interpreted from the light signal received by the sensory neurons of the eyes.

1999) went even further by arguing that action processing involved simulation mechanisms, mainly to anticipate the consequences of actions and choose the most appropriate motor sequence but simulation mechanisms can also be used when actions are observed or imagined. Now that we admit the existence of representations of actions in the brain (i.e., instances that can stand for actions even when actions are not currently produced or observed), we will discuss in more detail the different action components the field of action recognition usually refers to. The term representation implies some hypotheses regarding the element on which it is applied. The thesis does not address these questions. Thus, we will mainly use the term “component” or sometimes “dimension”. An action dimension or component refers to some part of action, which may partly correspond to the notion of representation but may also encompass the external information that is represented, or even some processes applied to this “part”.

2. Components of action

Action recognition theorists often assume that actions are goal-directed movements. Actions are also assumed to pursue both direct and long-term goals, which determine the way actions are performed. Actions are finally assumed to be purposefully self-initiated, in other words, intentional. Jacob and Jeannearod (2005) and Kilner (2011) identified four main action components: the kinematic component (which encompasses the overt movement and is not represented per se), the motor component (which encompasses the central representation of the motor act), the goal component (which encompasses the representation of the direct purpose of the action) and the intentional component (which encompasses the representation of the overall reason or motivation to perform the action). The intentional component is thought to organise and specify the goal component, which itself specifies the motor component which ultimately triggers the kinematic component of actions. In other words, they suggested that the different action components are hierarchically organised. Thill et al. (2013) will go for the same components and organisation, but will refer to the intentional component as the higher-goal component to emphasise the long-term goal of the actor rather than its motivation. Rizzolatti and Fogassi (2014) adopted a somewhat similar conceptualisation but retained the term “intention”: Movements referred to the overt component of the behaviour; motor acts referred to the sequence of movements required to achieve a single motor goal; actions referred to the sequence of motor acts required to achieve a (distant-) behavioural goal. They further distinguished motor goals as the outcome of the motor act, and motor intentions as the final outcome of an action. The notion of intention here is different from the one used by Kilner (2011) and Jacob and Jeannerod (2005) as Rizzolatti and Fogassi (2014) suggested that the overall reasons motivating an action should be encoded separately from the action or motor outcomes. The goal component was even further sub-divided by Ondobaka and Bekkering (2013), who dissociated the perceptuo-motor component in which bodily movements and goal location (i.e., where the movement should be directed to) are encompassed and the conceptual component in which the functional purpose of the action is encompassed. Csibra (2008), instead of explicitly specifying different components, emphasised the hierarchical organisation of actions, and will just assume that different levels of the hierarchy encode different types of representation (see for similar conceptions Bach, Nicholson, & Hudson, 2014; Bach & Schenke, 2017; and to some extent, see Vallacher & Wegner, 1987, 2012). Finally, or perhaps initially because they were the first to rise the importance of dissociating different components of action representation, Grafton and Hamilton (2007) identified i) the kinematic component, which encompasses the reach trajectory, grip configuration, or means of the action, ii) the goal object component , which encompasses the target object to grasp, and iii) the desired outcome component, which encompasses the physical consequences of the action. The kinematic component in Grafton and Hamilton’s (2007) terminology would rather correspond to the motor component in other terminologies, as they were primarily interested in the brain representation of actions, and not in the overt part of actions.

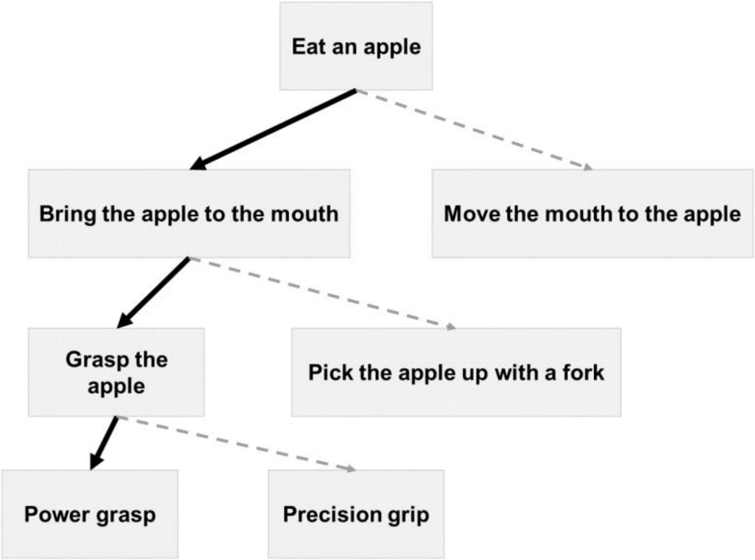


Figure 1. A simple action hierarchy. Inspired from Csibra (2008).

Summarising this short review, the motor components of actions seems to be dissociated from the goal components. Goals are usually dissociated between short-term and long-term goals. Reasons or motivations to perform a particular action seems sometimes to be equivalent to long-term goals, sometimes not. In the literature on action planning and motor control, different theories have tried to account for the different components. The importance of goals seems to be widely acknowledged, and their representational nature as well. The motor components of actions are more discussed, and although goals are commonly accepted as being at the origin of their organisation, the specification of motor components have sometimes been proposed to be based on representational processes, sometimes not. Intentions are also thought to be ultimately at the origin of our actions; however, whether intentions are responsible for the organisation of goals and motor acts has been discussed. One of the difficulties surrounding actions lies in their temporal and spatial complexity (e.g., Figure 1). Indeed, pushing a button (i.e., a single-step action) is considered as an action as much as making coffee is (i.e., a multistep action) or as fighting for freedom (i.e., an even more abstract action). In the following sections, we will discuss the notions of goals and the notion of motor acts. We will end the section on the notion of hierarchy. First, we discuss the notion of intention, and why we will not use it.

2.1. Aery intentions

When performing actions, humans have the feeling that they are at the origin of their actions. In other words, actions are not merely triggered by the environment, but by the individual. Many authors have agreed to attribute to intentions the ability to trigger actions, and to encompass the overall reasons and motivations underlying actions (Lashley, 1951; Scott, 2016). Intentions are generally thought to consist of "*a belief about the physical environment, a desire to change the environment, and an action plan to realise that change*" (p. 613; Uithol & Paulus, 2014). Intentions are also generally seen as discrete context-independent mental states (Pacherie, 2008; Uithol, Burnston, & Haselager, 2014). Causal action theories proposed that intentions both trigger and monitor the production of actions (Pacherie, 2008). Pacherie (2008) hypothesised three levels of intentions hierarchically organised. At the highest level, distal intentions are in charge of rational guidance and control, that is, they ensure that one's behaviour is in line with one's social and personal value (e.g., fighting for freedom would be at this level). Proximal intentions are in charge of situational guidance and control, that is they are one step further towards their actual implementation and are used to anchor one's behaviour in the now and then context (e.g., making coffee would be at this level). Finally, motor intentions are in charge of motor guidance and control, they are used to specify the motor parameters and monitor the execution of the motor acts (e.g., pushing a button would be at this level). The higher levels of intentions have the ability to influence the lower ones, and the lowers ones are able to provide feedbacks to higher ones. One of the advantages of Pacherie's model lies in its ability to speak about actions at very different levels of abstraction. Yet the context-independence and the discreteness of such construct have been criticised and seem to be irreconcilable with different sources of evidence suggesting an important role of contextual factors in determining actions (Uithol et al., 2014; Uithol & Paulus, 2014). For example, Scott (2016) provided extensive evidence for the influence of perceptual feedbacks in specifying the motor acts' parameters.

The main problem of intention as a concept lies in the various definitions it may refer to. The distinction of motor intentions and motor goals in Rizzolatti and Fogassi (2014)'s terminology may be similar to the distinction between proximal and motor intentions in Pacherie (2008)'s model. Yet it is unclear whether Rizzolatti and Fogassi (2014) keep the context-independence and discreteness characteristics of intentions. Besides, distinguishing goal and intention here adds unnecessary confusion, as motor intentions and motor goals seem to share the same format (i.e., they both represent "outcomes" but at different steps of action processing). One of the last characteristics of intentions is that they are usually considered as adopting a propositional format (Pacherie, 2008; Uithol et al., 2014; Uithol & Paulus, 2014). The notion of action intention in Rizzolatti and Fogassi (2014)'s terminology seems not to have this characteristic (Gallese, Rochat, Cossu, & Sinigaglia, 2009). When the propositional format is removed from this notion, most of the debates fades away evenly (Uithol & Paulus, 2014). The notion of intention as explicit, discrete mental states is adequate as such and following Uithol and Paulus (2014)'s suggestion, intentions may be best used to characterise how individuals make sense–consciously–of their behaviours (Vallacher & Wegner, 1989, 2012). Such conceptualisation of intentions may not only be used as some reappraisal processes, but may also help to plan our future actions: for example intention implementation is a procedure by which the use of an appropriate propositional format to formulate an intention helps its effective implementation later (Gollwitzer, 2012). From the action recognition point of view, such proposition is quite interesting because the mechanism individuals may use to attribute an intention to their own actions could be the same than the one used to attribute an intention to others. Intention attribution as we have defined it is rather an explicit and propositional phenomenon. Yet action recognition may not be all about intention attribution, and some implicit forms of action understanding through non-propositional format are indeed possible (Gallese et al., 2009; Uithol & Paulus, 2014). Here, we will just not consider these forms of action understanding as being intention recognition (Chiavarino, Apperly, & Humphreys, 2012; Uithol & Paulus, 2014).

2.2. The concept of goal

Goals are the most consensual component of actions. Goals are usually defined as a desired state of the environment the organism tries to achieve (Jacob & Jeannerod, 2005; Jeannerod, 1994; Kilner, 2011; Rizzolatti & Fogassi, 2014; Thill et al., 2013)–hence the notion of “desired outcome” of Grafton and Hamilton (2007). This idea takes its roots in the ideomotor principles (Greenwald, 1970; Shin, Proctor, & Capaldi, 2010). Ideomotor principles have been interpreted as followed (Greenwald, 1970; Shin et al., 2010): (1) a stimulus triggers a motor response that has some consequences on the environment; (2) with repetition of this phenomenon, an internal representation arises to represent the association between the stimulus and the consequences on the environment; (3) the internal representation gains the ability to trigger the appropriate motor response to provoke the consequences it represents. The term representation here refers to some central code stored in the brain. Shin et al. (2010) identified strong and weak ideomotor theories. Ideomotor theories are considered strong if the interface between action and perception does not require any additional cognitive processes. Ideomotor theories are considered weak if they require additional processes to translate perceptual input into motor output or to organise the motor pattern. Most of the theories we will discuss here can be considered weak forms of ideomotor theories, which are not *genuinely* ideomotor theories according to Shin et al. (2010). Rather, they merely exploit the general idea that movements are driven by their expected perceptual consequences.

In the motor control literature, the expected perceptual consequences are transformed into motor commands through inverse models (Miall, 2003; Wolpert, Doya, & Kawato, 2003). While such conception may account for the specification of the parameter of single motor acts, actions are often composed of sequences of several motor acts. Therefore, actions must be specified at a more complex level than just the single motor act (Jeannerod, 1994). In the middle of the 20th century, Lashley (1951) gave an influential lecture titled *"The problem of Serial Order in Behavior"*, considered by some as a turning point in Psychology (Botvinick, 2008; Summers & Anson, 2009). Lashley (1951) sheds light on the inability of the theorists of that time–who proposed roughly that any kind of complex behaviour could simply be explained by the mere juxtaposition of different behaviour's segments with each segment directly influenced by the one before (the associative chain theory)–to account for the organisation of complex sequences of behaviours. He observed that in most complex sequential behaviours, most of the units involved in a behavioural sequence could be organised in various ways, and that the *"individual items of the temporal series do not in themselves have a temporal valence in their associative connections with other elements"* (p. 116). Furthermore, different units in a sequence are usually performed too fast for any sensory feedbacks to trigger them one by one. Proprioceptive feedbacks require at least 25 ms to successfully affect the motor behaviour through the neurons in the spinal cord, 60 ms are necessary to observe an influence of visual feedbacks (Scott, 2016). Thus, Lashley (1951) hypothesised the existence of an organising agent that should be somewhat independent from the motor units (see Figure 2). Something he will refer later in the same discourse as *"the syntax of the act"* which he defined as *"a habitual order or mode of relating the expressive elements; a generalised pattern or schema of* *integration which may be imposed upon a wide range and a wide variety of specific acts”* (p. 122). Such syntax of the act implies that the brain should represent actions in more complex format than just at the level of motor acts.

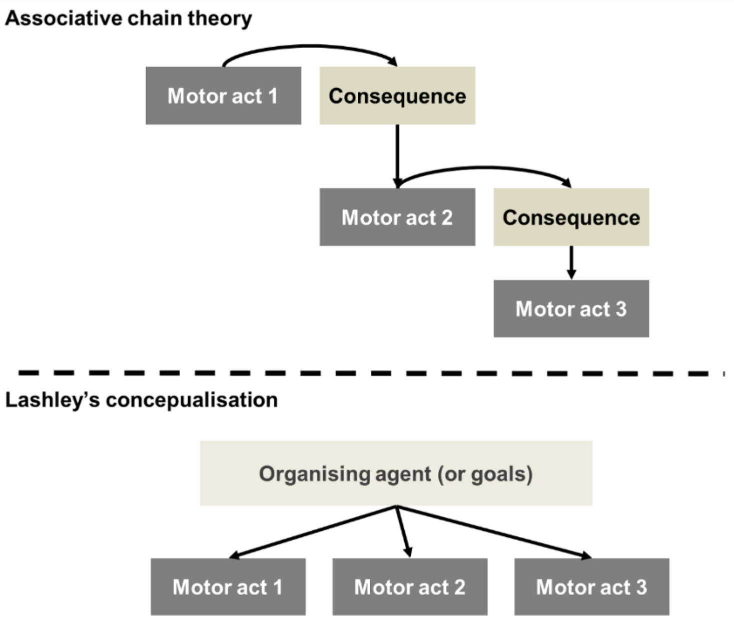


Figure 2. Schematic representation of conceptualisation before (upper part) and after Lashley (1951)'s proposal.

Between the decontextualised intentions and the very specific motor acts, Cooper and Shallice (2000) considered an intermediate level at which sequences of motor acts are implemented. In their model, complex actions are thought to be represented through schemas. Schemas consist “*of a goal, a triggering condition, an activation value, and a set of subgoals*” (p.892; R. P. Cooper & Shallice, 2006) Goals are conceptualised as “*a condition that may or may not be satisfied by the world*” (p. 305; R. P. Cooper & Shallice, 2000). Schemas are *“methods for achieving goals*” (p. 305; R. P. Cooper & Shallice, 2000). Several schemas can be selected to achieve the same goal, and the selection of a particular schema will depend on various parameters such as the available objects in the environment or some individual preferences. Schemas are conceptualised as methods, but each component of a schema is a subgoal. Whereas goals or subgoals can trigger different schemas, all components of a schema must be achieved for the schema to be completed. The ordering of each component is specified by preconditions that must be fulfilled for the component to be selected. Each level of the structure proposed by Cooper and Shallice (2000) is composed of goals (i.e., a set of preconditions) and schemas (i.e., methods to achieve goals). Goals (what to achieve) are there to bias the selection of schemas, schemas are there to organise an action sequence (how to achieve it). Only the lowest basic-level schemas are not composed of subgoals (i.e., schema's components) but correspond to discrete motor acts. The relations between the elements of the model of Cooper and Shallice (2000) are handled through excitatory and inhibitory relationships. The computational model has been developed to explain how routine actions such as making breakfast are performed, and more specifically to explain the errors patients could make during the performance of routine actions. Accordingly, the model was successful to explain omissions, anticipations, perseverations or object substitutions. The existence of such structuration of goals and sub-goals have also been supported by some clinical observation such as the phenomenon of "utilisation behaviours" (Lhermitte, 1983). Lhermitte (1983) reported patients suffering from frontal lesions who grasped and used objects of everyday life on mere visual presentation of these objects within their reachable space. This behaviour occurred even when explicitly instructed not to grasp the object, and without any motivational state to justify them. For example, patients could grasp a cigarette, then a lighter to light up the cigarette and smoke. Such behaviour occurs even if they have been told not to. Lhermitte (1983) suggested that these behaviours could be observed because the damaged frontal lobes could not inhibit the activity of the parietal lobe. Importantly, Lhermitte (1983) already emphasised that these behaviours could not simply be explained by online sensorimotor computations, as objects were used purposefully, but rather by some higher level action representations. Thus, actions are represented at higher level than just the motor acts.

The model of Cooper and Shallice (2000, 2006) was made to explain routine behaviours, and thus assumed that the use of attentional resources in carrying out everyday actions is minimal: goals and sub-goals can be activated without the constant need to deliberately activate them. Although goals are considered here as discrete representational units, they do not need to be consciously accessed and may not directly have a propositional format, as hypothesised for intentions. The definition of goals in Cooper and Shallice (2000, 2006) is in line with our own definition (i.e., “a desired state of the environment the organism try to achieve”), and with the notion of “desired outcome” from Grafton and Hamilton (2007). Goals have the ability to organise the motor behaviour, yet do not require the propositional format nor the conscious aspect of intentions. As such, we dissociate goals from intentions. In our view, goals are involved in performing actions, and are necessary to successfully organise different motor acts. In contrast, intentions are used to make sense of our behaviour. One criticism against intentions to organise the behaviour is that we can identify our actions and actions of others’ at very different levels (e.g., Figure 1). One may say “I am making sandwich” or “I am making dinner”, and making sandwiches being one step of making dinner. Identifying an action at a given level should be decisive in organising actions, yet there is no apparent reason to choose a level over another one if the consequences are the same (Uithol & Paulus, 2014). In addition, one is usually carrying out many actions at the same time, yet identifies one intention at a time (Uithol & Paulus, 2014). Consequently, intentions are inadequate to organise our behaviour, hence the need for some structures we are referring to as goals.

Goals are assumed to be, in the actor’s mind, internal representations and should not be directly observable. Therefore, from the point of view of action recognition, goals need to be inferred from visual information or knowledge the observer has about the actor. Attributing intentions is here considered as the explicit attribution of mental states to others, whereas action recognition aimed at making sense of others’ actions and help to draw expectations about others’ behaviours (Uithol & Paulus, 2014). The recognition of others’ goals could then be part of action recognition independently from intention recognition. Note that such proposition is fully in line with the definition of "intention" (i.e., long-term goals in our terminology) from Rizzolatti and Fogassi (2014).

2.3. The representation of the motor act

When Lashley (1951) introduced the notion of syntax of the act, he also suggested that some parts of the motor act were represented in the brain (Jeannerod, 1994; Summers & Anson, 2009). Clinical conditions such as the one reported by Lhermitte (1983) support the idea that some part of the motor act are prespecified. The motor components selected in the so called "utilisation behaviour" were indeed functional, and not merely based on the visual form of the objects. The literature on motor control has discussed such stored representations in terms of motor programs. Motor programs are usually seen as central representations of sequences of motor acts, that became stored in memory through practice (see Summers & Anson, 2009 for review). Theorists of motor programs usually admit that not every part of the sequence of motor acts are stored in motor programs and that some components of the motor act have to be specified during the task (notion of "generalised motor program"; see Schmidt, 1975)–although debates about how much is stored in the motor program are still undergoing. Acknowledging the paradox of the need to rely on both stored representations and online computations in action performance, early models from the limb apraxia literature hypothesised the existence of two routes by which actions could be performed (Gonzalez Rothi, Ochipa, & Heilman, 1991). Limb apraxia is considered as a disorder of "*skilled action, imitation, action recognition and (to a lesser degree) tool use.* [It] *occurs despite the relative unimpaired planning and programming of specific muscle movements. […] apraxia's primary hallmarks are the presence of errors in the shape, amplitude, and/or timing of movements […] in the pantomime of manipulable object-related […] actions to the sight of tools or upon command and in imitation of either or both meaningful and novel movements*" (Buxbaum & Kalénine, 2020). One route implies that the

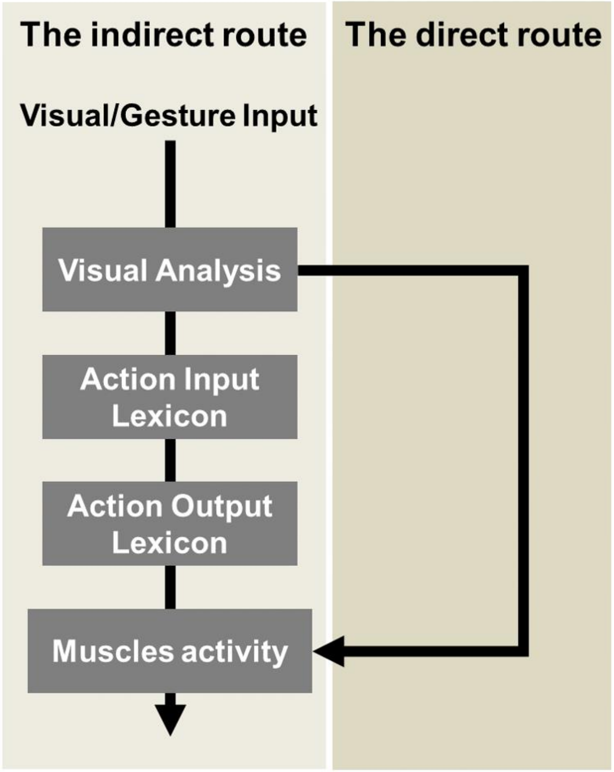


Figure 3. The two routes model of limb apraxia. Inspired from Rothi et al. (1991).

visual stimulus could be directly turned into a motor plan (the direct route). The other one involves the access to an action lexicon in which representations of known-actions are stored (the indirect route; see Figure 3). Inherited from these early accounts, several neurocognitive models of limb apraxia have tried to refine and accommodate the recent data from the neuroscientific literature. In this vein, Buxbaum and Kalénine (2010) proposed the "two action system" framework (see also Binkofski & Buxbaum, 2013; Buxbaum, 2017). The *structure-based* system is based on the structure of objects (e.g., size, localisation, shape etc.). The *structure-based* system is involved in object manipulation based on the online visuo-motor spatial transformations of visual input, and to produce transient representations of the motor act required to act on the object. The other *function-based* system is based on stored-representations of core features of skilled actions involved in the functional manipulation required to use an object. The *structure-based* system is supported by a dorso-lateral frontoparietal route involving the bilateral intraparietal sulcus (i.e., the dorso-dorsal stream). The *function-based* system is supported by a ventro-dorsal stream going through the left superior temporal sulcus and the inferior parietal lobule. The *structure-based* system is mainly useful in novel or unusual actions, whereas the *function-based* system is mainly useful in actions that has been performed multiple times.

The activation of the two systems is mainly driven by the environment itself, and such positions have been criticised (Osiurak & Badets, 2016). In contrast, Osiurak and Badets (2016) proposed that object-directed actions could be represented in terms of the effects they generate on the environment (i.e., our definition of goals). They propose that a motor pattern could be computed on the basis of the mechanical properties of the object in regards to the consequences one aims to generate. What is stored here is not motor components, but mechanical properties. In particular, Osiurak and Badets (2016) denied the role of any stored representation of information about the functional manipulation of objects, and they argued against the bottom-up view according to which the two systems may be activated without goals. The theoretical proposal of Osiurak and Badets (2016) is particularly useful to understand familiar and novel tool use. Yet it fails to account for the broad variety of evidence gathered in the literature on apraxia–in particular the various evidence supporting the role of action knowledge in purely perceptual task (Godard, Wamain, & Kalénine, 2019; Kalénine & Buxbaum, 2016; Kalénine, Wamain, Decroix, & Coello, 2016; Wamain, Sahaï, Decroix, Coello, & Kalénine, 2017)–which are best accounted by models hypothesising two routes for producing action knowledge (Buxbaum, 2017). Similar debates have been raised in the literature on motor control. Instead of motor programs, some have argued in favour of coordinative structures. Coordinative structures are built on the go to achieve a particular goal, and only for this particular goal (Summers & Anson, 2009). They do not require attention as their construction is based on the characteristics and constraints of the system (Summers & Anson, 2009). Roughly, if the system only allows to turn right, there is no need to specify this command. To some extent, the notion of inverse model seems quite close to these propositions, as the expected perceptual consequences (i.e., the goal) drive the specification of the motor parameters (Blakemore & Decety, 2001; Scott, 2016; Wolpert et al., 2003). Note that these propositions only increase the role given to goal representations and further strengthen its consensual importance.

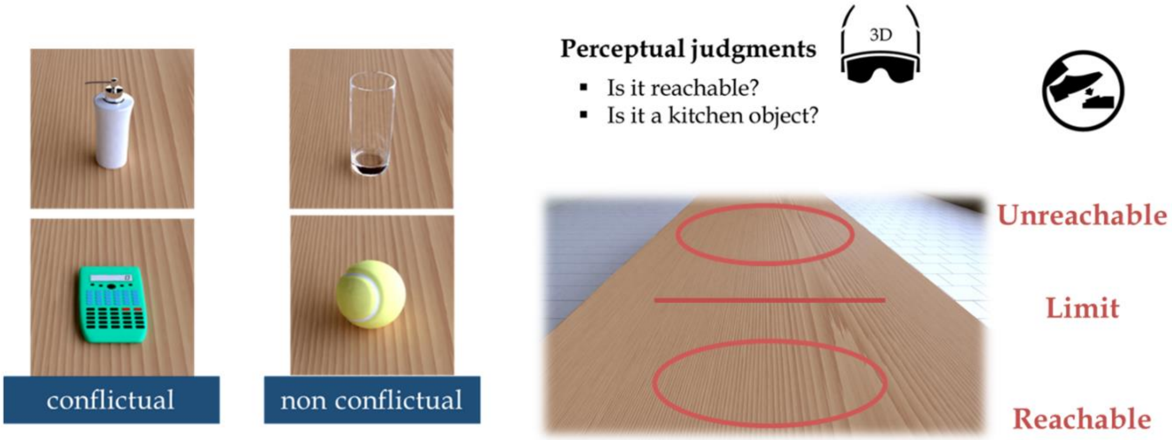


Figure 4. Example of task in which action knowledge influences perceptual judgement. Objects are visually presented either in the reachable space or not. Participants make perceptual judgements about the object (e.g., "is it reachable?" or "is it a kitchen object?") with their feet. Objects affording competing gesture representations (i.e., "conflictual", e.g., a calculator afford both a power grasp to move it and a poking gesture to make some calculations) elicit slower response times than objects affording non-competing gesture representations (i.e., "non-conflictual"). From Godard, Wamain, Delepoulle and Kalénine (2019).

In any case, all possibilities lead to representational processes (Pezzulo, 2011) and only differ in terms of the durability and the mode of activation of these processes. Coordinative structures and structure-based system involve transient representations of the motor acts, whereas function-based system involves stored and long-term representations of the motor acts. Coordinative structures and inverse models are triggered by goals whereas function-based and structure-based systems are triggered by the external environment. At the level of the motor act, debates are mainly related to what is actually represented and stored (Jeannerod, 1994; Schmidt, 1975; Summers & Anson, 2009).

From the point of view of action recognition, the idea of motor programs and stored-representations about prespecified motor parameters provides an easier solution than coordinative structures and inverse models (Buxbaum, 2017). Forward models (i.e., internal model allowing the prediction of the perceptual consequences of a planned motor act) has nonetheless been proposed to allow action recognition using, not the actor's motor plan, but the visual kinematics as a starting point (Blakemore & Decety, 2001; Oztop, Wolpert, & Kawato, 2005; Wolpert et al., 2003). Yet the specification of these models is complex and not easy to implement (Blakemore & Decety, 2001). In any case, even though most of the motor act parameters could be specified on the go, the impairments of patients suffering from limb apraxia goes outside of the mere action execution and affect tasks that do not require participants to perform actual actions (Buxbaum & Kalénine, 2020). The fact that some impairments affect tasks that do not require participants to actually perform an action suggests that some part of the motor acts may not be under the mere dependence of an action goal (Buxbaum & Kalénine, 2020; van Elk, van Schie, & Bekkering, 2014a). Instead, these observations suggest that some part of these motor acts are indeed stored. Such stored representations of motor acts have been proposed to be critical during object processing (Buxbaum, 2017; Buxbaum & Kalénine, 2010; van Elk et al., 2014a) and may serve action recognition as well (Bach et al., 2014; Bach & Schenke, 2017; van Elk et al., 2014a).

2.4. The hierarchical organisation of actions

When Lashley (1951) introduced his idea of a syntax of the act as an organising agent of the behaviour, he also introduced the idea that actions were hierarchically organised (Botvinick, 2008; Summers & Anson, 2009). Indeed, the organisation of the motor acts are thought to be under the dependence of goals. Most of our actions require different action steps, and thus different isolated goals are thought to be under the dependence of higher action goals (R. P. Cooper & Shallice, 2000, 2006). Although the necessary hierarchical organisation of action is widely acknowledged, its implementation is much more debated.

Uithol, van Rooij, Bekkering and Haselager (2012) discussed the role of action hierarchies in action processing. They identified two kinds of hierarchical organisation. On the one hand, a part-whole hierarchy in which every element of a given level of the hierarchy is constitutive of the elements higher in the hierarchy (e.g., Figure 1). On the other hand, a control hierarchy assuming that higher levels of the hierarchy control and modulate the lower levels in the hierarchy. Uithol et al. (2012) argued that although the two phenomena described by the two hierarchies are necessary for action processing, the two hierarchies in their form are conceptually incompatible. Part-whole hierarchies can only be built if the elements of the hierarchies are of the same nature. If the primary element is kinematic parameters, then the hierarchy cannot accommodate other formats such as perceptual consequences, goal representations or object representations. More importantly, if every part of the hierarchy is constitutive of every higher level, it is difficult to understand how any level could control the levels below. In contrast, control hierarchies, in which different levels are organised through causal relationships, imply that every level should be independent from one another, and thus reject the part-whole relationships. The problem lies in the specification of the levels of the control hierarchy. Although this problem is also present for the part-whole hierarchy, its importance for the control hierarchy is even bigger. If one speculates the existence of a sub-goal called “open fridge” to guide the selection of “full-hand grasp” and “pull handle”, then the “open fridge” needs to exist independently from the latter two elements. The control hierarchy also assumes that lower-levels are controlled by higher ones. It is known however that lower-levels are able to influence the higher ones, suggesting that the reverse communication is possible. Uithol et al. (2012) argued that in this case, it is not clear why the levels representing motor acts should be lower in the hierarchy than the levels representing the goals. Nonetheless, Uithol et al. (2012) did not argue against a hierarchical organisation. Instead, they argue for an implicit one. They advocate for a hierarchy in which higher levels are represented by neuronal populations with a more stable pattern of discharge than the lower ones. The more stable pattern of discharge would be able to influence or guide actions longer than the lower ones. Such kind of theoretical propositions remains to be experimentally evaluated. Besides, although the neuronal mechanisms are plausible, the number of levels in the hierarchy remains an unresolved issue.

The implementation of hierarchies in cognition is complex and difficult, and the problem spreads far outside the domain of action processing (Botvinick, 2008; Evans, 2008; Evans & Stanovich, 2013; Lashley, 1951; Toates, 2006). Although the existence of hierarchical organisations must be acknowledged, finding experimental evidence or implementing hierarchies remain challenging. As repeatedly acknowledged by several theorists, the number of levels in an action hierarchy is virtually infinite (Uithol & Paulus, 2014; Uithol et al., 2012; Vallacher & Wegner, 1989, 2012). Yet we believe that the infinite ways to define and describe actions are more related to the semantic organisation of actions than to the organisation of performed actions. As mentioned by some authors, the external and bodily environments provide sufficient information to greatly reduce the complexity of actions (Scott, 2016; Wolpert et al., 2003). The relationships between the organisation of actions in the semantic system and action recognition or action performance remain to establish. The part-whole hierarchy may be suitable for the organisation of action in the semantic system (Vallacher & Wegner, 1989, 2012), and this organisation may be more suitable for action recognition. Indeed, during action recognition there is no need for one level to control the other. Lastly, the format of each level of the hierarchy is more related to the question of the format of representations (Barsalou, Simmons, Barbey, & Wilson, 2003; Pezzulo, 2011) rather than to the hierarchical organisation itself.

3. Summary

Overall, the action recognition literature assumes that actions are goal-directed movements. Goals are seen as the representation of the expected consequences a given action aims to achieve, whereas movements, or motor acts, are the (stored or transient) representations of the motor parameters one needs to organise to achieve the action. Actions usually require different steps to be performed and the achievement of different goals. Thus, goals are thought to be themselves hierarchically organised. Theorists usually account for such organisation by dissociating short-term goals and long-term goals (Grafton & Hamilton, 2007; Jacob & Jeannerod, 2005; Kilner, 2011; Ondobaka & Bekkering, 2013; Rizzolatti & Fogassi, 2014; van Elk et al., 2014a). As we have briefly discussed above, the hierarchical organisation of action seems challenging to resolve. Regardless, in this thesis the hierarchical organisation of action will be assumed. We will not be interested in the hierarchy itself, but by how visual information received by the observer will affect their way actions are processed.

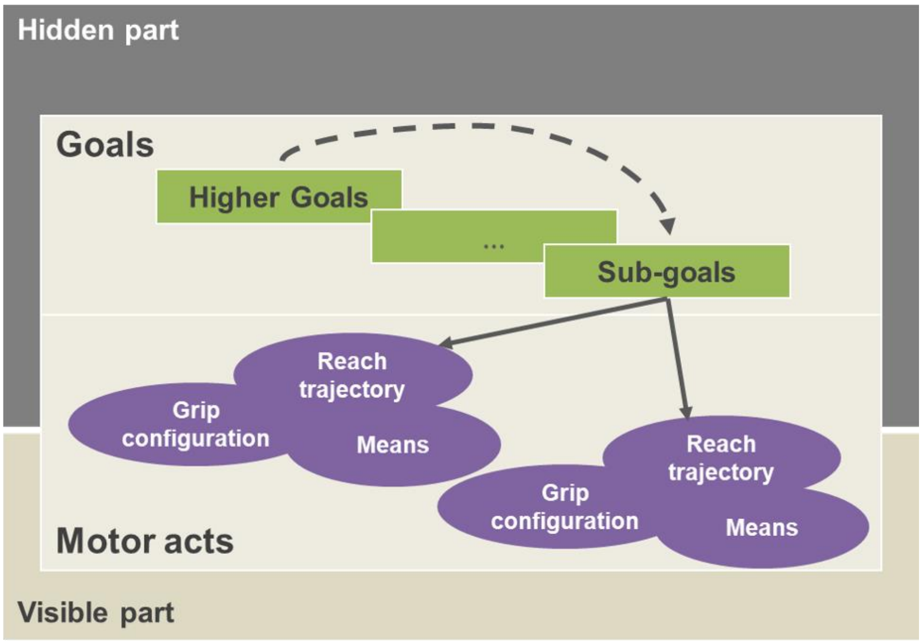


Figure 5. Summary of the different concepts used in the action recognition literature and their putative organisation.

To clarify our terminology, the terms dimension and / or component will be used to refer to both visual information and the representation that may be involved. Visual kinematics will be used to refer to the visual information provided by the movements. We will assume that some parts of the visual kinematics are stored as representations. We will also assume that goals are not visually accessible to the observer, and thus must be decoded from the visual information. To refer to the visual information that may inform about goals, we will use the term “goal-related information”. Action outcomes are indeed visual information accessible to the observer, but such information may not be represented as “goals” in the observers’ mind (i.e., they may not stand independently from the visual information).

The fact that some but not all of the visual kinematics may be represented in the brain indicates that action recognition is not a mere mechanism by which visual kinematics are matched with some stored representation. Instead, action recognition must involve some mechanisms in which the ongoing action is reconstructed in the brain. Many authors have proposed simulation as a mean to do so (Decety & Grèzes, 2006; Gallese, 2005; Gallese et al., 2009; Jeannerod, 1994; Pezzulo, 2011; Rizzolatti & Fogassi, 2014). Simulation is seen as a mechanism by which the brain recreates the action it is perceiving, which allows the recognition and the anticipation of others’ actions. How much simulation can inform about the actions of others during action recognition remains discussed. In the following chapters, we will explore the different ways theorists of action recognition have tried to conceptualise action recognition.

CHAPTER 2

THE PLACE OF MOTOR ACTS AND GOALS WHEN UNDERSTANDING OTHERS' ACTIONS: MIRROR NEURONS, SENSORIMOTOR AND PREDICTIVE APPROACHES OF ACTION RECOGNITION

In the previous chapter, we acknowledged that actions are complex phenomena involving different action components. Thereby, actions are composed of overt and hidden components. Visual kinematics are the main overt component. Representations of motor acts, and different levels of goal representation that are used to organise the motor acts are the main hidden components. Along with Uithol and Paulus (2014), we assumed in the first chapter that intentions may have a role in triggering or motivating an action, but little influence in directly specifying the parameters of the motor acts. We further postulated that action recognition is not all about intention recognition. Yet action recognition is not only about visual kinematics processing either. Several sources of empirical evidence suggest that when recognising an action, this action is understood in terms of goals and not only in terms of motor acts (Baldwin, Baird, Saylor, & Clark, 2001; Buresh & Woodward, 2007; Vallacher & Wegner, 1989, 2012; Zacks, Tversky, & Iyer, 2001). Even further, observers anticipate goals of actions well before the end of the action, and thus well before any observable outcomes (Ambrosini, Costantini, & Sinigaglia, 2011; Flanagan & Johansson, 2003; Flanagan, Rotman, Reichelt, & Johansson, 2013; Geangu, Senna, Croci, & Turati, 2015). Such observations have led researchers to consider action recognition mainly in terms of goal recognition. Making sense of others' behaviour is an old topic, but the discovery of a particular category of neurons in the monkey brain in the beginning of the 90's has fuelled a renew of interest in the topic (Heyes, 2010; Kilner & Lemon, 2013). Most of the theoretical considerations about action recognition are now stamped with the aim of integrating these neurons' behaviour in models of action recognition. In this chapter, we will review this literature, starting with the initial discovery of these neurons, the so called mirror neurons.

1. Discovery of mirror neurons and first interpretations

Mirror neurons were first discovered in the monkey brain using single-cell recording method, which remains the gold standard methodology to establish the presence of these neurons (see Figure 6). This method is challenging to use on humans, and most of the available data uses indirect evidence to demonstrate and to investigate mirror neurons in humans. Thus, we first discuss the discovery of mirror neurons in the monkey brain before discussing their existence in the human brain.

1.1. In the monkey brain

In the beginning of the 90’s, Rizzolatti and his colleagues (Di Pellegrino, Fadiga, Fogassi, Gallese, & Rizzolatti, 1992) discovered neurons in the monkey’s brain having the same pattern of discharge during both the execution and the perception of the same action. These neurons will be later labelled “mirror neurons” (Gallese, Fadiga, Fogassi, & Rizzolatti, 1996). Soon after their discovery, it was suggested that these neurons coded for representations of actions at the motor level, and that these representations could be used to learn (Jeannerod, 1994) or to understand others’ actions (Rizzolatti, Fadiga, Gallese, & Fogassi, 1996). Initially discovered in the equivalent of the inferior frontal gyrus in the monkey (Di Pellegrino et al., 1992), mirror neurons have now been reported throughout the motor system of the monkey, including the ventral and dorsal premotor cortices, the primary motor cortex and in various regions in the parietal cortex (see Kilner & Lemon, 2013 for a recent empirical review). Mirror neurons have different degrees of specificity, but they mostly show a “broad congruency” between action and perception. In other words, they discharge during the execution of particular actions (e.g., a precision grip) and during the observation of similar, but not identical actions (e.g., power grasp, precision grip or grasping with the mouth; Catmur, Press, & Heyes, 2016; Gallese et al., 1996; Rizzolatti & Fogassi, 2014). One of the specificities of mirror neurons is that they discharge during what has been referred to as "meaningful movements", that is when movements are directed towards a goal (Rizzolatti & Craighero, 2004; Rizzolatti & Fogassi, 2014; Umiltà et al., 2001). Some neurons also discharge differently depending on the action sequence in which the same motor act (e.g., grasping) is embedded (e.g., reach and grasp to place or reach and grasp to eat; Fogassi et al., 2005), which suggests that they could code actions beyond the motor act level. Thus, mirror neuron theorists first suggested that mirror neurons were primarily involved in the representation of motor goals (Rizzolatti & Craighero, 2004; Rizzolatti & Fogassi, 2014). They will later argue that mirror neurons could also be involved in the representation of intentions. Intentions according to mirror neuron theorists correspond to the distant perceptual consequences expected from the action sequence in which the observed motor act is embedded. Few studies have investigated this issue, as acknowledged by Rizzolatti and Fogassi (2014) and this proposal remains the most discussed claim in the literature.

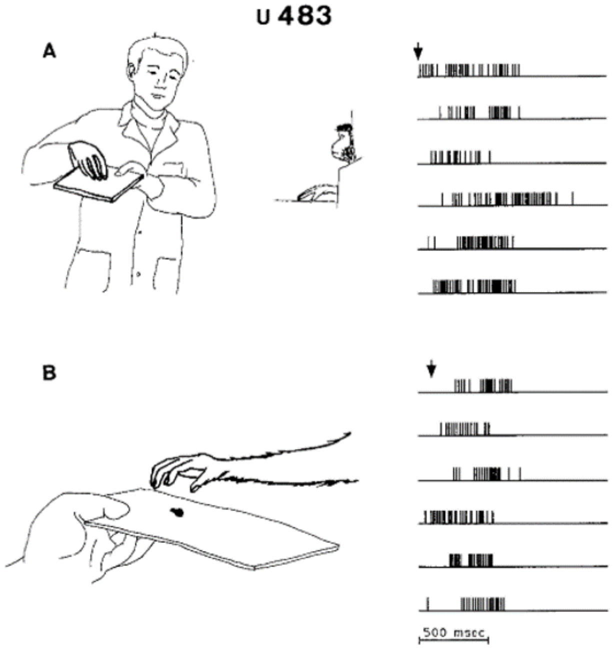


Figure 6. Example of a unit registered through single-cell recording selectively discharging during monkey grasping movements and during monkey observation of grasping movements made by the experimenter. A. The experimenter grasps the food. B. The monkey grasps the food. Arrows indicate the (approximate) onset of grasping. From di Pellegrino et al. (1992).

1.2. In the human brain

In humans, mirror neurons have been hypothetically localised in the inferior frontal gyrus and the anterior part of the inferior parietal lobule (see Figure 7; Rizzolatti & Craighero, 2004; Rizzolatti & Sinigaglia, 2008). Supports for such positions initially came from the neurophysiological literature. Soon after the first report of mirror neurons, and even before they were given their current name[4], Fadiga, Fogassi, Pavesi and Rizzolatti (1995) reported one of the first study on motor facilitation during action observation. They recorded the motor evoked potentials (MEP) of different muscles of the hand following transcranial magnetic stimulation (TMS) of the left primary motor cortex under different viewing conditions. TMS was used here to potentiate, and thus reveal, the covert activity of the motor cortex which under normal circumstances (i.e., without TMS) would not lead to any recordable (overt) activity of the peripheral muscles. Participants had to observe either (1) the experimenter grasping an object (different objects on different trials), (2) the object without any motion, (3) the experimenter moving his arm to draw a geometric shape in the air, or (4) to detect and signal a visual stimulus (e.g., a red dot flashing). Fadiga et al. (1995) reported higher MEP for action-related conditions (i.e., (1) and (3)) compared to visual stimulation non-related to action (i.e., (2) and (4)). Participants had no motor responses to provide, and the different conditions were equivalent regarding the attentional resources they required. Therefore, Fadiga et al. (1995) suggested that "*the observation of an action automatically recruits neurons that would normally be active when the subject executes that action*" (p. 2610).

---

[4] Di Pellegrino et al. (1992) first reported neurons discharging during both the observation and the execution of the same action, but Gallese et al. (1996) first used the "mirror-neurons" label.

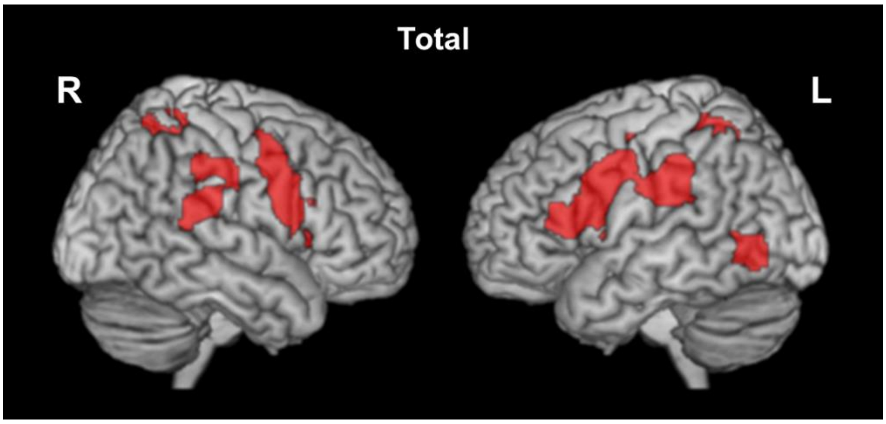


Figure 7. Mirror neuron network in humans according to the mirror neuron theorists. The human mirror neurons are thought to encompass mainly areas from the inferior frontal gyrus and the inferior parietal lobule. In red, the significant clusters of activation (FDR, p<.05) revealed by each ALE analysis superimposed on a rendered ch2better template using MRIcroN. From Molenberghs et al. (2012).

Subsequent studies will demonstrate some degrees of specificity both in terms of muscles and in terms of timing (e.g., Borroni, Gorini, Riva, Bouchard, & Cerri, 2011; Gangitano, Mottaghy, & Pascual-Leone, 2001). In other words, MEP patterns recorded during action observation are similar to the pattern of muscles activity that are recruited during the executed actions. Such modulations of the motor system activity are compatible with the mirror neuron proposal and have consequently been used to support such proposal. Functional magnetic resonance imagery (*f*MRI; e.g., Molenberghs, Cunnington, & Mattingley, 2012 and Figure 7) or mu-rhythm desynchronisation[5] (e.g., Fox et al., 2016) have also been used and keep being used to argue in favour of the existence of a mirror neuron system in humans. The only direct evidence to support the existence of mirror neurons comes from single-cell recording studies (Rizzolatti & Fogassi, 2014). As far as we know, only one single-cell recording study provides such direct evidence in humans (Mukamel, Ekstrom, Kaplan, Iacoboni, & Fried, 2010). Neurons from the temporal and medial frontal cortices were recorded in epileptic patients implanted with intra-cranial depth electrodes. Patients were asked to

---

[5] The suppression or diminution of a particular frequency-band of oscillation of the brain activity recorded over the central areas using electroencephalography (EEG).

observe video-taped or to perform a grasping action towards a mug. Neurons demonstrating mirror properties (i.e., discharge during both observation and execution) were observed in the supplementary motor area and the hippocampus. These data directly demonstrate the existence of mirror neurons in humans, although outside of the classical mirror neuron system.

1.3. *Proof of existence does not mean proof of function*

Concerns have been raised regarding the extent to which any of the neurophysiological methodologies used to investigate the involvement of mirror neurons can *actually* mark the activity of mirror neurons (see for example Dinstein, Thomas, Behrmann, & Heeger, 2008; Hickok, 2009). For example, Naish and colleagues (2014) have recently reviewed the literature on motor evoked potential (MEP) induced by TMS, and concluded that the mere observation of increased MEP during action observation is not a sufficient measure of mirror neuron activity. For example, MEP modulations measured before 200 ms of action processing show no muscle specificity (i.e., increased MEP can be obtained on a given muscle regardless of its involvement in the observed action), which should be expected if MEP would *actually* reflect mirroring activity. Thus, the mere activation of the motor system during action observation is not sufficient to attribute this activity to mirror neurons. Similar conclusions have been drawn independently for each of the neurophysiological indicators used to favour or to evaluate mirror neuron activity in the brain. Thus, evaluating mirror neuron activity requires particular paradigms, and mere similarities between action observation and action execution should not directly be attributed to mirror neuron activity. The critical point here is not to debate about the *existence* of mirror neurons in the human brain. We do have at least one piece of direct evidence which demonstrates their *existence* (Mukamel et al., 2010). Yet as some have previously noted (Dinstein et al., 2008; Hickok, 2009), the *existence* of such neurons should be dissociated from their *functional role*. In the following sections, we will discuss the role given to mirror neurons by different theoretical approaches, and the consequences of these theories for action understanding.

2. Sensorimotor approaches: Motor simulation guides action recognition

The discovery of mirror neurons and their potential abilities in coding short-term (motor goal in Rizzolatti's terminology) and long-term (motor intention in Rizzolatti's terminology) goals have led researchers to consider that action recognition could bypass any explicit and conscious appraisal of others' actions. In particular, it has been proposed that observers could rely on their own motor system to understand others' actions. We will start with discussing the initial interpretations of mirror neurons. These interpretations were quite ambitious in their formulation and claimed for a full-fledged understanding of others' actions (i.e., both goals and motor acts) through the motor system alone. In the second part, we will discuss more careful interpretations that now admit that, if the motor system may be involved in representing actions at the level of the motor acts, some additional structures may be required for representing goals. Finally, the involvement of the motor system in action recognition has been mainly conceptualised through mirror neurons, but action recognition may rely on the motor system without mirror neurons. We will exemplify such possibilities through the insights provided by motor control theories.

2.1. Initial sensorimotor accounts

Due to their involvement in representing actions both in action performance and in action recognition, mirror neurons have greatly contributed to the rise of the sensorimotor approaches of action understanding as a basis of social cognition (Gallese, 2005; Gallese et al., 2009; Rizzolatti & Craighero, 2004; Rizzolatti & Fogassi, 2014). In the field of action execution and motor (explicit) imagery (i.e., the voluntary and conscious activity of internally imagining an action), some authors have defended the idea that the same action representation could be recruited in different cognitive activities involving actions (Decety & Grèzes, 2006; Jeannerod, 1994, 1999, 2009). Jeannerod (1994) in particular suggested that any performed action should be preceded by a motor simulation in order to anticipate its perceptual consequences. Such implicit simulation should be spontaneous and effortless. Early on, mirror neurons were proposed as the neuronal candidates for such action representation (Jeannerod, 1994; Rizzolatti et al., 1996). Thus, Gallese (2005) suggested that actions could be understood with the same processes as individuals use for planning their own actions, namely by simulating their expected perceptual consequences. As the simulation taking place in action performance involves some equivalence between the motor act and the perceptual consequences, one may simply benefit from the same mechanism in understanding actions performed by others (see Rizzolatti & Craighero, 2004; Rizzolatti et al., 1996; Wolpert et al., 2003 for similar position). Gallese (2005) employed the term simulation as "*an automatic, unconscious and pre-reflexive functional mechanism*" (p. 41). Thus, action understanding according to the classical mirror neuron theorists is seen as a form of bottom-up propagation from the perception of the visual kinematics to the simulation of the expected perceptual consequences of the direct motor act and of the action sequence in which it is embedded–the "motor goal" and the "motor intention" respectively in Rizzolatti's and Gallese's terminology (Gallese et al., 2009; Rizzolatti & Fogassi, 2014)–within the motor system (see Zentgraf, Munzert, Bischoff, & Newman-Norlund, 2011 for similar description).

2.2. Mirror neurons represent associations not goals

Initial interpretation about mirror neurons assumed that they encode motor acts and goals. Yet mirror neurons may not encode the two types of action representations. Their place during action recognition can also be questioned. The associative learning account assumes that mirror neurons encode sensorimotor contingencies between visual experience and motor experience, instead of goals and motor acts (Catmur et al., 2016; Heyes, 2010). At the beginning, the links between the motor neurons of the frontoparietal network and the sensory neurons able to code for high-level visual properties of actions would be weak and inconsistent. Motor production of an action is frequently associated with the visual experience of the same action, such as during imitation or self-initiated action under visual monitoring (Jeannerod, 2009; Rizzolatti & Sinigaglia, 2008). The frequent co-activation of sensory neurons and motor neurons would lead to motor neurons being able to discharge during the mere perception of actions because of the link established with the sensory neurons (see Figure 8). Thereby, mirror neurons are not designed to be visuo-motor neurons, but rather, they acquire this ability with visuo-motor experience (Catmur et al., 2016; Heyes, 2010). This proposition is well-suited to explain why some mirror neurons show strict congruency between action performance and action observation, whereas others show less restrictive congruency (“broad congruency”) between performed and observed actions. The associative account suggests that mirror neurons in this case encode different sensorimotor

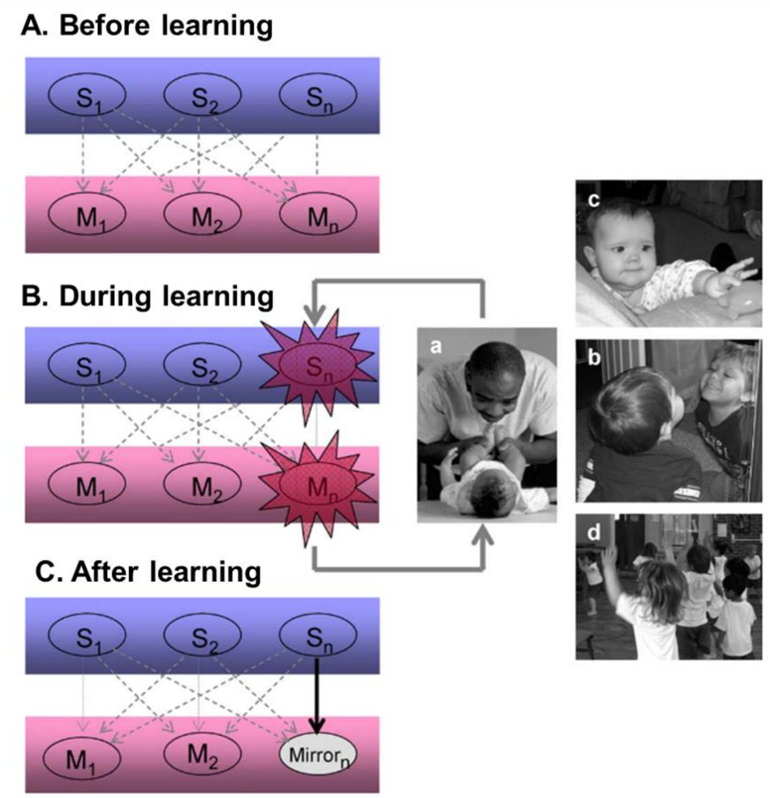


Figure 8. The three steps leading to the birth of mirror neurons according to the Associative Learning account. A. Weak connection between sensory and motor neurons. B. Frequent associations between motor neurons and sensory neurons. C. Motor neurons are now also discharging during the mere perception of action. From Heyes (2010).

relationships, rather than information about goals or motor acts. Following this framework, mirror neurons are not “made” for action recognition, though they may “happen” to be used during action recognition. As stated by Catmur et al. (2016), this account is indeed quite neutral on whether mirror neurons should have any role to play in action understanding or to what extent they are involved in action understanding.

Recently, Thompson, Bird and Catmur (2019) identified three different meanings underlying action recognition: the identification of the motor acts, the identification of goals, and the identification of intentions. They concluded that mirror neurons could be involved in the identification of the motor act component, but that there was little evidence to support their involvement in the identification of goals and intentions. Motor acts may then be retrieved from visual kinematics through mirror neurons activity before the involvement of additional processes to retrieve goals and intentions (Catmur, 2015).

2.3. Simulation and mirror neurons: an insight from motor control

The idea that action recognition may simply reuse the same principles and mechanisms as action execution has also been proposed in the motor control literature. Instead of turning goals into kinematics, visual kinematics are to be turned into goals. Motor control has been seen as the “*study of sensorimotor transformation*” (p. 593, Wolpert et al., 2003), and formalised as loops linking motor commands and sensory feedbacks (Scott, 2016; Wolpert et al., 2003). Through these loops, motor control is mainly concerned with the control of our own body. Wolpert et al. (2003) proposed that instead of trying to control the state of our body, we could as well use sensorimotor loops to control the state of others’ body. Wolpert et al. (2003) defined “state” as “*a set of variables which vary over time and when taken together with fixed parameters of the system [...] and the world are sufficient to predict the system’s future behaviour*” (p.593). In the motor loops, the system is one’s own body, whereas in the “social loops”, the system is the body and mental states of others. However, efficient motor control cannot rely on sensory feedbacks only, as these feedbacks take time (Scott, 2016; Wolpert et al., 2003). Instead motor control uses internal models to simulate the perceptual consequences of a given motor command. This allows the motor system to know whether the motor command should be successful or not without the need of sensory feedbacks. These models are developed through sensorimotor coupling, that is through the frequent association between motor commands and sensory feedbacks, and have been proposed to be at the origin of the motor representation of the world (e.g., Coello & Delevoye-Turrell, 2007; Hunnius & Bekkering, 2014). The brain could also develop models and representations about others using the same principles. Critically, Wolpert et al. (2003) argued that such outcome is possible only because our brains are similar. Assuming some similarities between how we carried out our own actions and how others performed theirs would be necessary for such models to be efficient. As mirror neurons seem to encode similarity between observed and executed actions, they may be good candidates to represent the required similarities between our own motor system and the motor system of others. Similar to the initial interpretations about the functional role of mirror neurons, internal “social” models importantly rely on visual kinematics to compute a model of others’ behaviours. Internal models are nonetheless built to provide precise predictions about the action outcome (e.g., is the football player about to shoot in the right or the left corner of the goal?). Miall (2003) argued that this characteristic of internal models is incompatible with the majority of mirror neurons that show broad congruency between executed and perceived actions. In addition, motor simulation through internal models has also been shown to strongly involve the connexion between the cerebellum and the posterior parietal cortex (see Miall, 2003 and Figure 9). In this case, motor simulation would not rely on the frontoparietal network and thus not on the putative human mirror neuron network. Internal models were first intended to explain the planification and control of actions at the level of the motor act. Wolpert et al. (2003)

nonetheless suggested that the principles occurring at the level of the motor act could be extended at different levels of goal representations (e.g., lower goals such as cutting a tomato and higher goals such as making a sandwich). In an attempt to reconcile the various propositions, Miall (2003) argued that if mirror neurons are inappropriate for representing the motor acts, they may be appropriate for representing different levels of goals. Motor simulation of the motor act through the processing of visual kinematics would then first rely on a cerebellum-parietal network, and goal representations would then arise from the mirror neuron activity in the frontoparietal network. It is interesting to note that such proposition would go against the conclusion drawn by Thompson et al. (2019). Indeed, motor simulation using mirror neurons would be involved in processing goals, and not motor acts.

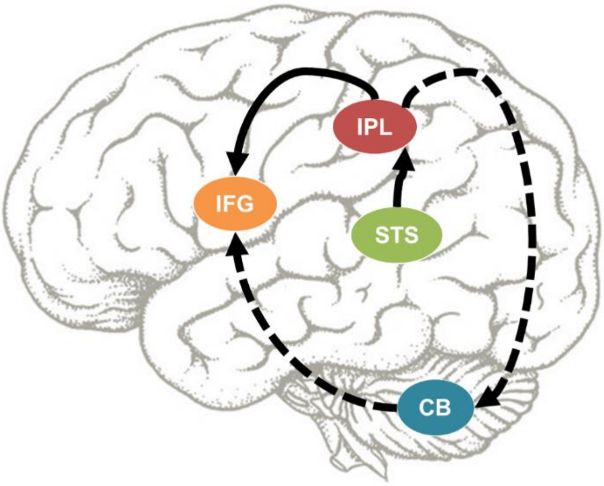


Figure 9. During action recognition, internal models could be computed through the mirror neuron network (Inferior Frontal Gyrus, IFG; Inferior Parietal Lobule, IPL; Superior Temporal Sulcus, STS; solid line) or computed through the temporal, parietal and cerebellar (CB) complex (dashed line). Inspired from Miall (2003).

2.4. Synthesis

The initial interpretations about the function of mirror neurons mainly defended their role in representing goals and motor acts, and that such shared functions were pretty convenient to recognise others’ actions. Later accounts discussed the extent to which mirror neurons could actually represent goals and motor acts. For some, mirror neurons are too specific to represent goals (cf. associative account), for others, mirror neurons lack of specificity to represent motor acts (cf. motor control insight). Motor acts may be recognised by mirror neurons to later lead to the representation of goals by different areas (cf. associative account). Motor acts may also be recognised through motor simulation occurring outside of the mirror neuron system (cf. motor control insight) before goal recognition occurring within the mirror neuron system. In any case, what gathers these different positions is that action recognition is driven by the motor simulation of visual kinematics. For that reason, these approaches will be referred to as the sensorimotor approaches.

3. Predictive approaches: Motor simulation does not guide action recognition

Soon after the formalisation of the first interpretations about the function of mirror neurons, several authors argued that motor acts and goals should be represented independently (Grafton & Hamilton, 2007; Jacob & Jeannerod, 2005). In Chapter 1, we argued that goals were required to organise motor acts, and that motor acts could not merely be under the dependence of environmental constraint (i.e., the execution of one motor act inside an action sequence cannot just wait perceptual feedback from the previous one to be triggered). Goals also provide flexibility to the actor, as the same goal can be achieved through different combination of motor acts. The availability of different strategies to achieve the same goal also implies that, in most situations, there is no one-to-one relationship between goals and motor acts. This one-to-many problem for action recognition has repeatedly been highlighted (Hunnius & Bekkering, 2014; Jacob & Jeannerod, 2005; Tidoni & Candidi, 2016). This problem is one of the main, and most common, criticism against the sensorimotor approaches. The ability of visual kinematics and motor acts to directly inform about goals is not straightforward, and motor simulation may not help to recognise the goal of others. In a first section, we discuss whether motor simulation should have any role in action recognition. In the two subsequent sections, we discuss the possibility that motor simulation may not inform about goals but instead be informed by goals. Following this position, action recognition should be driven by goals instead of being driven by the extraction of motor acts from visual kinematics.

3.1. The emulation account

Several empirical evidences suggested that the part of the mirror neuron network involved in representing goals were different from the part involved in representing motor acts (see Grafton & Hamilton, 2007 for review). Breaking down the mirror neuron network raised the question of how the brain is *actually* able to turn visual kinematics into motor representations. Rizzolatti and Craighero (2004) first proposed some kind of matching mechanisms where visual kinematics are simply mapped upon the corresponding motor representations. As discussed above, this mechanism barely stands against the one-to-many problem, and it is unclear how matching mechanisms could be distributed over the different parts of the mirror neuron network. Csibra (2008) came up with the same conclusions, and highlighted that it was unclear *what* was actually mirrored by the mirror neurons. Inspired by the imitation literature, Csibra (2008) dissociated *imitation*, which corresponds to the exact reproduction of an action, and *emulation* – or goal emulation –,which corresponds to the reproduction of the outcome of the action using the observer’s own means. Because actions can be described at different levels of generality, reproducing the same action could be done at different levels as well. If such statement goes for overt action reproduction, then it could go for covert action reproduction within the motor system. Thus, Csibra (2008) argued that action mirroring within the motor system is not the mechanism underpinning action understanding, but rather the product of the analysis. Actions are first processed and interpreted outside of the motor system, and then, the results of this process is fed into the motor system for emulation.

Importantly, Csibra (2008) did not deny the presence or the importance of simulation and mirroring. However, he did deny its primary, early and necessary role in action understanding. Therefore, simulation and mirroring would principally serve action anticipation.

With Csibra (2008), motor simulation through mirror neurons does not feed action recognition but instead is fed by it. The place of motor simulation, and even further of motor acts and visual kinematics, in Csibra (2008)'s account are almost anecdotic for action recognition. Indeed, he also suggests that visual kinematics may not be used at all, as it is possible to understand actions that do not involve visual kinematics. For example, Heider and Simmel (1944) developed a now classical paradigm in which participants are asked to describe the behaviour of triangles as they move around in a video clip. Based on the movement of these triangles, participants usually infer goals, intentions or emotions to each triangle, though there are no motor acts to retrieve. Consequently, motor acts seem to be a non-obligatory step for making sense of triangles' actions. Csibra (2008) did not deny that in some cases visual kinematics can be used, although visual kinematics should be processed outside the mirror neuron network; they are just unnecessary. Csibra (2008)'s account is quite powerful and thought provoking, but may be challenging to falsify. The use of visual kinematics may be informative or not, and the activity of the frontoparietal network can be observed without discarding this account. One possibility to falsify this account would be to demonstrate the causal role of the frontoparietal network during action recognition, which is out of reach of many experimental paradigms.

3.2. The predictive coding framework

Among the strongest accounts against sensorimotor approaches comes the predictive coding framework of Kilner, Friston and Frith (2007). These authors suggested that merely mapping visual kinematics onto one's own motor system would be possible only if there was a one-to-one relationship between goal and visual kinematics. As there are few situations in which such relationship can be established (Hunnius & Bekkering, 2014; Jacob & Jeannerod, 2005), this mechanism would be inefficient in most situations. Instead, they argued that the brain could make sense of visual kinematics through action predictions. Predictions about expected visual kinematics would be passed down from the higher levels of the action hierarchy to the lower levels–hypothetically from the inferior frontal gyrus to the superior temporal sulcus respectively. At each node of the mirror neuron network (i.e., frontal, parietal and temporal nodes), predicted information about kinematics would be compared to the actual visual information from observed kinematics. Predictions would be optimised using predictive errors until reaching a satisfactory threshold. Predictions would be generated through contextual information, which would be carried away from the occipito-temporal cortex up to the inferior frontal gyrus, as later proposed by Kilner (2011). Kilner et al. (2007) provided a formal computational and neurocognitive model along with their theoretical considerations (see for example Figure 10).

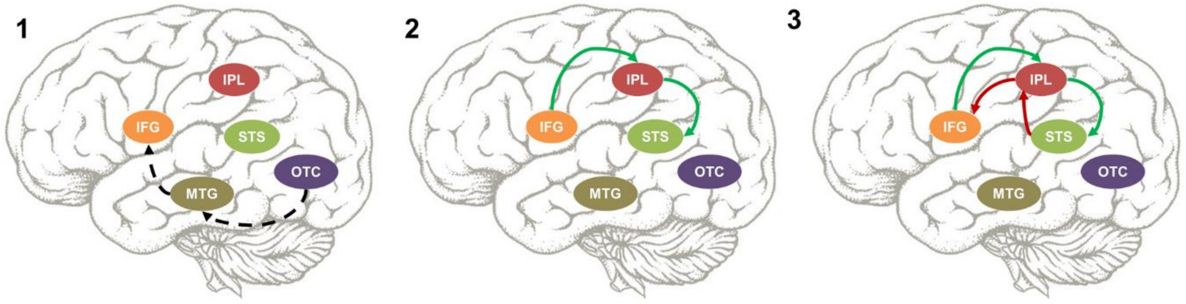


Figure 10. The three steps of action processing following the predictive coding framework. 1. Contextual information is conveyed from the occipital-temporal cortex (OTC) through the medial temporal gyrus (MTG) up to the inferior frontal gyrus (IFG; dashed arrows). 2. Contextual information in the IFG allows the computation of predicted kinematics, the prediction is passed down to the inferior parietal lobule (IPL) and the superior temporal sulcus (STS; green arrows). 3. Prediction errors are sent back to each node of the mirror neuron network (IFG, IPL) to improve the prediction (red arrows).

The model of Kilner et al. (2007) drastically differs from the initial interpretations about the role of mirror neurons as its formulation prevents any possibility to make sense of the visual kinematics in the absence of a prediction about the actor's goal. This approach takes its roots in the more general predictive coding framework of visual perception, which now spreads over all cognitive abilities (Allen & Friston, 2018). This model was later concretely evaluated with eye movement data (Donnarumma, Costantini, Ambrosini, Friston, & Pezzulo, 2017). When contextual information allows predictions about the action to be accurately generated, eye movements were proactive, that is the eye anticipated the next step of the observed action. In contrast, when predictions could not be generated, eye movements were reactive and mostly tracked the observed movements. In any case, Kilner's model implies that visual kinematics should not be the first information used by the observer. Instead, contextual information should be first processed to later guide the processing of visual kinematics.

3.3. The affordance-matching hypothesis

Developed by Bach, Nicholson and Hudson (2014), the affordance-matching hypothesis can be seen as an inherited account of the predictive coding framework. One of the important requirements for the predictive coding framework is the need for contextual cues from which observers could derive predictions about goals. Bach et al. (2014) proposed that during object-directed actions, objects could fulfil this requirement. The literature on object processing has established that both information about object function and information about object manipulation (i.e., motor information about how to purposefully use an object; e.g., a precision grip to write with a pencil) were included in object knowledge (see Buxbaum, 2017; van Elk et al., 2014a for extensive discussions on this topic). They suggested that the object alone is sufficient to provide information about both what kind of goals (i.e., through object functional knowledge) and what kind of motor acts (i.e., through object manipulation knowledge) one could expect from the use of the object. Functional knowledge would support action interpretation, as the object function allows to derive information about what kind of action goal one could expect. Manipulation knowledge would support action prediction, as how objects are handled can be used to anticipate the forthcoming action (e.g., a power grasp on a hammer may lead to expect a vertical up and down movement with the hammer). Bach et al. (2014) further suggested that both kinds of knowledge are related to one another, and that the activation of one leads to the activation of the other. For example, expecting a vertical up and down movement with a hammer from a power grasp is possible because the hammer is also associated with hammering a nail. A similar power grasp on a remote control would not lead one to expect the same vertical up and down movement. Depending on the flow of information during action recognition, functional knowledge and manipulation knowledge may help each other to reactivate goals and motor acts.

In this approach, the relationship between goals and motor acts is somewhat mediated, or at least moderated, by object knowledge. If we put together the predictive coding framework and the affordance-matching hypothesis, the functional and manipulation knowledge about objects could be used to derive prediction about goals and about the visual kinematics one can expect from the actor. The predicted visual kinematics could then be used to interpret the actual visual kinematics using the mechanisms proposed by the predictive coding framework.

3.4. Synthesis

After the first initial interpretations about mirror neurons, several authors questioned the necessary place of motor simulation and of the processing of visual kinematics in action recognition. In response, it has been proposed that motor simulation could be optional (cf. the emulation account) or at least involved much later during action recognition (cf. predictive coding framework and affordance matching hypothesis). In any case, in these approaches, visual kinematics cannot be interpreted through motor simulation. Instead, motor simulation requires some inputs that have already been interpreted to some extent. The only way to prove/disprove the emulation account would be to evaluate the critical role of the motor simulation during action recognition. The predictive coding framework and the affordance-matching hypothesis allow for more clear-cut hypotheses. As such, predictions about goals are thought to drive the recognition of others' actions, and thus, goal-related information should be prioritised during action recognition. These positions will be referred to as the predictive approaches.

4. General summary

The discovery of mirror neurons has deeply and durably impacted the field of action recognition, up to the point where dissociating questions about action recognition and questions about mirror neurons is not always easy. However, as recently stated by Enticott (2015), "[d]*espite a raft of studies using various methodological approaches, we appear to be moving further away from any form of consensus* [in the mirror neuron literature], *particularly concerning what this mirror system actually "mirrors" (e.g., low-level motor representation, goal or intention coding), and the functional significance (if any) of this mechanism*". Summoning mirror neurons outside of an identifiable framework seems devoid of meaning. Yet the developments about mirror neurons have contributed and keep contributing to highlight the role of the motor system during action understanding. In any case, some of the theoretical considerations discussed above can stand without mirror neurons. The motor system could be involved with or without mirror neurons. Observers could rely first on visual kinematics or not, with or without mirror neurons. What the action recognition literature seems to struggle with is the role of motor components in action recognition, and the extent to which visual kinematics can inform about action understanding on their own, and what kind of information they can provide. In this chapter, we have identified two main approaches differing on the place they give to motor simulation and visual kinematics. The sensorimotor approaches conceptualise action recognition with the first and critical involvement of the motor simulation and of the visual kinematics. In contrast, predictive approaches place motor simulation later in the action recognition process. In both approaches, motor simulation is involved and in charge, at least, of representing motor acts. Critically however, whereas the processing of motor acts allows for goal representation in the sensorimotor approaches, it is the processing of goals that allow for motor act representation decoding in the predictive approaches. In this thesis, we will mainly be concerned with the place given to visual kinematics during the recognition of others' actions, and thus, indirectly, with dissociating predictive from sensorimotor approaches. In the next chapter, we will review some of the evidence supporting each of the two approaches.

CHAPTER 3: WHAT CAN WE LEARN FROM BEHAVIOURAL AND NEUROPHYSIOLOGICAL STUDIES ON ACTION RECOGNITION? CONSIDERATIONS AND CURRENT DIRECTIONS

In Chapter 1, we introduced the different components of actions, and their theoretical importance for action recognition. Motor acts, goals, their visible and hidden parts (i.e., representations) have been admitted. In Chapter 2, we dissociated two groups of theories according to the place they give to visual kinematics and motor simulation during action recognition. Visual kinematics are thought to drive the recognition of others' actions for sensorimotor approaches, whereas observers cannot make sense of visual kinematics without a prediction about the action goal according to predictive approaches of action recognition. In Chapter 3, we will argue that the available evidence is not conclusive about whether the recognition of others' actions is driven by the first processing of visual kinematics or not.

1. The Action Observation Network

Neurosciences, especially motivated by the questions surrounding mirror neurons in humans, have importantly contributed to the improvement of our understanding of action recognition. Studies from this literature have led to the identification of a widespread network involved in the processing of others' actions, the action observation network (AON). The AON was endorsed by various methodologies and crystallised around three main meta-analyses: one for *f*MRI studies (Caspers, Zilles, Laird, & Eickhoff, 2010), one for transcranial stimulation studies (Avenanti, Candidi, & Urgesi, 2013) and one for brain-lesions studies (Urgesi, Candidi, & Avenanti, 2014). The AON is mainly organised around three nodes: the inferior frontal gyrus (IFG), the inferior parietal lobule (IPL) and what we will refer to as the lateral occipitotemporal cortex (LOTC; see Lingnau & Downing, 2015 for a detailed delimitation). *f*MRI studies in particular have further been used to demonstrate that different nodes of the AON should be involved in the processing of different components of observed actions (de Lange, Spronk, Willems, Toni, & Bekkering, 2008; Grafton & Hamilton, 2007; Hamilton & Grafton, 2006, 2008; Ondobaka, de Lange, Wittmann, Frith, & Bekkering, 2015; Southgate, Begus, Lloyd-Fox, di Gangi, & Hamilton, 2014; Spunt, Falk, & Lieberman, 2010; Spunt, Kemmerer, & Adolphs, 2016; Spunt & Lieberman, 2012; Wurm, Ariani, Greenlee, & Lingnau, 2015). For example, de Lange et al. (2008) used a violation paradigm in which action photographs of an actor using an object could be a) typical (e.g., using a cup near the mouth with precision grip on the handle), b) with an unexpected goal (e.g., using a cup near the ear with precision grip on the handle), or c) with an unexpected grip (e.g., using a cup near the mouth with power grasp on the cup’s body). Participants were asked in some trials to judge whether the action goal was typical, and in others to judge whether the grip applied on the object was typical. Irrespective of the task, they found that violations of the goal component increased activity in the inferior frontal gyrus (IFG) compared with a typical action. Conversely, a violation of the grip component increased activity in the lateral occipitotemporal cortex (LOTC; see Figure 11). Although these studies could provide some information about the place of visual kinematics in action recognition–visual kinematics processed by the LOTC may indicate early perceptual processing, and thus that visual kinematic processing drives action recognition–, the role of the different nodes of the AON in processing visual kinematics and goals is not always clear (Marneweck & Vallence, 2015). For example, the IFG has been found activated during both goals (Hamilton & Grafton, 2008; Iacoboni, Molnar-Szakacs, Gallese, Buccino, & Mazziotta, 2005; Romaiguère, Nazarian, Roth, Anton, & Felician, 2014) and grip processing (Hamilton & Grafton, 2008; Wurm & Lingnau, 2015). Similarly, the LOTC has also been linked to both goals (Romaiguère et al., 2014; Wurm et al., 2015) and grip processing (de Lange et al., 2008; Hamilton & Grafton, 2006). We already concluded in Chapter 2 that the distinction between sensorimotor and predictive approaches could not be settled through the mirror neuron literature. *f*MRI literature does not seem to be able to settle the problem either. The *f*MRI literature is not an isolate case: both reviews on transcranial magnetic studies and brain-lesions studies stressed out that the available evidence are barely sufficient to draw any clear conclusions on the role of each node of the AON during action recognition (Avenanti, Candidi, et al., 2013; Urgesi et al., 2014).

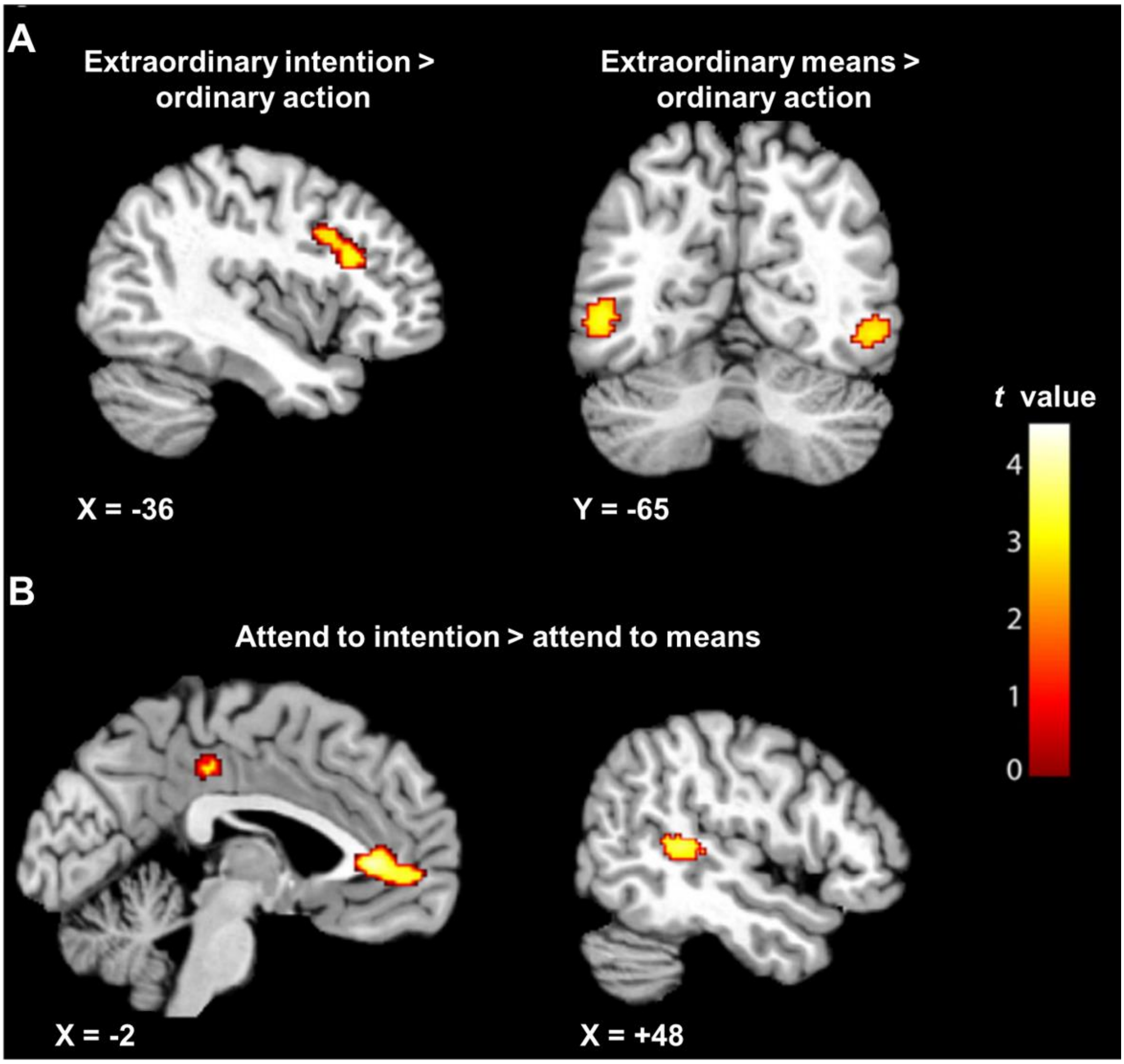


Figure 11. Results from de Lange et al. (2008). A. Goal violation (extraordinary intention) induces more activity in the inferior frontal gyrus (IFG) whereas grip violations (extraordinary means) induces more activity in the occipitotemporal regions. B. These results occur irrespective of the task given to the participants (i.e., attending to intentions or attending to means).

2. Empirical support for sensorimotor and predictive approaches

2.1. Evidence in favour of sensorimotor approaches

Support to sensorimotor approaches mainly comes from experimental evidence suggesting that visual kinematics are sufficient for observers to decode or to anticipate others' actions. Classically, in these paradigms, an actor is asked to perform an action with different goals in mind (e.g., grasping an apple to eat or to place; grasping an apple to use of oneself or to give to someone else). Visual kinematics are isolated from the action, either by modifying the video-clips or by using point-light displays. Finally, the modified stimuli are presented to observers that are usually asked to discriminate between two possible action goals: "is it to eat or to place?" or "is the action intended for oneself or someone else?" (see Figure 12). Using these paradigms, it has been demonstrated that participants are able to discriminate actions that are intended for oneself versus for someone else (e.g., Lewkowicz, Quesque, Coello, & Delevoye-Turrell, 2015; Quesque & Coello, 2015), or that observers could discriminate cooperative actions from competitive actions (Manera, Becchio, Cavallo, Sartori, & Castiello, 2011; Sartori, Becchio, & Castiello, 2011), or that observers could discriminate between grasp-to-use versus grasp-to-place actions (Naish, Reader, Houston-Price, Bremner, & Holmes, 2013; Ortigue, Thompson, Parasuraman, & Grafton, 2009). Some data even suggest that this ability could be reflected in the observers' own performed actions (Quesque & Coello, 2015). This ability is made possible by the fact that visual kinematics vary as a function of the actor's goals (Ansuini, Cavallo, Bertone, & Becchio, 2014; Ansuini, Santello, Massaccesi, & Castiello, 2005; Quesque, Lewkowicz, Delevoye-Turrell, & Coello, 2013) and provide sufficient information to discriminate between two different goals (e.g., Cavallo, Koul, Ansuini, Capozzi, & Becchio, 2016; Koul, Soriano, Tversky, Becchio, & Cavallo, 2019). Although recent theoretical accounts argue for more controlled experimental designs (Becchio, Koul, Ansuini, Bertone, & Cavallo, 2018a, 2018b), these data have nonetheless been interpreted in support of sensorimotor approaches, and corroborate the idea that observers spontaneously access to others' goals through the processing of visual kinematics (Ansuini et al., 2014; S. Gallagher, 2008; Rizzolatti & Craighero, 2004).

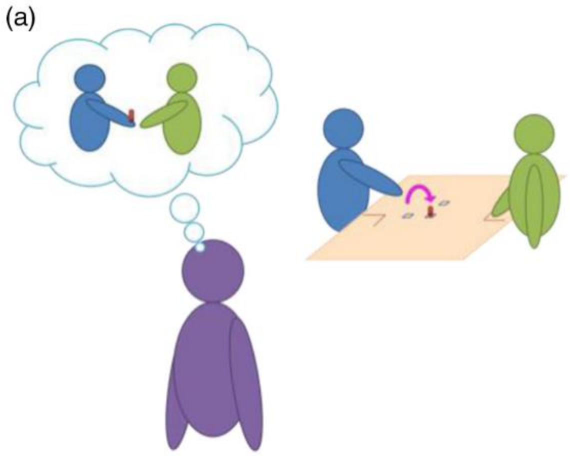


Figure 12. Example of classical tasks used in studies supporting sensorimotor approaches. Participants have to guess the goal of the actor (in blue) on the sole basis of the visual kinematics. These tasks usually involve objects, without any semantic content (e.g., a wooden stick). From Quesque & Coello (2015).

2.2. *Evidence in favour of predictive approaches*

Predictive approaches hypothesise that observers cannot make sense of visual kinematics without a prediction about the action goal. Already in the first theoretical accounts, proponents of predictive approaches stressed out the need of contextual information to make sense of visual kinematics (e.g., Kilner et al., 2007). The evidence claimed in support of these approaches mainly comes from studies demonstrating the influence of contextual information or task demands on action recognition. Thus, recognising that someone is squeezing a lemon is easier when the action is carried out in a kitchen than in a bathroom (Wurm & Schubotz, 2012, 2016). Some studies have demonstrated that the pattern of muscle activation during action observation (recorded using motor evoked potentials) first follows the environmental constraints rather than merely simulates the actor's visual kinematics (Cavallo, Bucchioni, Castiello, & Becchio, 2013; Koul et al., 2019). For example, if the actor can reach a wine bottle and grasp it on its lower part with a power grasp, the motor system will simulate a power grasp; now if there an obstacle forcing the actor to reach and grasp the wine bottle on its higher thinner part with a precision grip, the motor system will simulate a precision grip. In other words, motor simulation follows the prediction about the actor's goal based on the environmental constraints. In the same vein, motor simulation (as indexed by motor evoked potentials) was found facilitated (i.e., increased amplitude of the motor evoked potential) when actions are embedded in a congruent context (Amoruso, Finisguerra, & Urgesi, 2016; Riach, Holmes, Franklin, & Wright, 2018) whereas motor simulation was impaired when actions are embedded in an incongruent context (Amoruso et al., 2016; Amoruso & Urgesi, 2016). Using a different strategy, it has also been shown that the processing of visual kinematics can be biased by predictions about the actor's goal (Hudson, Bach, & Nicholson, 2018; Hudson, Nicholson, Ellis, & Bach, 2016; Hudson, Nicholson, Simpson, Ellis, & Bach, 2016). In these paradigms, participants were informed about the actor's forthcoming goal (e.g., "I'll take it" or "I'll leave it"), and were then asked to judge the position of a moving hand: the hand was perceived closer to the object than it really was if the announced goal was "to take it" and the hand was perceived further from the object than it really was if the announced goal was "to leave it" (see Figure 13). Interestingly, the effect of knowing the forthcoming goal was not affected by the statistical probability that the hand really fulfilled the announced goal (i.e., it did not matter that the hand faithfully produced the announced goal 75%, 50% or 25% of the time; Hudson et al., 2018; Hudson, Nicholson, Ellis, et al., 2016), neither by the probability that the action could be fulfilled in the displayed environment (i.e., it did not matter that the object to grasp was a cactus; Hudson, Nicholson, Simpson, et al., 2016). It was, however, affected by the announced predictability of the actor. In other words, participants were explicitly told that the actor could not be trusted, and that his words may be counter-predictive (i.e., "to take it" could now predicts a leaving action). Following these explicit instructions, knowing the actor's goal before judging the position of the hand had no more the expected effects (Hudson et al., 2018). In these experiments, explicitly knowing the actor's forthcoming goal and its trustworthiness (i.e., explicitly knowing whether he/she provides accurate information) was able to overcome any statistical relationships and affordances provided by the environment. Overall, these experiments suggest that the processing of visual kinematics can be biased by various non-motor goal-related information.

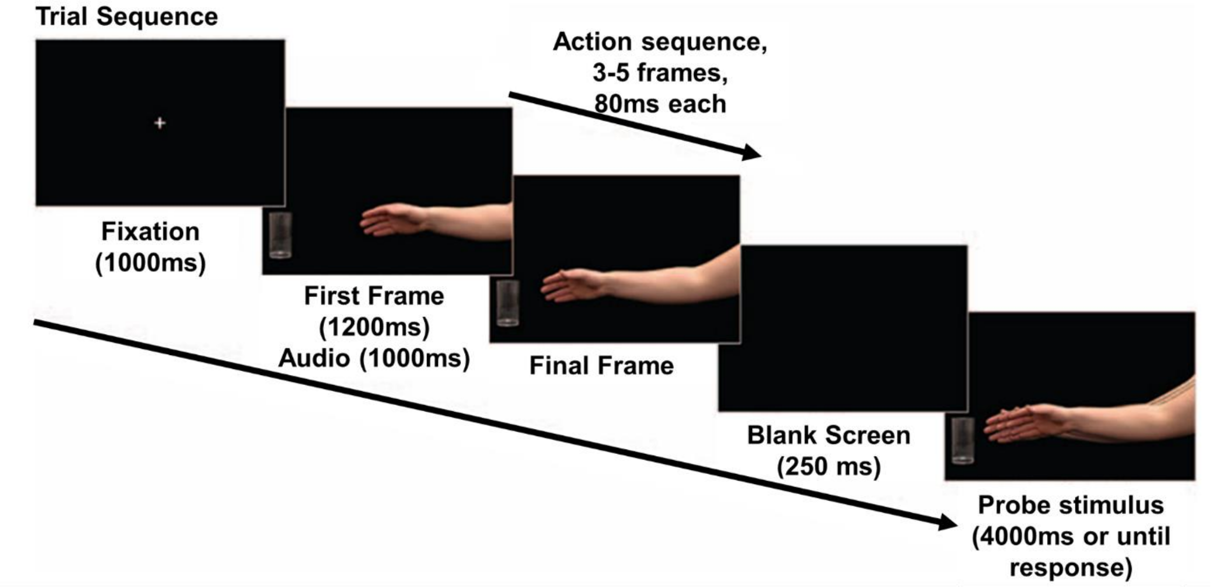


Figure 13. Example of trial sequence in the paradigm developed by Hudson et al. (2016, 2018). Once the fixation cross disappeared, participants heard "I'll take it" or "I'll leave it". Then, participants watched a hand moving back and forth towards the object until it stopped. After a blank screen, a static frame of the hand and the object appeared and participants had to say whether the hand was closer or further than the last frame of the video. Typically, participants will say "closer" if the announced goal was "I'll take it" and "further" if the announced goal was "I'll leave it". From Hudson et al. (2018).

2.3. Interim summary

To sum up, both sensorimotor and predictive approaches have their own sources of evidence. Studies in line with sensorimotor approaches have essentially contributed to demonstrate that visual kinematics provide sufficient information to discriminate different action goals, and that this information can be used by the observers, although the kinematic cues on which the actor relies on seem to vary from one study to another. Studies in line with predictive approaches have essentially contributed to demonstrate that the processing of visual kinematics could be affected by non-motor sources of information. These different sources of evidence have importantly contributed to demonstrate that both visual kinematics and non-motor goal-related information were involved in action recognition. Yet, they did not allow to determine which of the two approaches is favoured when recognising others' actions.

Experiments supporting the sensorimotor approaches usually involve poor contextual information about the action. Still, despite the poor contextual information provided, kinematic processing may still be guided by predictions about possible action goals, especially in studies requiring decisions between a limited subset of possible goals. In Lewkowicz et al. (2015), for example, video clips displayed an actor's arm reaching and grasping a dowel to move it away or take it for himself. Observers may first pre-activate the only two possible goals (move away versus take it) and then use kinematics information to verify each hypothesis. Actions would be processed in a predictive manner. Regarding the experiments supporting predictive approaches, the non-motor information is usually provided before the action, or the information provided by the visual kinematics is experimentally deteriorated. For example, in Wurm and Schubotz (2012, 2016)'s studies, although the actions are embedded in a visual context, the visual kinematics are blurred.

Consequently, although the processing of visual kinematics contributes to subsequent decisions about action goals, it does not inform about whether movement analysis is driven by the prior activation of a representation of the possible action goal. In addition, although contextual information influences the processing of observed actions, it remains to clarify whether predictions about possible action goals could guide action decoding in the absence of externally driven expectancies (visual scene, prior knowledge about other's goals). The two approaches are not necessarily mutually exclusive: predictive mechanisms could rather be involved when non-motor sources of information are available, whereas sensorimotor approaches would rather be at play when visual kinematics are the only source of information.

3. The temporal dynamics of action recognition

In the present work, we propose that sensorimotor and predictive approaches could be dissociated on the basis of the dynamic involvement of visual kinematics and non-motor sources of information during the processing of others' actions. Important theoretical accounts have previously highlighted the need to consider the processing of others' actions as a dynamic phenomenon (Catmur, 2015; Grafton & Hamilton, 2007; Kilner & Frith, 2008; Thioux, Gazzola, & Keysers, 2008) that cannot be fully uncovered without considering the dynamic processing of visual kinematics and non-motor goal-related information in action recognition. Thus, we believed that the literature should orient its questioning towards a more dynamic conception, trying to identify no more what action dimensions are involved in action recognition, but rather when they are involved, and for what purpose. The temporal dynamic of action recognition remains poorly investigated, despite its theoretical importance. Various methodologies can nonetheless be very informative about the dynamic organisation of the steps involved in action recognition. The evidence provided by some of these different methodologies are reviewed below. First, we discuss evidence coming from neurophysiological methodologies, then the evidence coming from behavioural methodologies.

3.1. Contribution of neurophysiological studies

The temporal dynamics of action recognition have been approached with different neurophysiological methodologies. Electroencephalography methodologies are particularly appropriate given their precise temporal resolution. Recently, motor evoked potentials induced by transcranial magnetic stimulation (TMS) over the motor cortex have also provided valuable information regarding the temporal dynamics of action recognition. Some evidence from these two methodologies are discussed here.

3.1.1. Electroencephalography studies

Beyond the theoretical divergences, the observation of several successive periods of stability in the brain activity (i.e. “micro-state”) when visually processing others’ actions (Avanzini et al., 2013; Ortigue, Sinigaglia, Rizzolatti, & Grafton, 2010) have further supported the idea that the recognition of others’ actions is a multistep process (see Figure 14).

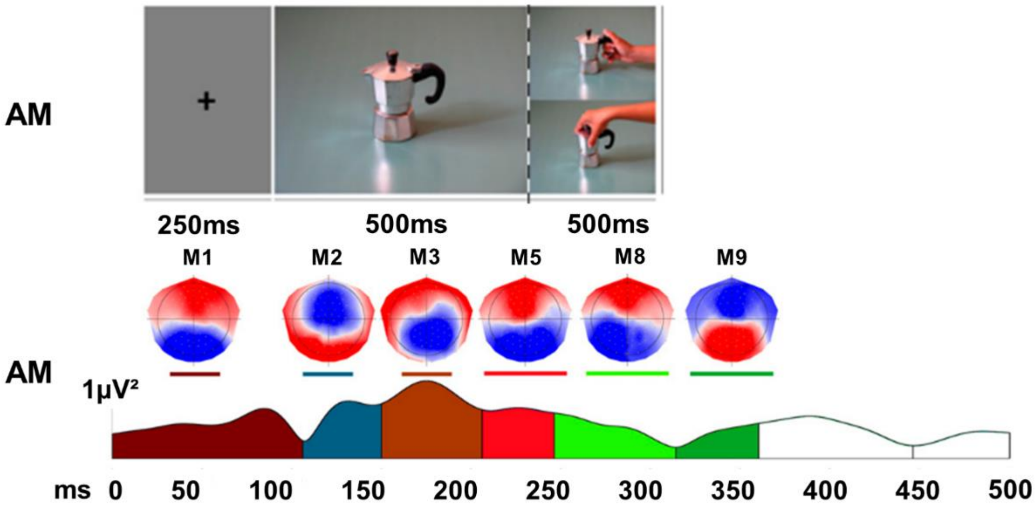


Figure 14. From Avanzini et al. (2013). Example of stimuli (upper part) for the apparent motion (AM) condition where a picture of an object is immediately followed by a hand-on-object picture. Below, the microstate segmentation results in the AM condition. Each micro-state is represented with a different colour. Micro-state represents temporary period of stability in the brain activity. Above each micro-state is displayed the corresponding map topography.

Electroencephalography studies have also helped to set up the boundaries of action visual processing (Avanzini et al., 2013; Ortigue et al., 2010, 2009; van Elk, Bousardt, Bekkering, & van Schie, 2012). In these studies, participants are generally asked to determine the goal of an object-directed action (e.g., is the hairdryer grasped to be used or to be moved away?). Determining others’ action goals impacted the brain activity in different time windows, the earliest starting from 60 ms of action visual processing (Avanzini et al., 2013; Ortigue et al., 2009), the latest emerging from 300 ms of action visual processing (Avanzini et al., 2013; Ortigue et al., 2010, 2009; van Elk et al., 2012). Event-related paradigms have further identified late neurocognitive mechanisms involved in the integration of different action components (see Amoruso et al., 2013 for review; Bach, Gunter, Knoblich, Prinz, & Friederici, 2009). These studies provide important empirical evidence defending the dynamic nature of action recognition. Nonetheless, as far as we know, the different action components have not been dissociated, which prevent us to draw any clear conclusions about which of the two mechanisms (sensorimotor or predictive) is favoured.

3.1.2. Studies using Motor Evoked Potentials

Some recent studies have suggested that the evolution of motor evoked potentials (MEP) across time when visually processing others' actions could reveal different processes involved during action recognition (Naish et al., 2014). In these studies, transcranial magnetic stimulation (TMS) is applied over the hand area of the motor cortex, and MEP are recorded from different muscles of the hand. The application of TMS allows to potentiate the activity of the motor cortex and to make visible the motor activity that would be normally not detected in periphery (i.e., in the muscles). TMS stimulations are time-locked to different steps of visual action processing (using videos or static photographs). Using such methodology, Cavallo et al. (2013) found that the pattern of muscle activity recorded on the observer's hand reflected first the action possibilities offered by the visual environment (e.g., using a full hand clench to grasp a wine bottle by its body / using a precision grip to grasp a wine bottle by its neck because of an obstacle) and only later the pattern of action visual context reflected the actual visual kinematics of the observed actions (see Figure 15). Consequently, in the critical condition, the actor is grasping the bottle without any obstacle with a precision grip on the bottle's neck: the pattern of muscle activity of the observer first reflected a full hand clench at the beginning of the movement, but at the end of the movement, the pattern of muscle activity of the observer reflected the observed precision grip. Such results indicate that the action context first influences the motor simulation, which is then modulated by the visual kinematics.

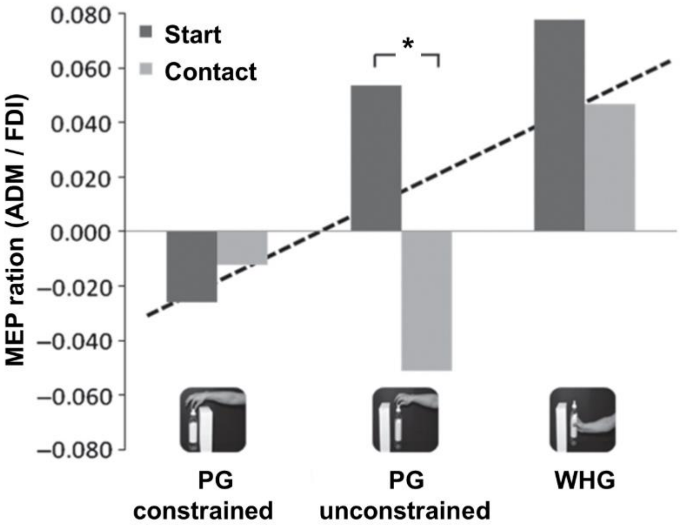


Figure 15. From Cavallo et al. (2013). Motor Evoked Potentials were recorded while participants watched an action video in which an actor reached and grasped a bottle. The bottle could be behind or in front of an obstacle. The actor could use a precision grip (PG) or a whole hand grasp (WHG). Two muscles of the hand were recorded (FDI: first dorsal interosseous; ADM: abductor digiti minimi). FDI is involved in both PG and WHG but ADM is much more involved in WHG. Thus, their ratio represents the relative simulation of PG versus WHG. MEP were recorded at the start of the video and when the hand reached the object. Results showed that in the PG-unconstrained, MEP first reflect the visual kinematics one could expect from the context (i.e., a WHG, indexed by more activity of the ADM), and only later reflect the actual visual kinematics (i.e., a PG, indexed by more activity of the FDI).

In a similar study, Amoruso, Finisguerra and Urgesi (2016) found that motor evoked potentials were facilitated around 240 ms of action visual processing when actions were embedded in a congruent visual context when compared to actions embedded in a neutral context. For 240 ms of action visual processing, no difference was observed between actions embedded in an incongruent context and actions embedded in a neutral context. In contrast, for 400 ms of visual processing, actions embedded in a congruent visual context were not different from neutral context, however actions embedded in an incongruent visual context impaired the generation of motor evoked potentials when compared to actions embedded in a neutral context. These results suggest that the influence of visual context is not identical at every time-point of action recognition. Although these data provide valuable insight about action recognition, the information they provide is necessarily bound to the activity of the motor cortex, and thus, are understood through the influence of visual kinematics and non-motor goal-related information on a particular step of action processing: the motor simulation. In Chapter 2, we have demonstrated that the role of motor simulation and its involvement in action recognition was unclear. In particular, it is not clear at which action step motor simulation occurs.

3.2. Contributions of behavioural experiments

Behavioural experiments refer to a range of various methodologies and paradigms. In the present section, we will first discuss the few behavioural studies that have explicitly varied visual kinematics and goal-related information independently, and subsequently show the limit of such approach. In a second part, we will discuss evidence from priming experiments and how they allow to provide some information about the timing of visual kinematics and non-motor goal-related information processing.

3.2.1. The relative involvement of visual kinematics and goal-related information

Assessing the timing of visual kinematics and goal activations during action perceptual processing requires to experimentally manipulate information about action kinematics and action goal independently. A few behavioural experiments have dissociated kinematic and goal-related information during the processing of visual actions (Kalénine, Shapiro, & Buxbaum, 2013; van Elk, Van Schie, & Bekkering, 2008). Kalénine et al. (2013) found that healthy participants were slower to determine that two videos displayed different actions when the two actions differed only in terms of kinematic parameters (e.g. applying detergent with circular versus straight wipe), or in terms of goals (e.g., applying versus removing detergent), as compared to when the two actions differed in terms of both kinematics and goals (see Figure 16). van Elk et al. (2008) showed that response times to determine that the goal of the action displayed in a picture was correct ("Is the object held at the correct goal location?" e.g., cup near the ear versus near the mouth) were slower when the grip component of the action was incorrect. Conversely, response times to decide whether the grip was correct or not (i.e., "Is the object grasped with an appropriate grip?" e.g., pencil grasped with power clench versus precision grip) were slower when the goal was incorrect. Moreover, interference from incorrect goals was greater than interference from incorrect grips. Overall, both studies support the hypothesis that visual kinematics and action goal representations play distinct but complementary roles in the processing of observed actions. However, though they provide empirical evidence about the relative "weight" of the different action dimensions in decisions about perceived actions, these studies do not directly inform about the timing of processing of grip and goal information during action decoding.

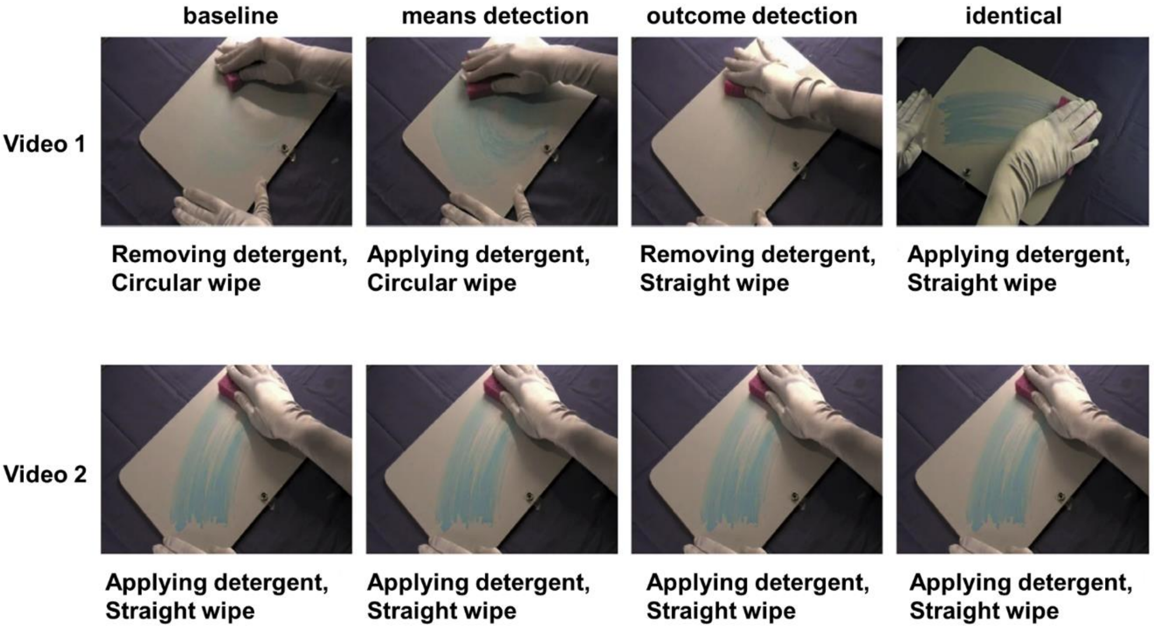


Figure 16. Stimuli used by Kalénine et al. (2013). Participants watched two videos and had to judge whether the two videos were the same or different. In the baseline pairs, the two videos were different both in terms of action goal (i.e., outcome) and in terms of kinematics (i.e., means) used to perform the action. In the means detection pairs, participants had to detect that the kinematics used to perform the action (i.e., applying detergent) were not the same in the two videos. In the outcome detection pairs, participants had to detect that the goal of the action was different between the two videos, although the kinematics used to do it were the same. In the identical pairs, participants had to say that the two videos were the same, although the visual perspective was not the same between the two videos. Overall participants were slower to answer when one of the two action dimensions were different (i.e., means detection pairs or outcome detection pairs) than when the two actions dimensions were different (i.e., baseline pairs).

3.2.2. Timing of action recognition processes: priming experiments

Among the behavioural strategies available, priming paradigms can provide valuable information on the temporal dynamics of action recognition. We will go back later to priming paradigms, but briefly these paradigms are based on the influence of the first presentation of a picture (the prime) on the processing of a second picture presented later (the target). By varying the duration of the prime, the experimenter can vary the availability of the visual information and then identify how much time is necessary for the different sources of information presented in the picture to influence target processing. Very few priming studies have used such procedures in the domain of action recognition. Cattaneo (2010) coupled a priming paradigm with transcranial magnetic stimulation (TMS) to investigate the critical (causal) role of the premotor cortex in processing visual kinematics. The TMS part is not of interest here, but this study was preceded by a behavioural priming experiment. Participants had to categorise target grip configurations (i.e., “has the hand moved slowly or quickly towards the object”; the task was mostly arbitrary, and participants had to make their decision subjectively) that could be primed by photographs displaying similar or dissimilar grips. Primes were available for 100 ms of visual processing. Cattaneo (2010) found that participants were faster to answer on target photographs when the same grip configuration as the target was presented in the prime. Timing was not an issue in this study, but results showed that 100 ms of visual presentation was sufficient to process static grip configurations. In another series of experiments, Costantini, Committeri and Galati (2008) investigated the role of the effector and the target object during the recognition of object-directed actions. In a go/no-go task, participants had to respond only when the presented action photograph was meaningful. The paradigm was not a proper priming procedure, but the experiment was designed in such a way that the action photograph to identify could be preceded by an action photograph that shared the same action goal or not (e.g., grasping, pushing, pressing etc.). In addition, the preceding action photographs could also share the same effector but a different target object or the same target object but a different effector, or both the same effector and the same target object. Action photographs were available for 150 ms of visual processing. The authors found that participants were faster to judge actions as being meaningful when the preceding action photographs shared the same action goal. Critically, whether the target action was preceded by an action photographs with the same effector or not, or the same target object or not, had no influence on facilitative effects. In other words, neither the effector, nor the target object were at the origin of the facilitation effect induced by the repetition of the same action goal. As acknowledged by Costantini, Committeri and Galati (2008), these findings need to be interpreted carefully as the effect stood on a null effect (i.e., the conclusions were drawn on the absence of significance, which can never really be proven). Nonetheless, 150 ms of visual processing seems to be sufficient for action goals to be processed. Overall, the available priming experiments were not directly designed to evaluate the timing of action recognition, but happen to incidentally provide information about it. In particular, they suggest that 100 ms of visual processing may be sufficient to process grip configuration, and 150 ms of visual processing may be sufficient to process action goals. Even more substantial, it draws attention towards priming paradigms as valuable paradigms to investigate the relative timing of visual kinematics versus non-motor goal-related information during action recognition.

4. General summary

In this chapter, we reviewed some experimental sources of evidence to evaluate which of the sensorimotor and predictive approaches were the most applicable to action recognition. We identified that sensorimotor and predictive approaches both had their own sources of evidence, relying on different types of paradigm. As a consequence, we proposed that the two approaches could better be dissociated on the basis of the temporal dynamics of visual kinematics and non-motor goal-related information. We further reviewed the available evidence providing information about the temporal dynamics of the processes involved in action recognition. Our investigation led us to the conclusion that the available evidence was not conclusive on when visual kinematics and non-motor goal-related were processed and involved in the recognition of others' actions. In particular, most paradigms did not dissociate visual kinematics and non-motor goal-related information, preventing us to draw any clear-cut conclusions about the origin of the effects. In addition, most paradigms were not directly designed to evaluate the timing of visual kinematics and non-motor goal-related information involvement in action recognition. In the next chapter, we will introduce a method to vary visual kinematics and non-motor goal-related information independently, and all along the thesis we will present different implementations of this method that will allow us to evaluate the temporal involvement of visual kinematics and non-motor goal-related information. The first involvement of visual kinematics would favour the sensorimotor approaches whereas the first involvement of non-motor goal related information during action recognition would be aligned with predictive approaches.

General Methodology

1. Overview of the section

As we have discussed all along the previous chapters, both visual kinematics and non-motor goal-related information are involved during the processing of others' actions. Evaluating the contribution of these two dimensions requires to tackle several issues:

(a) Making the visual information relevant for the two dimensions available at the very same time.

(b) Carefully dissociating the contribution of each dimension independently.

(c) Using a task that addresses equally the two dimensions.

Some of these challenges were addressed by van Elk, van Schie and Bekkering (2008), and we based our manipulations on their paradigm. In this chapter, we will first describe the paradigm developed by van Elk and colleagues (2008) and review the assumptions underlying such paradigm. We will then describe and justify our methodological choice. Second, we will describe the stimuli we developed that will be used all along this thesis. Finally, we will discuss and describe our statistical strategy.

2. An action violation paradigm: choice and theoretical justification

First, we describe the paradigm developed by van Elk and colleagues (2008) and review the assumptions underlying this paradigm. Subsequently, we detail and justify each part of the paradigm, namely why we choose object-directed action, which visual kinematics are targeted, which non-motor goal-related dimension we focused on, and the task participants will have to perform. Finally, the stimuli are described.

2.1. van Elk and colleagues (2008)'s paradigm

In the study of van Elk et al. (2008), participants had to evaluate the correctness of a) the grip and b) the action goal of object-directed action photographs. Action photographs involved an actor using an object across four situations: (1) applying a correct grip on the object with a correct goal; (2) applying an incorrect grip with a correct goal; (3) applying a correct grip with an incorrect goal; and (4) applying an incorrect grip with an incorrect goal (see Figure 17). Grips and goals were correct with respect to the prototypical use actions associated with the object. Grip configurations were correct if they displayed the typical grip associated with the typical use of the object. For example, a precision grip applied to a pen is typical ("correct grip"), whereas a power grasp is not ("incorrect grip"). Goals were correct if they displayed the object at the typical location of its use. For example, a glass is usually used near the mouth ("correct action goal location") whereas a glass near the eye is not ("incorrect action goal location"). In the following sections, we will favour the term "typical" instead of "correct" as all grip configurations are biomechanically possible and may be "correct" but "atypical" in some situations. For example, a child may use a pen to draw using a power grasp. Similarly, action goals may still be "correct" yet "atypical". A child may use a pen near his/her mouth to imitate an adult that applies lipstick on his/her mouth. In addition, note that we assumed that the goal dimension is not directly observable, instead it is the way visual information is varied that provides information about the actor's action goal. Thus, the location of the object in van Elk et al. (2008) provide visual information about the goal of the action.

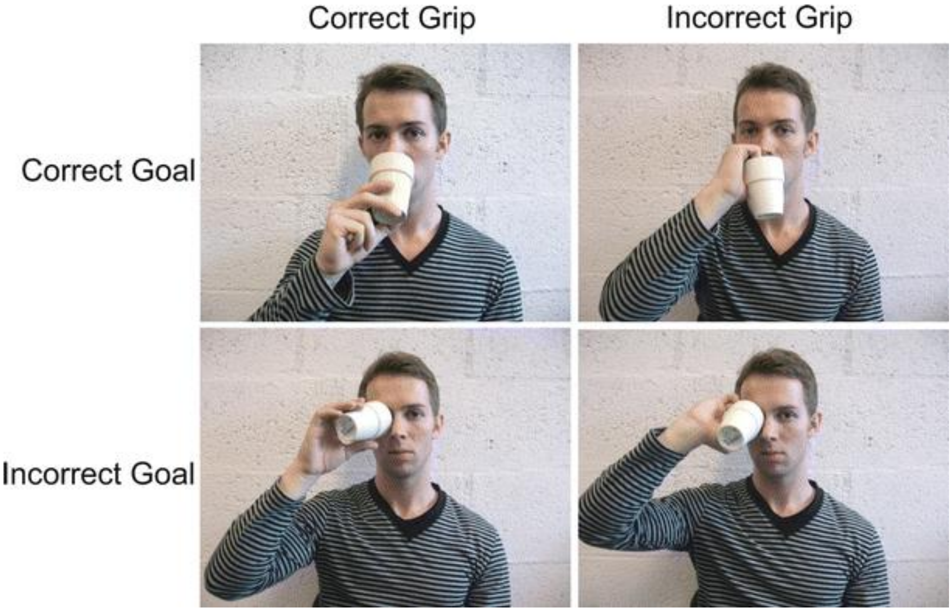


Figure 17. Example of the stimuli used by van Elk et al. (2008).

2.2. The choice of object-directed actions

The specificity of using object-directed actions is that goal-related information does not solely rely on the visual kinematics but is also spread upon the visual processing of the object, and its associated knowledge. It is then possible to evaluate the contribution of motor-related information (through visual kinematics) and the contribution of non-motor-related information (through visual processing of objects) during action recognition. Critically, the successful evaluation of object-directed actions depends on the knowledge we have about objects. The paradigm assumes that objects are associated with knowledge about how to use objects (i.e., manipulation knowledge) and for what purpose (functional knowledge). This assumption is well supported from research on object processing (see Bach et al., 2014; Bach & Schenke, 2017; Buxbaum, 2017; Osiurak & Badets, 2016; van Elk et al., 2014a for recent reviews). Bach et al. (2014) recently proposed that functional and manipulation knowledge about objects could mediate action recognition of object-directed actions. Functional knowledge could help the observer to process the actor’s goal whereas manipulation knowledge could help the observer to process the visual kinematics. Nevertheless, they highlighted that object knowledge is not sufficient to understand others' actions. An object can have different functions (e.g., a bottle can be used to pour water but also be used to drink directly); in this case, visual kinematics are necessary to understand the action. Similarly, manipulation knowledge can be shared between different objects (i.e., power grasp can be used to grasp a cell phone or a remote controller); in this case, the object function is necessary to understand the action. One may argue that because object processing involves motor-related information (through manipulation knowledge), the terminology "non-motor goal related information" to describe information derived from objects may be not suitable. Nonetheless in our design, we believe that the motor-related information related to the object is mixed with the visual kinematics. Indeed, the motor-related information associated with the object is necessary to recognise that the grip-configuration is appropriate or not. Overall, we reasonably assume that during the recognition of object-directed actions, the processing of visual kinematics and non-motor goal-related information are mediated both by functional knowledge and by manipulation knowledge associated with objects[6].

2.3. *Targeting the visual kinematic component: grip configuration*

The use of action photographs allows displaying grip and goal-related information at the very same time, but prevents from using of dynamic kinematic parameters. Grip configuration may not be as predictive of the outcome of the action as the full kinematics. Yet, grip configurations are an important part of the visual kinematics (Grafton & Hamilton, 2007). During the observation of others' object-directed actions, *f*MRI studies have found different brain regions associated with the processing of grip configuration (independently from hand trajectory) and with the processing of hand trajectory (independently from the grip

---

[6] We are indeed referring to the restricted case of familiar objects. It is clear that for novel objects, such knowledge is not available and action recognition may involve other/additional mechanisms.

configuration) (Grafton & Hamilton, 2007; Hamilton & Grafton, 2006, 2008; Southgate et al., 2014). Naish, Reader, Houston-Price, Bremner and Holmes (2013) further demonstrated that processing grip configuration was required for the observers to be able to discriminate two action goals (“to eat or to place”). Participants were asked to decide whether reach-to-grasp hand movements aimed at eating or placing an object. Despite the fact that the two action categories were characterised by early differences in terms of peak acceleration, participants succeeded to discriminate the two action goals only after the grip configuration was complete. Finally, participants tend to rely more on grip configuration than on hand trajectory when asked to judge whether two realistic object-directed action videos (e.g., reach to grasp action such as reaching and grasping a cup, or aimed dropping action such as dropping a piece of sugar in the cup) were identical or different (Loucks & Baldwin, 2009; Loucks & Pechey, 2016; Loucks & Sommerville, 2013). Overall, these data suggest that, although it may represent only one portion of the visual kinematics, grip configuration is an important component to discriminate and to understand different actions.

2.4. Targeting the non-motor goal-related dimension: visual information about the functional goal

As we have mentioned earlier, van Elk et al. (2014a, 2008) defined “*the goal of an action as the spatial location towards which an action is directed”* (2014, p.223). Action recognition have nonetheless been proposed to be *“based on conceptual expectations about the relationship between objects and the purpose of objects that are involved in the action (e.g., using objects in a purposeful way to have coffee) and more concrete perceptuomotor predictions of bodily movements directed to changing the location of a particular object (e.g., transporting a cup to the mouth)”* (Ondobaka & Bekkering, 2013, p. 2966). Thus, operationalising action goals as goal locations targets the perceptuomotor, but not the conceptual dimensions (i.e., functional goals) of action recognition. Yet goal locations and functional goals can be dissociated (Bach, Gunter, Friederici, Knoblich, & Prinz, 2005; Bach et al., 2009; Grafton & Hamilton, 2007; Hamilton & Grafton, 2006; Ondobaka et al., 2015; Southgate et al., 2014) and it may be important to distinguish between these two aspects of action goals. For example, Ondobaka et al. (2015) asked their participants to judge whether an object-directed action photograph matched with an action verb previously presented. Participants could either judge the functional relationship between the action verb and the object-directed action (e.g., *verb:* smelling and *action*: grasping a flower), or judge the spatial location relationship between the action verb and the object-directed action (e.g., *verb*: smelling and *action*: a flower near the nose). Action photographs could contain violations in terms of functional relationships (i.e., at the conceptual level; e.g., grasping a pencil near the nose) or in terms of their spatial location relationships (i.e., at the perceptuomotor level; e.g., grasping a rose near the ear). Brain regions sensitive to the functional relationships (i.e., posterior cingulate cortex) were different from brain regions sensitive to the spatial location relationships (i.e., inferior frontal gyrus and inferior parietal lobe). These results suggest that goal locations and functional goals may be sub-served by different brain networks, and more importantly that they jointly impact behavioural performance during action recognition. Thus, in our design, we attempted to target the functional goal of the object-directed action, and not the goal-location. Because goals are hidden in the actor’s mind, we needed to provide visual information about the functional goal. To do so, we varied the position of the object relative to the hand in such a way that the functional goal could be achieved or not. In our paradigm, we will refer to visual goal to speak about the visual information about the possibility to achieve the functional goal of the object-directed action.

2.5. Task instruction: overall correctness judgement about the action

Finally, most of the previous studies used tasks that require participants to explicitly process one dimension, while ignoring the other (Bach et al., 2005, 2009; Jacquet & Avenanti, 2015; Ondobaka et al., 2015; van Elk et al., 2012, 2008). Such tasks have generally demonstrated that the to-be-ignored dimension still influences the processing of the dimension participants had to judge. This strategy is particularly useful to speak about which action dimensions are automatically involved during action processing. Yet it does not allow evaluating at what moment of the recognition process which dimension is used. In addition, action processing seems to be particularly sensitive to task demands. For example, Flanagan, Rotman, Reichelt and Johansson (2013) found that, during observation of reach-to-grasp actions, participants' gaze behaviour was essentially composed of fixations on the object to be grasped when they had to predict which object will be grasped, whereas participants' gaze behaviour was essentially composed of saccades between the hand and the object to grasp when they had to evaluate the weight of the object. Processing the same object-directed action seems to involve very different strategies depending on the task. In the same vein, the involvement of the frontoparietal network (as indexed by the mu/beta-desynchronisation) has been found greatly reduced (Muthukumaraswamy & Singh, 2008; Perry, Troje, & Bentin, 2010; Schuch, Bayliss, Klein, & Tipper, 2010; Woodruff & Klein, 2013) when participants had to perform tasks irrelevant for the action they were looking at. Pobric and Hamilton (2006) also found that disrupting the activity of the inferior frontal gyrus impairs weight judgements of a hand lifting a box, but neither weight judgements of a bouncing ball (no hand) nor judgements of action irrelevant features (e.g., “How long was the hand visible”). These data suggest that even the neuronal mechanisms underlying action processing seem to differ as a function of task demands. For all of these reasons, we looked for a task that would not orient participants' attention specifically towards one of the two dimensions. Indeed, we may expect participants to process the dimension relevant to the task demands first, which would have prevented us to identify which dimension participants spontaneously favoured. In our protocol, participants had to judge whether the action photographs were displaying a typical use of the object ("*L'action correspond-elle à une utilisation typique de l'objet ?*"). In French, the verb "*use*" ("*utiliser"*) equally addresses both how to perform an action (i.e., related to the typical grip configuration) and for what purpose the action is performed (i.e., related to the functional goal of the action). Thus, our task remains agnostic in regard to which dimension participants may rely on when visually processing actions.

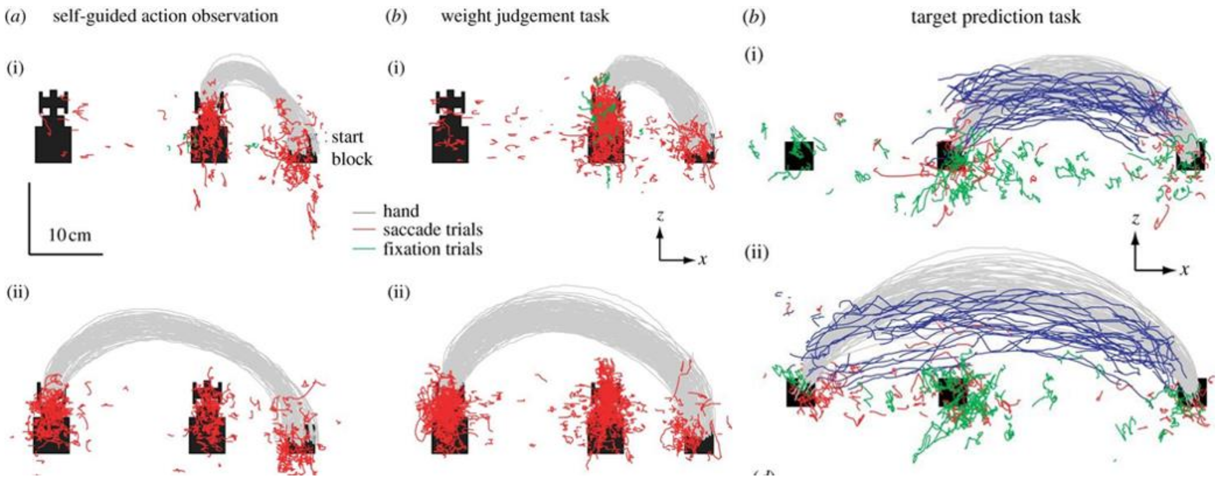


Figure 18. Pattern of gaze fixation as a function of the task. For judging about the weight of the object, participants essentially used saccade whereas to predict which object is about to be grasp, participants essentially tracked the hand movement and used fixation. From Flanagan et al. (2013).

2.6. *Stimuli*

Two sets of stimuli were designed. In each set, twenty objects were selected (see Appendix 4). There was only one main difference between the two sets: Set 1 was composed of photographs of the upper body of an actress (face included) performing hand-on-object actions (see Figure 19; Appendix 5); Set 2 was composed of photographs of a hand and an object only (no upper body; see Figure 20; Appendix 6). Actions were performed by the same right-handed actress for all photographs of the set 1. The right-handed actress for the set 2 was different, but also performed all the photographs of this set. All photographs always included only the tool-object, but never the recipient object on which the tool acts on (e.g., a nail for an action with a hammer). All information apart from the actress and the tool-object was eliminated so that object-directed actions were displayed in a context as neutral as possible. The twenty objects of each set were extracted from our everyday life environment, actions associated with these objects were then well-known from all participants. Visual and/or motor expertise could vary between participants, but as we will see below, our statistical strategy took into account such variation.

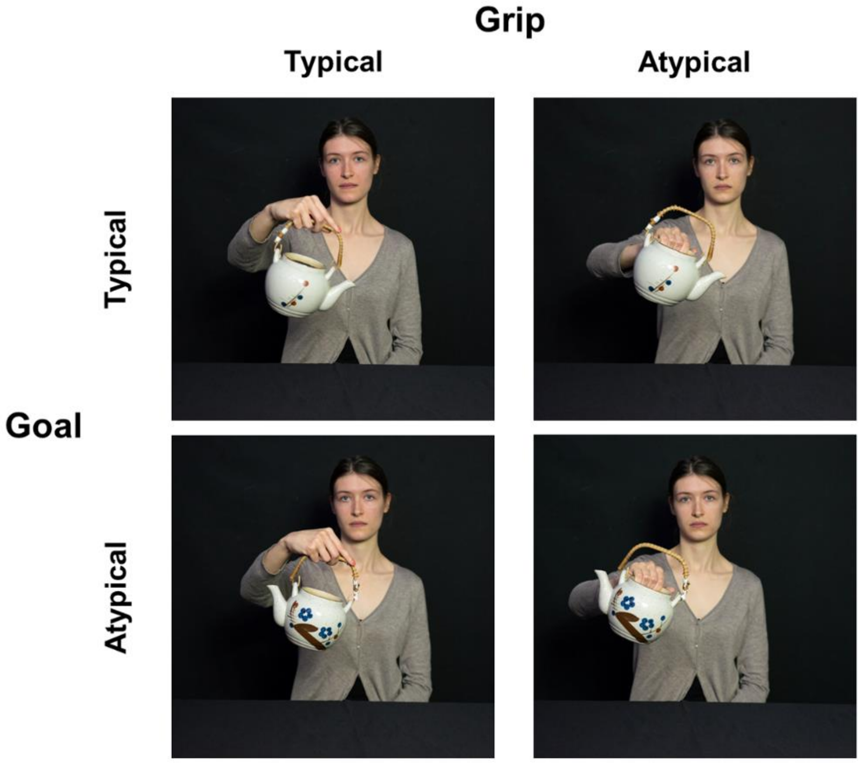


Figure 19. Example of the stimuli of Set 1.

For each reference object, actions could be typical or not along the grip dimension and/or the functional visual goal dimension. Grip configuration applied on the object could be typical or not according to the typical manipulation associated with the object. For instance, a precision grip applied to a pencil is typical, whereas a power grasp is not. The visual goal typically associated with the object could be achieved (goal-typical) or not (goal-atypical). The visual goal was mainly varied using different ways to orient the object. For example, a pencil in upright position allows writing (typical goal possible), whereas a pencil upside-down does not (typical goal impossible). Importantly, the typical goal could still be achieved even when the grip was atypical, and vice-versa. Thus, grip configuration and visual goal dimensions were manipulated independently.

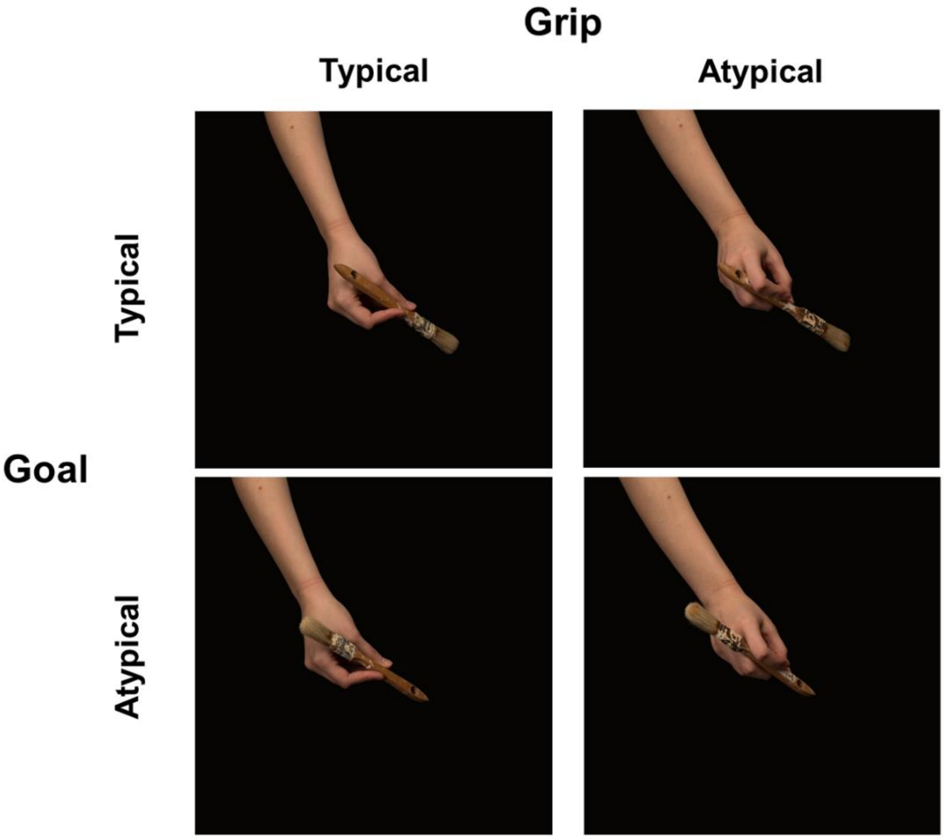


Figure 20. Example of the stimuli of Set 2.

The effect of grip and goal typicality on the perception of the overall typicality of the action photographs of the set 2 were verified in a pre-test. For each action photograph, nine participants were asked to determine whether the action was typical or not according to the typical use of the object. Participants were able to successfully classify typical (i.e. with both typical grip and goal) and typical (with either or both atypical grip and goal) photographs (M_{accuracy} = 90 % +/- 10 *SE*). A Chi-square test for independence indicated that performance was equally distributed between conditions $\chi 2(3) = 0.55$, $p = .907$. Thus, participants were able to recognise typical and–fully or partially–atypical action photographs, and there was no systematic bias towards one condition in particular.

3. Effects of interest

All along this thesis, the effects of interest were always the same. First, we were looking for the independent processing of the visual kinematics, then the independent processing of the non-motor goal-related information and finally we expected those two dimensions to be related to one another in some way. The visual kinematics were operationalised through the typicality of the grip configuration applied to the object. Thereby, statistical main effects of grip (typical versus atypical grips) were considered to reflect the processing of grip-configuration independently from non-motor goal-related information. Non-motor goal-related information were operationalised through the possibility to achieve the typical functional goal associated with the object in the way it is presented. Thus, statistical main effects of visual goal (typical versus atypical goals) were considered to reflect the processing of the non-motor goal-related information independently from the grip configuration. Finally, if grip configurations and visual goals relate to one another, we may expect the processing of grip configurations to be modulated as a function of the visual goal or vice-versa. Statistical interactions between grip and goal (congruent versus incongruent grip and goal dimensions) were considered to reflect the integration of the two action dimensions. The term "*integration*" here is just used to mean that at some point the two dimensions are related to one another. Such statistical interactions do not help to decide which of the two dimensions genuinely modulates the other.

4. Statistical strategy

All along this thesis, all the reported experiments will use repeated measure designs. In such designs, all participants go through all experimental conditions, which allow avoiding potential spurious results due to unbalanced samples of participants between the experimental conditions. Besides, it improves the signal-to-noise ratio and reduces the required sample size. One major problem with repeated-measure designs is that they produce non independent data. Data in condition A for participant Z is indeed non independent from data in condition B because it is produced by the same participant Z. In addition, we aimed to validate results with regard to the variability between participants, but also with regard to the variability between the stimuli used. Classical ANOVA strategies do not allow to account for both sources of variations simultaneously, and do not address the issue of data non-independence.

Mixed-effect linear models have been proposed as an alternative approach. These approach explicitly estimates the random effects, and allows to characterise several sources of random variation in a single analysis (Baayen, Davidson, & Bates, 2008; Barr, Levy, Scheepers, & Tily, 2013). The literature on mixed-effect linear models is still growing, and guidelines on how to specify them, how to use them, and how to evaluate them are somewhat blurred. Mixed-effect linear models are particularly challenging regarding (1) how to select the random structure, (2) how to evaluate the significance of the results and (3) how to compute and report effect size measures. In the following sections we will first introduce mixed-effect linear models, then discuss the three points raised just above. Parts of these sections are freely inspired from the paper of Brauer and Curtin (2018)–we also recommend this paper for a recent and complete introduction to mixed-effect linear models.

4.1. Mixed-effects model, a brief introduction

In mixed-effect linear models variables can be either “fixed” or “random”. Fixed variables roughly refer to the independent variables of an experiment. A variable is considered random *“when it has many possible levels and when the researchers’ interest is in all possible levels, but only a random sample of levels is included in the data*” (Brauer & Curtin, 2018, p. 4). As such, subjects and items will be considered as random variables in our experiments. In mixed-effect linear models, three main sources of variability are considered: first, as in any linear regression, the error term; second the random intercepts (both for participants and items), and third the random slopes (both for participants and items). In the error term lies all the variations unexplained by the fixed and random variables. The random intercept assumes that different levels of a random variable (participants and/or items) may account for the variation in the data: one subject may be overall faster than another one, or an item may be overall easier to process than another one. The random slope assumes that a random variable (participants and/or items) may be differently affected by the fixed variables. For example, participant 1 may be very efficient in processing grip configurations and will be highly sensitive to the difference between typical grips and atypical grips, whereas participant 2 will have poor abilities to process grip configurations and will never notice the difference between typical grips and atypical grips. Similarly, item 1 may elicit very different responses for typical versus atypical grips, whereas item 2 may elicit very similar responses for typical versus atypical grips. Therefore, it will be very easy to discriminate typical versus atypical grip for item 1 but very difficult to discriminate typical versus atypical grip for item 2. These random sources of variation are considered in the evaluation of the fixed effect (the main effect of grip in the example) and avoid the contamination of the fixed effects by the non-independence of the data. An example of mixed-model equation applicable to our design would be:

VD = (Intercept) + *(Fixed Main Effect Variable 1) + (Fixed Main Effect Variable 2) + (Fixed Variable 1 x Variable 2 Interaction Effect)* **+(Intercept + Main Effect Variable 1| Participants) + (Intercept | Items)** + (residual errors).

The italic part displays the fixed effects. The bold parts are the random effects. Here, Items have only a random intercept, whereas Participants have both a random intercept and a random slope for the main effect of the variable 1.

In the statistical software R, and with our labels, the same model would look like:

VD = *Grip-typicality* + *Goal-typicality* + *Grip-typicality:Goal-typicality* + **(Grip-typicality | Participants) + (1| Items)**

The intercepts are usually not displayed, unless they are the only parameters (e.g., for Items here). The error term is never explicitly specified. The formalism can be further simplified with the equivalent formulation:

VD = *Grip-typicality* * *Goal-typicality* + **(Grip-typicality | Participants) + (1| Items)**

The R formalism is provided here because it is easy to use and we will use this formalism hereafter.

4.2. Identifying the random structure

4.2.1. Theoretical principles

The specification of the fixed effects is inherited from the more classical linear regression and is quite easy to achieve. In contrast, the specification of the random effects has been debated (Barr et al., 2013; Bates, Kliegl, Vasishth, & Baayen, 2015; Brauer & Curtin, 2018; Matuschek, Kliegl, Vasishth, Baayen, & Bates, 2017). Barr et al. (2013) advocate for including "*the maximal random effects structure justified by the design*" (p. 255). Indeed, they argue that it is difficult to know in advance which effects will account for most of the variation in the data structure and choosing among the different possible sources of variation necessarily bias the results. This strategy allows one to be highly rigorous from a theoretical point of view. The full random structure for our case would look like in R formalism:

> VD = *Grip-typicality* + *Goal-typicality* + *Grip-typicality:Goal-typicality* + **(Grip-typicality + Goal-typicality + Grip-typicality:Goal-typicality | Participants)** + **(Grip-typicality + Goal-typicality + Grip-typicality:Goal-typicality | Items)**

In the above model, 13 critical parameters have to be estimated (the overall intercept, 3 fixed effects, 8 random parameters and the residual errors. Data sets in typical psychology experiments are often not able to support the resulting complexity of such models, which results in models that are not able to converge (i.e., the model does not find a reliable solution). Therefore, Bates et al. (2015) advocate for a more parsimonious approach in which we should select a random structure supported by the data. Matuschek et al. (2017) further demonstrate that keeping the random structure maximal, while lowering the type 1 error (i.e., the risk to conclude that an effect exists when it does not), also decreases the power (i.e., the ability to detect an effect if this effect does exist).

Taken into consideration the aforementioned elements, we selected all along the thesis the maximum random structure supported by the data. Our selection was based on the advices provided in the literature (Barr et al., 2013; Bates et al., 2015; Brauer & Curtin, 2018; Matuschek et al., 2017). Higher-order interactions in the random structure were first removed. Redundant factors (e.g. visible by a high correlation between the random intercept and the random slope for a given effect) were then removed from the random structure using principle component analysis of the random structure of the model. Then, we identified the random factors that were most represented on the components that contributed the less to the model. If those factors were not best represented on any other component, we removed them because they contribute poorly to the model (Bates et al., 2015). Random factors were removed one by one until the model converges. In case of hesitation on which factors to remove, we also considered the theoretical value of each factor of the random structure. For example, if we wonder whether we should remove grip similarity or response type, and our hypotheses are all oriented towards grip similarity, we will try to keep grip similarity and will first remove response type from the random factor structure.

4.2.2. An Example

Here is a practical example from our first experiment. We had four independent variables: grip similarity (GRIP), goal similarity (GOAL), response type (RESP) and prime duration (DURATION). As a dependent variable, we recorded the response times (RT). If we cross every participant, every item and the conditions of each independent variables, we have one data point. The fixed effect structure was:

RT = GRIP + GOAL + DURATION + RESP + GRIP:GOAL + GRIP: RESP + GRIP:DURATION + GOAL:RESP + GOAL:DURATION + RESP:DURATION + GRIP:GOAL:RESP + GRIP:GOAL:DURATION + GRIP:RESP:DURATION + GOAL: RESP:DURATION + GRIP:GOAL:RESP:DURATION

Or simplified:

RT = GRIP * GOAL * DURATION * RESP

The two structures are strictly equivalent, only the formalism is different. The first formalism allows us to visualise the complexity of even somewhat simple models. Here, just with the fixed effects structure, the model has 15 fixed effects + the intercept + the residual errors = 17 critical parameters to estimate. The full structure of the random factors would include:

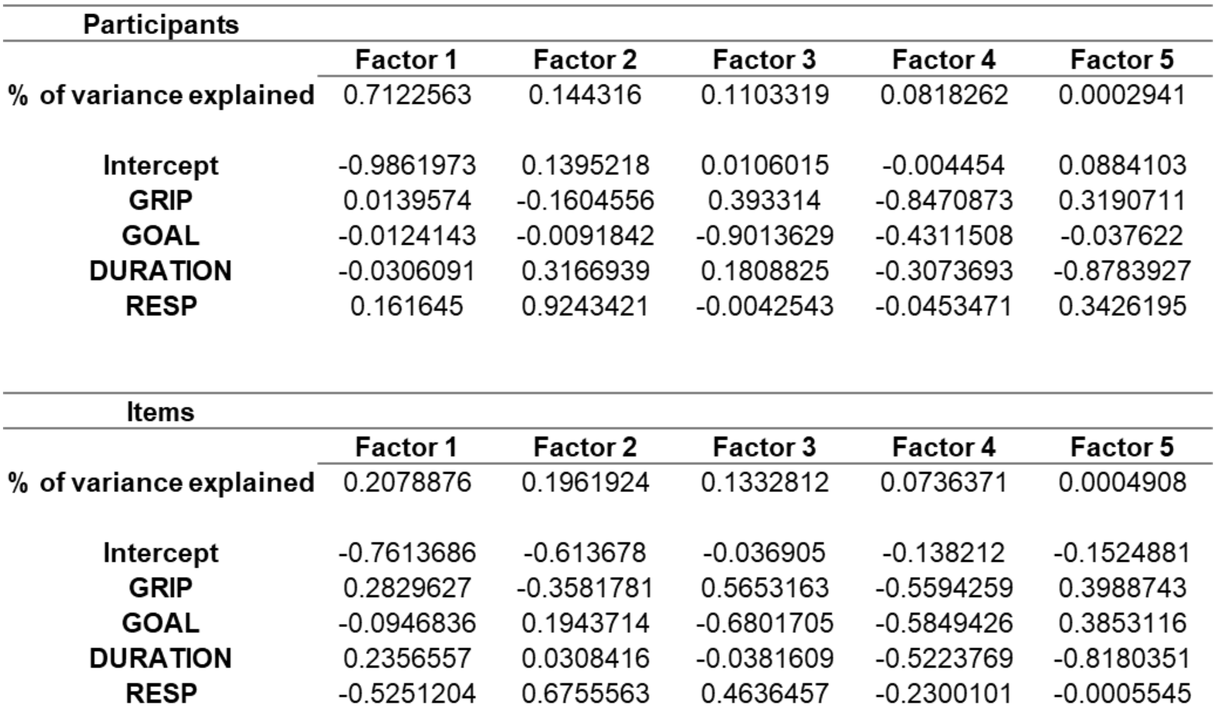
RT = GRIP * GOAL * DURATION * RESP + (GRIP * GOAL * DURATION * RESP |Participants) + (GRIP * GOAL * DURATION * RESP |Items)

In this model, we added the parameters of the random structure to the parameters of the fixed structure, that is 32 random parameters + 15 fixed effects + the intercept + the residual errors, so 49 parameters to estimate. Such model, unsurprisingly, did not converge (i.e., did not find a reliable solution). We first removed the higher order interaction from the random structure, note that the "*" will became "+" in R's formalism:

RT = GRIP * GOAL * DURATION * RESP + (GRIP + GOAL + DURATION + RESP |Participants) + (GRIP + GOAL + DURATION + RESP |Items)

In this structure, the model needs to estimate 27 parameters (fixed + random factors). Although the model was already deeply simplified, it still not converged. Thus, we ran the principal component analysis.

Table 1. Results of the principal component analysis.



Participants					
	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
% of variance explained	0.7122563	0.144316	0.1103319	0.0818262	0.0002941
Intercept	-0.9861973	0.1395218	0.0106015	-0.004454	0.0884103
GRIP	0.0139574	-0.1604556	0.393314	-0.8470873	0.3190711
GOAL	-0.0124143	-0.0091842	-0.9013629	-0.4311508	-0.037622
DURATION	-0.0306091	0.3166939	0.1808825	-0.3073693	-0.8783927
RESP	0.161645	0.9243421	-0.0042543	-0.0453471	0.3426195

Items					
	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
% of variance explained	0.2078876	0.1961924	0.1332812	0.0736371	0.0004908
Intercept	-0.7613686	-0.613678	-0.036905	-0.138212	-0.1524881
GRIP	0.2829627	-0.3581781	0.5653163	-0.5594259	0.3988743
GOAL	-0.0946836	0.1943714	-0.6801705	-0.5849426	0.3853116
DURATION	0.2356557	0.0308416	-0.0381609	-0.5223769	-0.8180351
RESP	-0.5251204	0.6755563	0.4636457	-0.2300101	-0.0005545

The above table gave us the results of the principal component analyses for the model with 27 parameters for the random structure of participants and items respectively. The first line represents the percentage of variance explained by each factor of the principal component analysis. For participants, Factor 5 explains the least variance in the model and thus have the least importance. In this column, we can identify DURATION with the best representativity on this 5th factor. Thus, we decided to remove the random slope DURATION from the random structure associated with the random factor participants. For items, DURATION was also identified as the best represented parameter on the least explicative factor of the principal component analysis. It was removed from the random structure associated with the random factor items. The model resulting from this procedure is:

RT = GRIP * GOAL * RESP * DURATION + (GRIP + GOAL + RESP |Participants) + (GRIP + GOAL + RESP |Items)

We keep applying this procedure until the model converged. This procedure led us to select quite conservative random structures in comparison to the one found in the literature.

4.3. Evaluating significance

Evaluating the significance of mixed-effect linear models may be challenging. As pointed out by Luke (2017), *"it is unclear whether the number of observations or the number of subjects and/or items or the number of grouping factors (i.e., the number of random effects), or some combination of these, would define the denominator degrees of freedom"* (p. 1494; see also Baayen et al., 2008). As a reminder, classical ANOVA uses *F* statistics to evaluate the significance of the statistical effects. Such statistics require the denominator degrees of freedom in their computation, which usually consider to the number of participants (in a by-subject analysis) or to the number of items (in a by-item analysis) weighted by the number of conditions in the design. In mixed-effect linear models, such denominator degrees of freedom are not easily available, which prevent the straightforward use of *F* statistics. One popular approach uses model comparison. In such approach, a full model is compared to another model with all but the fixed parameter we are interested in. If the difference between the full model and the reduced model is significant (as evaluated by the -2log likelihood ratio test, which does not require the computation of the denominator degrees of freedom), then we conclude that the removed variable has a significant effect. It is also possible to directly evaluate the significance of the parameter estimates of the model: dividing the estimate with its standard errors is used to obtain a Wald *t*-value, and significance of this *t*-value is evaluated using the *z* distribution. As the degrees of freedom increase, the *t* distribution can be approximated by the *z* distribution. Model comparisons and *t*-as-*z* approach are usually used together. Indeed, the *t*-as-*z* approach is limited in terms of interpretation, as parameter estimates only provide paired comparisons. Therefore, it is impossible to obtain the overall effect of a variable with more than two conditions.

An alternative approach implies to estimate the degrees of freedom of the denominator. This can be achieved either with the Kenward-Roger approximation (Kenward & Roger, 1997) or with the Satterthwaite approximation (Satterthwaite, 1941). The exact statistical justification and computational background supporting those two approximations will not be presented here. Using an extreme simplification, the Satterthwaite approximation makes some assumptions about the distribution of the model parameters. These assumptions are inadequate in the case of small sample size, and the Kenward-Roger approximation addresses these issues by providing a more general framework to estimate the denominator degrees of freedom. For this reason, the Kenward-Roger approximation is sometimes recommended over the Satterthwaite approximation (Brauer & Curtin, 2018). In practice, the two approximations provide very similar results, with slightly better performances of the Kenward-Roger approximation in case of small sample size (Brauer & Curtin, 2018; Kenward & Roger, 1997; Luke, 2017). In addition, the Kenward-Roger approximation is computationally heavier than the Satterthwaite approximation, and usually requires larger sets of data (Brauer & Curtin, 2018).

Recently, Luke (2017) found that approximations of denominator degrees of freedom produced acceptable type 1 error rates (as compared with the standard of alpha = 0.05 in Psychology), whereas model comparison and *t*-as-*z* approaches were somewhat anti-conservative (i.e., reject the null-hypothesis more often than they should). In addition, we believe that model comparison and *t*-as-*z* approaches may be more challenging to apply. The model comparison approach requires to fit several models and to find a suitable random structure that could be used across the different models of interest. In mixed-effects models, unbiased parameter estimates can be obtained using the Restricted Maximum Likelihood (REML) procedure. Model comparisons prevent the use of this procedure. In the *t*-as-*z* approach, it is not clear how large the data should be to justify its use (see Luke, 2017 for similar statement). As we have stated above, it is also not possible to evaluate the overall effect of a variable (as estimate fixed parameters only allow paired comparisons). In contrast, approximations of denominator degrees of freedom can be used along with the REML procedure and require fitting only one model (and thus the selection of only one random structure). With denominator degrees of freedom, overall main effects and interactions can be evaluated using an *F*-statistic. A priori comparisons can be directly specified in the model structure and evaluated using *t*-statistics on the estimate parameter of the model, and statistical interaction can be decomposed using classical post-hoc analyses such as the Tukey-procedure for paired comparisons.

In early work of this thesis, statistical significances of the mixed-effects models were first evaluated using model-comparison and *t*-as-*z* approach (see Decroix & Kalénine, 2018), but we quickly switched for the approximation of the degrees of freedom of the denominator. As we stated above, Kenward-Roger approximation is sometimes favoured, but provides overall similar results compared to the Satterthwaite approximation. The latter was computationally less demanding. Consequently, we decided to apply the Satterthwaite approximation.

4.4. Effect size measures

A last challenge concerns the computation of the sizes of the effects. The problem is similar to the one raised for the identification of the degrees of freedom of the denominator: there is currently no agreement on which sources of variation should be taken into account in the computation of the effect sizes. An additional difficulty was that until recently there was no proposition to compute effect sizes for mixed models at all. Some authors now argue that the computation of effect sizes should take both participants and items as sources of variations (Brysbaert & Stevens, 2018; Judd, Westfall, & Kenny, 2017; Westfall, Kenny, & Judd, 2014). In this approach, effect sizes on individual parameter estimates are computed using an adapted *d* of Cohen for mixed-effect models, hereafter "Westfall's *d*" (Brysbaert & Stevens, 2018; Judd et al., 2017; Westfall et al., 2014). Westfall's *d* is computed by dividing the difference of estimated means by the square root of the sums of the variance of the random factors. Following the current recommendations from the American Psychological Association, both unstandardised and standardised effect sizes should be reported (see also Pek & Flora, 2018). Unstandardised effect sizes simply correspond to the difference of means between two conditions of interest. They are expressed in their scale of origin and are easy to interpret (Pek & Flora, 2018), compared to the Westfall's *d*, considering the current state of the literature on those questions. In the following work, standardised effect sizes will be reported as the Westfall's *d*. It is not clear yet what should be the guideline to interpret the Westfall's *d*, so its report is mainly intended for future attempt to summarise experimental results of the field in meta-analyses. Overall effect sizes of *F* statistics are usually reported with partial eta-squared.

As far as we know, such measures are not available for mixed-effect linear models, thus we will only report effect sizes for our contrasts of interest. In early work of this thesis, effect sizes were reported using pseudo *R*-squared (see Decroix & Kalénine, 2018).

5. Summary of the methodological strategy

In sum, we aimed at evaluating the contribution of the visual kinematics and non-motor goal-related information during the processing of others' actions. Object-directed actions allow us to spread the goal-related information over both the visual kinematics (motor information) and the objects (non-motor information). The violation paradigm allows us to dissociate an important component of the visual kinematics (i.e., the grip configuration) from the object visual goal information through the way the object is used (e.g., upright or upside-down). The denomination "visual goal" was chosen to highlight that our design does not directly focus on the goal representation (on which visual kinematics can be informative too) but on the visual non-motor goal-related information. Analyses focus on the independent main effect of the grip configuration as a statistical marker of the contribution of the grip-configuration, the independent main effect of the goal-related information as a statistical marker of the contribution of the visual goal, and the Grip x Goal interaction as a statistical marker of the integration of the two action dimensions. Following the most recent statistical recommendations for repeated-measures designs, we used mixed-effect linear models. Among the different procedures to use and evaluate these models, we identified the ANOVA-like procedure (i.e., using *F*-statistics thanks to an approximation of the denominator degrees of freedom to evaluate the significance of our effects) as the most suitable procedure to analyse our data.

Experimental Contribution

CHAPTER 1: TIMING OF GRIP AND GOAL ACTIVATION DURING ACTION RECOGNITION: BEHAVIOURAL EXPERIMENTS

In the previous chapters, we established that sensorimotor and predictive approaches could be dissociated from the predictions they made on the place of visual kinematics. Sensorimotor approaches predict that action recognition should be driven by the processing of visual kinematics, whereas predictive approaches claim that predictions about the action goal are necessary to interpret the visual kinematics. We admitted in Chapter 1 that goals, by definition, are not directly observable and should be inferred from different sources of information. In Chapter 3, we established that visual kinematics could be used to derive goal information, and thus, in the general methodology we designed a paradigm to spread the source of goal information over both visual kinematics (i.e., the grip) and non-motor (i.e., the position of the object) sources of information. The two sources of information were manipulated orthogonally in order to address the two dimensions separately. Following predictions of sensorimotor approaches, participants should rely on visual kinematic information first, whereas predictive approaches predict that participants should rely on non-motor information first.

These predictions were evaluated using two different strategies. First, we used a priming procedure. Target action photographs could be primed by action photographs sharing the same grip and/or the same non-motor goal-related information. Hence, it was possible to selectively address one action visual dimension independently from the other. The duration of the prime was varied in order to evaluate the time required for grip and visual goal information to influence action recognition. In the second part, we considered this issue from a different point of view and evaluated how grip and visual goal information was used by the visuo-attentional system during a visual search task. We followed gaze behaviours as participants were looking for action photographs displaying a correct object-directed action. Distractors could share the same grip, or the same visual goal information as the target action. Thus, we evaluated the extent to which each distractor would capture visual attention during the search of the correct action target. Dynamic variation of visual attention on each distractor allowed us to evaluate at what moment each action dimension would be at play during action recognition.

PART 1: Priming Experiments

Most of this work was published in

***Decroix, J.**, & Kalénine, S. (2018). Timing of grip and goal activation during action perception: a priming study. Experimental Brain Research, 236(8), 2411–2426. https://doi.org/10.1007/s00221-018-5309-0*

The relative place of visual kinematics and non-motor goal-related information during the recognition of others' actions has been previously addressed by a few behavioural experiments (Kalénine et al., 2013; van Elk et al., 2008). Kalénine et al. (2013) found that healthy participants were slower to determine that two videos displayed different actions when the two actions differed only in terms of kinematic parameters (e.g. applying detergent with circular versus straight wipe), or in terms of goals (e.g., applying versus removing detergent), as compared to when the two actions differed in terms of both kinematic and goal. Van Elk et al. (2008) showed that response times to determine that the goal of the action displayed in a photograph was correct (i.e., "Is the object held at the correct goal location?" e.g., cup near the ear versus near the mouth) were slower when the grip information of the action was incorrect. Similarly, response times to decide whether the grip was correct (i.e., "Is the object grasped with an appropriate grip?" e.g., pencil grasped with power clench versus precision grip) were slower when the visual goal information was incorrect. Moreover, interference from incorrect goals was greater than interference from incorrect grips. Overall, both studies support the idea that visual kinematics and non-motor goal dimensions play distinct but complementary roles in the processing of observed actions. However, although they provide empirical evidence about the relative "weight" of the different action dimensions in decisions about perceived actions, they do not directly inform about the timing required to process grip and goal information during action recognition.

To address the question of the relative timing of grip and visual goal information during action recognition, a priming paradigm was employed. Priming paradigms have a long history in psychology and can be used in various ways. Roughly, priming is just a phenomenon in which a target event is affected by the previous presentation of another event. A priming effect is observed when the prior presentation of an event (hereafter the prime) influences the processing of the target event. Varying the relationship between the target and the prime, the time of presentation of the prime, or the delay between the prime and the target have been used to investigate various cognitive processes. The relationship between the target and the prime helps to address the nature of the information the researcher is interested in. The time of presentation of the prime informs about how much time is required for an information to be processed. The delay between the prime and the target informs about how much time the information remains in the cognitive system. As we were interested in the timing at which each of the visual kinematics and non-motor goal-related information influences the recognition of others' actions, we decided to fix the delay between the prime and the target. Instead, we varied the duration of the visual presentation of the prime. Varying the relationship between the target and the prime was used to evaluate one dimension independently from the other.

In the present series of experiments, we implemented the photographs we designed in a priming paradigm to evaluate the relative timing of grip and goal dimensions during action recognition. Target actions were judged correct or incorrect after being primed by action photographs sharing the same grip and/or the same goal. The duration of the prime varied with the idea that short prime durations should only allow the processing of the first steps of action decoding. Depending of which is processed first (visual kinematics for sensorimotor approaches; non-motor goal-related information for predictive approaches), selective grip and goal priming effects should be observed at short prime durations, with faster judgements for action photograph pairs sharing either the same grip or the same visual goal. At long prime durations, both grip and goal representations should be activated and should elicit action priming effects. In addition, photographs with partial grip violations (e.g., grip atypical but goal typical, see General Methodology section) or partial visual goal violations (e.g., grip typical but goal atypical, see General Methodology section) in which grip and visual goal information are incongruent should be more difficult to process. Consequently, when the two pieces of information are processed (i.e., for long prime duration), prime photographs with incongruent grip and visual goal information should interfere with the target action processing, leading to slower judgement than when target actions photographs are primed with photographs in which both pieces of information are congruent (i.e., both grip and goal typical or grip and goal atypical).

This study is divided in three experiments. The first two aimed at specifying the involvement of grip and visual goal information at different time points of visual processing. The third experiment aimed at replicating the main findings of these two experiments using different stimuli.

1. First Experiment: 66 and 300 ms prime durations

In the first experiment, we wanted to evaluate two extreme prime durations. EEG studies are among the few studies providing information about the temporal dynamics of action recognition. Thus, we designed the prime duration on the basis of EEG studies investigating the perceptual processing of object-directed actions. In these studies, brain activity of participants is recorded while they are asked to infer the goal of an observed action (Avanzini et al., 2013; Ortigue et al., 2010, 2009; van Elk et al., 2012). Note that the non-motor goal-related information is not dissociated from the visual kinematics, which prevent them to draw any clear-cut conclusion on which dimension is processed first. Overall, decoding others' action goals impacts brain activity in different time windows, the earliest starting around 60 ms (Ortigue et al., 2009) and the latest starting around 300 ms (Avanzini et al., 2013; Ortigue et al., 2009; van Elk et al., 2012) of visual action processing. Thus, 60 and 300 ms were chosen as prime duration boundaries.

1.1. Methodology

1.1.1. Participants

Thirty-one healthy participants took part in the study[7]. Three of them were left-handed or ambidextrous according to the Edinburgh handedness inventory (EHI; Oldfield, 1971) and were excluded. The final sample included 28 participants (M_{age} 22, age range 18-36, 23 females). All were right-handed (M_{EHI} = 87%, from 45% to 100%), reported normal or corrected-to-normal vision and provided written inform consent. They were not paid but could receive extra course credit for their participation. The protocol was approved by the Ethical Committee of the University of Lille and was in accordance with the declaration of Helsinki (1964, revised in 2013).

1.1.2. Stimuli

Set 1 of stimuli was used in this experiment. Briefly, for each of the twenty reference objects, four 1024*683 pixels coloured photographs were taken, all involving the same right-handed actress interacting with the object in one of four conditions. For each reference object,

---

[7] A sample size of about 30 participants was chosen to ensure sufficient statistical power (.80) for anticipated moderate effect sizes (Cohen d = .50 for the critical paired comparisons).

actions could be typical or not along the grip dimension and/or the goal dimension. Detailed description can be found in the General Methodology section. An example of the stimuli can be found in Figure 21.

Pictures were implemented in a priming paradigm. The four types of pictures could be presented as prime. Only the full correct actions showing both typical grip and typical visual goal (“correct targets”) or the full incorrect actions showing both atypical grip and atypical visual goal (“incorrect targets”) could be presented as targets. On these targets, participants had to judge whether the displayed action was correct (“yes response”) or incorrect (“no response”) according the typical use of the object. Overall there was four prime-target relations for each reference object: “grip similar, goal similar”; “grip similar, goal dissimilar”; “grip dissimilar, goal dissimilar”; “grip dissimilar, goal dissimilar”.

There was a total of 2 grip similarity (grip-similar; grip-dissimilar) x 2 goal similarity (goal-similar; goal-dissimilar) x 2 response type (yes = “correct target”; no = “incorrect target”) x 20 objects = 160 trials. The design of the different priming conditions is presented on Figure 21.

Prime duration was manipulated in two conditions, a short (66 ms) and a long (300 ms) prime duration. The 160 pairs of photographs were then presented twice, once with a prime duration of 66 ms, and once with a prime duration of 300 ms, leading to 320 experimental trials. Trials were divided into four blocks of 80 trials, each block containing an equal number of trials of the four prime-target pairs x two response types x two prime duration conditions.

Response Type	YES				
Prime-target pairs	Grip and goal similar	Grip similar only	Goal similar only	All different	Neutral **(Experiment 2)**
Prime					
Target					
Grip similarity Priming effect	**Shared**	**Shared**	**Different**	**Different**	**NA**
Goal similarity priming effect	**Shared**	**Different**	**Shared**	**Different**	**NA**
FSIM perceptual index mean(SD)	1	0,91 (+/-0,03)	0,87 (+/-0,04)	0,86 (+/-0,04)	0,82 (+/-0,03)
Response Type	NO				
Prime-target pairs	All different	Goal similar only	Grip similar only	Grip and goal similar	Neutral **(Experiment 2)**
Prime					
Target					
Grip similarity Priming effect	**Different**	**Different**	**Shared**	**Shared**	**NA**
Goal similarity priming effect	**Different**	**Shared**	**Different**	**Shared**	**NA**
FSIM perceptual index mean(SD)	0,86 (+/-0,04)	0,87 (+/-0,04)	0,91 (+/-0,03)	1	0,81 (+/-0,03)

Figure 21. Example of action photographs containing grip and/or goal violations for a given stimulus. In this example, the actress grasps the upright (full and closed) water bottle with a typical clench grip, a correct action regarding the typical function of the object (drink). In the photograph presenting a goal violation, the actress grasps the bottle with the same typical clench grip but the bottle is upside down, preventing her from drinking. In the photograph presenting a grip violation, the actress grasps the bottle with an atypical grip from below that does not prevent her from drinking. The photograph combining grip and goal violations displays the actress grasping the upside-down bottle with the atypical grip. Action photographs are then divided into eight prime-target pairs according grip similarity, goal similarity and response type. An additional neutral prime-target pair was included in Experiment 2. The FSIM index provides a measure of low-level perceptual similarity between prime and target.

1.1.3. Control measure of perceptual similarity

An objective index of perceptual similarity between prime and target was computed using the FSIM algorithm (Zhang, Zhang, Mou, & Zhang, 2011) to assess potential differences in terms of low-level visual features between prime-target pairs in the different conditions. According to this measure, pairs in the “grip-similar, goal-dissimilar” condition presented a higher degree of perceptual similarity than pairs in the “grip-dissimilar, goal-similar” and “grip-dissimilar, goal-dissimilar” conditions (p = .001), the two having the same degree of perceptual similarity (p = .09; see also Figure 21). Note that the perceptual similarity index in the “grip-similar, goal-similar” condition had no variance as the pair was composed of the same exact pictures. Perceptual similarity scores were taken into account in a complementary analysis (see below).

1.1.4. Procedure

Participants were seated at 100cm from the screen so that the action in the picture would appear within 5 degrees of visual angle. They were required to judge as fast and accurate as possible whether the target action photograph was correct or not according to the typical use of the object (forced choice). Instructions were displayed on the screen and carefully explained by the experimenter to ensure that participants understood the correct/incorrect distinction.[8] Twelve representative practice trials with feedback were provided with 3 objects that were not included in the actual experiment. The experimental session was the same as the practice session but without feedback. The experiment was conducted with E-Prime V2.0.10.353 software (Psychology Software Tools, Pittsburgh, PA).

---

[8] “On each trial, you will see two successive pictures showing an actor using an object. The first picture will always be briefly presented. You will have to judge the second photograph. You will have to determine, as fast and as accurate as possible, if the presented action is correct or not according to the typical use of the object. The use of an object is atypical when the object is used for another purpose or in another manner as the typical one. You will start with a training in which you will have feedback.”

Each trial began with a 2500 ms fixation cross, then a prime was displayed during 66 or 300 ms, followed by a 66 ms grey screen mask, and finally the target until participants' response. Participants answered “yes” (correct) or “no” (incorrect) using their left and right hands on left and right extreme keys of the response box. Response mapping (yes / no, left / right) was counterbalanced between participants. Response times (RT) and correct responses (accuracy) were recorded. The four blocks and the 80 experimental trials within each block were randomly presented. Short breaks were proposed between blocks. The experiment lasted ~1 hour.

1.1.5. Data Analysis

Correct RT were analysed as a function of grip similarity between prime and target (grip-similar, grip-dissimilar), goal similarity between prime and target (goal-similar, goal-dissimilar), prime duration (66 ms, 300 ms) and response type (yes, no); the four factors being repeated between participants and items. Response type was not a factor of interest but was introduced in the model since effect of yes / no response have been frequently reported. We expected the effects of grip and goal similarities to be modulated by prime duration.

As detailed in the General Methodology section, we were interested in the independent processing of grip (here statistically reflected by the grip similarity main effect), the independent processing of visual goal (here statistically reflected by the goal similarity main effect), and the integration of the two dimensions (here statistically reflected by the Grip similarity x Goal similarity interaction). As we were interested in the timing at which the two dimensions were activated, we looked for interactions between any of these effects and prime duration. Sensorimotor approaches predict an early access to grip over visual goal dimension, whereas predictive approaches predict the reverse pattern. The critical timing condition was then when primes were visually presented for 66 ms. Presence of an effect of grip similarity at 66 ms in absence of effect of goal similarity would favour the sensorimotor approaches. In contrast, presence of an effect of goal similarity at 66 ms in absence of effect of grip similarity would favour the predictive approaches. The Grip similarity x Goal similarity interaction was expected only when the two dimensions showed an independent activation.

1.2. Results

1.2.1. Data preparation for response times analysis

All analyses were carried out with R version 3.4.4 (R Core Team, 2019) and R Studio version 1.1.447. Overall, participants made few errors (M_{accuracy} = 2%, range 0.3% - 6%). Since the task was relatively easy to perform and errors were not of primary interest, errors were solely analysed in order to verify the absence of any trade-off effect. A chi-square test for independence indicated that errors were equally distributed between conditions $\chi 2(3) = .014$, $p = .99$. Errors and RT superior to 1500 ms and inferior to 150 ms were considered outliers and removed. Overall, 2.91 % of the data were removed. Final data are presented on Figure 22.

1.2.2. Mixed-effect model analysis of correct action recognition RT

As detailed in the General Methodology section, mixed effects linear models were used to analyse the RT. The model included grip similarity (GRIP), goal similarity (GOAL), prime duration (DURATION), response type (RESP) and related first-order, second-order, and third order interactions as fixed effects. The random structure was selected using the approach describe in the General Methodology section. The final random structure included random intercepts and random slopes for GRIP, GOAL, RESP and DURATION for both participants and items.

1.2.3. Effect of grip and goal similarity priming

The 4-way interaction between GRIP, GOAL, DURATION and RESP was not significant $F(1,9424.6) = 1.60$, $p = .206$. The effect of RESP will not be considered any further.

The GRIP x GOAL x DURATION interaction was significant *F*(1,9424.0) = 122.29, *p* < .001.Thus, the results were analysed separately for each prime duration.

The model fitted for 66 ms of prime duration included random intercepts for both participants and items. The analysis revealed no GRIP x GOAL interaction *F*(1,4758.2) = 0.02, *p* = .887, Westfall's *d* = 0.01, and no GRIP main effect *F*(1,4758.2) = 2.90, *p* = .089, Westfall's *d* = 0.04. The main effect of GOAL was significant *F*(1,4758.2) = 9.13, *p* = .002, Westfall's *d* = 0.07, and was explained by the fact that goal-similar trials yielded faster response times than goal-dissimilar trials (goal-dissimilar minus goal-similar = 12.65 ms, *SE* = 4.19).

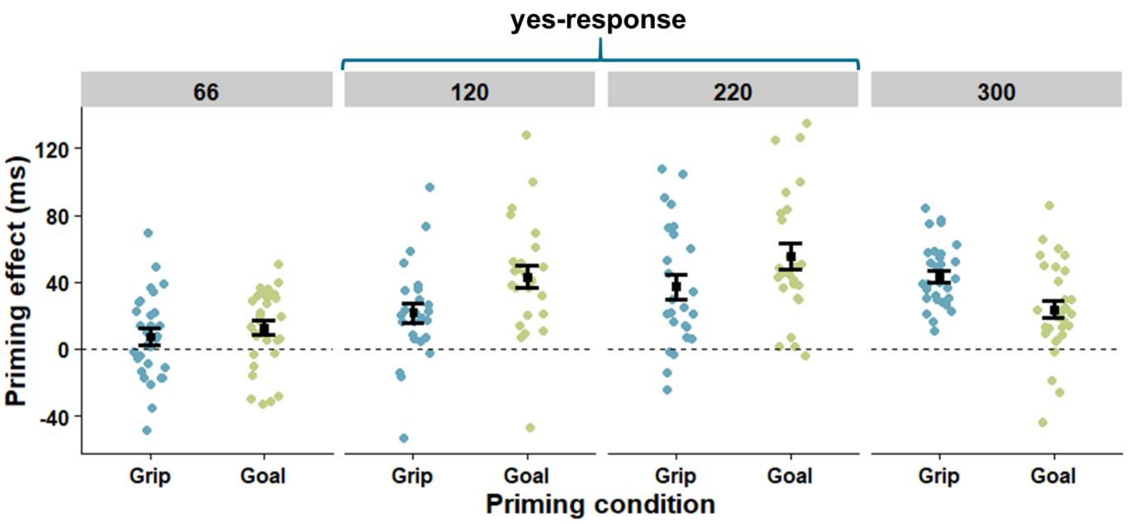


Figure 22. Individual mean priming effects in milliseconds according to grip, goal and prime duration. Prime durations of 66 and 300 ms were tested in Experiment 1 (both yes and no response here), and prime durations of 120 and 220 ms were evaluated in Experiment 2 (only yes-response displayed). Small dots represent different individuals. Black dots represent the mean priming effects. Error bars represent standard error.

The model fitted for 300 ms of prime duration included random intercept for participants and GOAL as a random slope for items. The analysis revealed a significant GRIP x GOAL interaction *F*(1,4769.1) = 243.45, *p* < .001, Westfall's *d* = 0.71, a significant main effect of GRIP *F*(1,4769.1) = 118.00, *p* < .001, Westfall's *d* = 0.25, and a significant main effect of GOAL *F*(1,4769.2) = 34.92, *p* < .001, Westfall's *d* = 0.13. In both cases, the similar dimension yielded faster response times than the dissimilar dimension (goal-dissimilar minus goal-similar

= 23.52 ms, *SE* = 3.97; grip-dissimilar minus grip-similar = 43.15 ms, *SE* = 3.97). In contrast, the Grip similarity x Goal similarity interaction was driven by the fact that primes involving incongruent grip and goal dimensions (Grip-similar Goal-dissimilar; Grip-dissimilar Goal-similar) yielded slower response times than primes involving congruent grip and goal dimension (Grip-similar Goal-similar; Grip-dissimilar Goal-dissimilar; estimate of the interaction = -123.96 ms, *SE* = 7.95).

1.2.4. Complementary analysis: influence of perceptual similarity

Spearman's rank correlations were computed between the goal and grip similarity priming effects obtained in each response condition and the corresponding perceptual similarity indices. There were no significant correlations (all $p > .66$, see Appendix 7).

1.3. Interim discussion

In this first experiment, we sought to provide some information about the relative timing at which grip and goal dimensions would be activated when visually processing object-directed actions. We used a priming paradigm in which target action photographs were primed by action photographs sharing either both grip and visual goal dimensions, only the grip dimension, only the visual goal dimension, or neither of the two dimensions. Importantly, the object involved in the target and in the prime action photographs was always the same, and the same actress was always involved. Thus, any effects are necessarily due to one of the two action-relevant dimensions. Information about the timing was investigated through the variation of the prime duration. Our results demonstrated that priming an action target with the same goal for 66 ms of visual processing facilitates action recognition of the target. Priming an action target with the same grip induces a facilitation, but only for 300 ms of visual action prime processing. Consequently, the visual goal dimension influences action processing before the grip dimension. In addition, judgements on target action were slower when action primes contained incongruent grip and visual goal dimension–primes with partial grip or goal violation–than when they did not–primes with both grip and visual goal dimensions typical or both atypical. As predicted, when both dimensions could influence action processing (i.e., 300 ms of prime visual processing), an interaction emerged.

Priming effects for the 300 ms prime duration are in line with previous experiments (Kalénine et al., 2013; van Elk et al., 2008) with the co-activation of both grip and visual goal dimensions during action decoding. Here, as long as participants had sufficient time to process the prime (300 ms), both grip and visual goal information influenced the target action processing. Interestingly, the magnitude of grip similarity priming was greater than the one of goal similarity priming. In our study, participants did not have to explicitly process the different grip and visual goal dimensions manipulated in the primes. Whereas grip information may be more difficult to process when judged explicitly (cf. Van Elk et al., 2008), it may strongly impact action processing when implicitly processed. In addition, our priming paradigm highlights a cost on target processing when primes contained incongruent grip and goal information (partial action violations). This result mirrors the slower action decisions on action stimuli containing partial violations reported in the literature (Kalénine et al., 2013; van Elk et al., 2008). However, contrasting with previous studies, participants were not explicitly informed about the action dimensions manipulated. Thus, the cost entailed by partial action violations indicate that once activated, grip and goal representations incidentally interact with one another during action visual processing. This extends the claim of the involvement of distinct but interrelated kinematic parameters and goal representations in action perception.

When primes were presented for 66 ms, only goal similarity priming was facilitating action judgements. The absence of early grip priming effects may be interpreted in the light of previous studies founding that sensorimotor simulation of specific kinematic parameters performed by others may not arise before 200 ms (Naish, Houston-Price, Bremner, & Holmes, 2014). Although modulations of corticospinal excitability by observed actions have been found as early as 90 ms (e.g., Lepage et al. 2010; Cavallo et al. 2014 see Naish et al., 2014 for review), these early modulations would only correspond to a general pre-activation of the motor system. Only later corticospinal excitability modulations (after 200 ms) would reflect the mapping of the specific kinematic parameters perceived onto the observer's own motor system. Therefore, even though the grip may have been processed during the first 66 ms of prime processing, this processing may have not been sufficient to influence target action judgements.

Regarding the activation of visual goal information, early activation has been previously reported. In particular, EEG studies showed that discriminating between the different possible action goals of an observed action starts influencing brain activity after 60 ms of visual processing (e.g., Ortigue, Thompson, Parasuraman, & Grafton, 2009). Yet these early responses have often been associated to low-level perceptual processes (Catmur, 2015) rather than early activation of goal-related information. Then, earlier and stronger facilitative priming effects may be expected between pairs of action pictures that are more perceptually similar. However, this possibility was ruled out using the FSIM measure of low-level perceptual similarity. Grip-similar pairs were more perceptually similar than goal-similar pairs, and there was no correlation between the amplitude of grip and goal similarity priming effects and perceptual similarity differences between prime-target pairs. Thus, it is unlikely that the early facilitative goal priming effect merely reflects the greater perceptual similarity between action pairs in this condition. In contrast, goal priming results are consistent with an early activation of goal representations. Critically, goal activation was not only visible early but first. Hence, the relative timing of goal and grip activation supports a top-down processing of observed actions in our task. This suggests that participants may first activate a prediction about the possible action goal of the actor that will guide the subsequent processing of the kinematic parameters of the action, which occurred between 60 and 300 ms.

2. Second Experiment: 120 and 220 ms prime durations

In the second experiment, we wanted to further investigate the time-course of grip and visual goal processing. Specifically, we were looking for the timing at which both grip and visual goal dimension started to influence the recognition of others' actions. Intermediate prime durations between 66 and 300 ms were selected. The first prime duration was set up at 120 ms of visual presentation. This timing corresponds to the late boundary of first time window of action understanding identified at the neurophysiological level (e.g., Ortigue et al., 2009). At such timing facilitative effects of goal similarity were expected, which would confirm the earliness of goal activation. The second prime duration was set up at 220 ms in order to fit the timing at which specific motor responses have been reported during action observation (Naish et al., 2014). Effects of grip similarity were expected to emerge between 120 and 220 ms. Experiment 2 followed the same rationale and paradigm as Experiment 1, asides from two additions:

- The strength of the perceptual mask used between prime and target was enhanced for a greater control of low-level perceptual priming effects.
- A neutral priming condition presenting the object without action was added to further evaluate the role of object information in priming effects involving object-related actions.

2.1. Methodology

2.1.1. Participants

Thirty healthy participants took part in the study. Five of them were classified as left-handed or ambidextrous according to the EHI (Oldfield, 1971), and were excluded from the study. The final sample included 25 participants (M_{age} 21, age range 18-30, 16 females). All were right-handed (M_{EHI} = 84 %, from 43 % to 100 %), reported normal or corrected-to-normal vision and provided written inform consent. They were not paid but could receive extra course credit for their participation. The protocol was approved by the Ethical Committee of the University of Lille and was in accordance with the declaration of Helsinki (1964, revised in 2013).

2.1.2. Stimuli

Experiment 2 used the same prime-target action pairs as Experiment 1, divided in 4 conditions: “grip-similar, goal-similar”; “grip-similar, goal-dissimilar”; “grip-dissimilar, goal-dissimilar”; “grip-dissimilar, goal-dissimilar”. A control condition was added in which the target was coupled with a neutral action-free prime in which the object was displayed in front of the actress with her hands resting on the table (Neutral). This control condition was designed to evaluate to what extent mere object priming could account for the expected effects of goal and/or grip similarity priming (see complementary analysis below). Overall, there was a total of 5 prime-target pairs x 2 response types x 20 objects = 200 prime-target pairs (see Figure 21). Trials were divided into 4 blocks of 100 trials, each block containing an equal number of trials of the 5 prime-target pairs x 2 response types x 2 prime duration conditions.

2.1.3. Procedure

Experiment 2 evaluated two intermediate prime durations of 120 ms and 220 ms. A few additional changes were applied to the procedure. The grey mask used in the first experiment was replaced by a black and white pixelated mask and the target picture was spatially shifted from the prime picture by 0.5 degree of visual angle. These subtle changes have been made to the procedure in order to provide a stronger control of low-level visual effects such as visual persistence and low-level spatial overlap. Apart from those modifications, the procedure and the rationale of Experiment 2 were the same as Experiment 1. The four blocks and the 100 experimental trials within each block were randomly presented. Short breaks were proposed between blocks. The experiment lasted ~1 hour.

2.1.4. Data analysis

As in Experiment 1, RTs were analysed as a function of grip similarity between prime and target (grip similar, grip dissimilar), goal similarity between prime and target (goal similar, goal dissimilar), prime duration (120 ms, 220 ms) and response type (yes, no); the four factors being repeated between participants and items.

Following the results obtained in the first experiment, we did not expect the effect of goal similarity to be modulated by prime duration, as it was expected at both 120 and 220 ms. The critical effect tested in this experiment was the timing at which the effect of grip similarity would be observed. An emergence of grip similarity effect at both 120 and 220 ms would be reflected by a main effect of grip similarity while an emergence of grip similarity effects at 220 ms only would be reflected by an interaction between grip similarity and prime duration. As previously, the grip similarity x goal similarity interaction was expected as soon as both dimensions would be activated.

2.2. Results

2.2.1. Data preparation for response times analysis

Again, error rate was relatively low (M_{accuracy} = 3%, from 0.4% to 12%). A chi-square test for independence revealed that errors were equally distributed between conditions $\chi 2(4)$ =

.0126, *p* > .99. Errors and RT superior to 1500 ms and inferior to 150 ms were considered outliers and removed. Overall, 3.19 % of the data were removed. The final data are reported on Figure 22.

2.2.2. Mixed-effect model analysis of correct action recognition RT

As detailed in the General Methodology section, mixed effects linear models were used to analyse the RT. The model included grip similarity (GRIP), goal similarity (GOAL), prime duration (DURATION), response type (RESP) and related first-order, second-order, and third order interactions as fixed effects. The random structure was selected using the approach described in the General Methodology section. The final random structure included random intercepts and random slopes for RESP for both participants and items.

2.2.3. Effect of grip and goal similarity priming

The 4-way interaction between GRIP, GOAL, DURATION and RESP approached significance *F*(1,7635.2) = 3.78, *p* = .052. Thereby, the remaining analysis was conducted separately for yes and no responses.

For yes responses, the model included both random intercept and random slopes for GRIP, GOAL and DURATION for both participants and items. The GRIP x GOAL x DURATION interaction was not significant *F*(1,3743.5) = 1.24, *p* = .266, Westfall's *d* = 0.12. The analysis revealed a significant GRIP x GOAL interaction *F*(1,3744.9) = 31.82, *p* < .001, Westfall's *d* = 0.30, a significant main effect of GRIP *F*(1,19.9) = 28.61, *p* < .001, Westfall's *d* = 0.18, and a significant main effect of GOAL *F*(1,28.3) = 75.63, *p* < .001, Westfall's *d* = 0.30. In both cases, the similar dimension yielded faster response times than the dissimilar dimension (Goal-dissimilar minus Goal-similar = 48.96 ms, *SE* = 5.63; Grip-dissimilar minus Grip-similar = 29.58 ms, *SE* = 5.53). In contrast, the GRIP x GOAL interaction was driven by the fact that priming situations in which grip and goal dimension were not congruent (Grip-similar Goal-dissimilar; Grip-dissimilar Goal-similar) yielded slower response times than priming situations in which grip and goal dimension were congruent (Grip-similar Goal-similar; Grip-dissimilar Goal-dissimilar; estimate of the interaction = -48.68 ms, *SE* = 8.63).

For no responses, the model included random intercepts and random slopes for GOAL and DURATION for participants, and random intercepts and random slopes for GRIP, GOAL and DURATION for items. The analysis revealed a significant GRIP x GOAL x DURATION interaction *F*(1,3767.6) = 15.89, *p* < .001, Westfall's *d* = 0.42. The remaining analysis was then conducted separately for 120 and 220 ms.

The model analysing the data at 120 ms for no responses included random intercepts for participants, and random intercepts and random slopes for GRIP for items. For 120 ms of visual processing, only the GRIP x GOAL interaction was significant *F*(1,1880.4) = 6.19, *p* = .013, Westfall's *d* = 0.19. As previously observed, priming situations in which grip and goal dimensions were not congruent (Grip-similar Goal-dissimilar; Grip-dissimilar Goal-similar) yielded slower response times than priming situations in which grip and goal dimension were congruent (Grip-similar Goal-similar; Grip-dissimilar Goal-dissimilar; estimate of the interaction = -30.44 ms, *SE* = 12.23).

The model analysing the data at 220 ms for no responses included both random intercepts and random slopes for GRIP and GOAL for both participants and items. The analysis revealed a significant GRIP x GOAL interaction *F*(1,1833.89) = 23.55, *p* < .001, Westfall's *d* = 0.36, a significant main effect of GRIP *F*(1,24.05) = 23.41, *p* < .001, Westfall's *d* = 0.23, and a significant main effect of GOAL *F*(1,24.41) = 45.64, *p* < .001, Westfall's *d* = 0.34. In both cases, the similar dimension yielded faster response times than the dissimilar dimension (Goal-dissimilar minus Goal-similar = 54.89 ms, SE = 8.13; Grip-dissimilar minus Grip-similar = 37.21 ms, *SE* = 7.69). In contrast, the GRIP x GOAL interaction was driven by the fact that priming situations in which grip and goal dimension were not congruent (Grip-similar Goal-dissimilar; Grip-dissimilar Goal-similar) yielded slower response times than priming situations in which grip and goal dimension were congruent (Grip-similar Goal-similar; Grip-dissimilar Goal-dissimilar; estimate of the interaction = -58.45 ms, *SE* = 12.04).

2.2.4. Complementary analysis: Effect of object priming

If participants only rely on object identity to derive information about the goal or grip of the action, then the pattern of goal and grip similarity priming effects should follow the pattern of object priming effects. To evaluate this hypothesis, we replaced the conditions in which grip overlapped between prime and target with the neutral priming condition to compute grip similarity effects and we replaced the conditions in which goal overlapped between prime and target with the neutral priming condition to compute goal similarity effects. In other words, neutral prime-target pairs (object similar) were compared, first, with prime-target pairs that differed on grip dimension (grip dissimilar conditions) and subsequently, with prime-target pairs that differed on goal-related dimension (goal dissimilar conditions, see Table 1). Those comparisons were tested in situations where we observed an effect of grip similarity and goal similarity in the main analysis of Experiment 2.

For YES responses, object similarity did not show any advantage over grip dissimilarity $F(1,3828.2) = 1.46$, $p = .227$, Westfall's $d = 0.03$. This stands in contrast with the grip similarity priming effect observed in the main analysis and indicates that object repetition is not responsible for the grip similarity priming effect. Object similarity did show a 15 ms advantage over goal dissimilarity $F(1,3826.2) = 11.93$, $p < .001$, Westfall's $p = 0.09$; Goal dissimilar minus Neutral = 15.23, *SE* = 4.41. This suggests that object similarity priming could at least partially account for the effect of goal similarity in the main analysis, though only partially as the object similarity effect is 3 times less important than the goal similarity effect (Goal-dissimilar minus goal-similar = 48.96 ms, *SE* = 5.63).

For NO responses, the dissimilar grip condition (for 120 ms : *F*(1,1902) = 13.17, *p* < .001, Westfall's *d* = 0.13; Grip dissimilar minus Neutral = -20.02, *SE* = 5.52; for 220 ms : *F*(1,1879.9) = 26.48, *p* < .001, Westfall's *d* = 0.19; Grip dissimilar minus Neutral = -29.4, *SE* = 5.71) and the dissimilar goal condition (for 120 ms *F*(1,1903) = 18.84, *p* < .001, Westfall's *d* = 0.16; Goal dissimilar minus Neutral = -22.77, *SE* = 5.25; for 220 ms: *F*(1,1883.9) = 33.05, *p* < .001, Westfall's *d* = 0.19; Goal dissimilar minus Neutral = -33.29, *SE* = 5.79) were systematically more facilitative than the object similar condition. Facilitative effects of grip or goal similarity for no responses were not due to object similarity priming.

2.2.5. Complementary analysis: influence of perceptual similarity

As in Experiment 1, Spearman's rank correlations were computed between the goal and grip similarity priming effects obtained in each response condition and the corresponding perceptual similarity indices. There were no significant correlations (all *p* > .29, see Appendix 7).

2.3. Interim discussion

In this second experiment, we sought to investigate intermediate timing between 66 and 300 ms, using the same priming paradigm as in Experiment 1. The availability of the prime for visual processing was set up at 120 ms to reflect the latest boundary of the first time-window identified in EEG experiments, and at 220 ms to reflect the timing at which specific motor responses were reported. Results demonstrated the impact of grip and visual goal dimensions on action recognition from 120 ms of action visual processing. Sharing the same visual goal or the same grip as the target facilitated the recognition of the action. The facilitation observed at 120 ms of visual processing was especially detected for yes responses with an effect more important for goal similarity than for grip similarity. The facilitation was also observed for no responses, but slightly delayed as its emergence required 220 ms of visual processing.

The effect of response type in Experiment 2, although not predicted, is not at odds with the literature on action processing (e.g., Yoon, Humphreys, & Riddoch, 2010). van Elk et al. (2008) also found an effect of response type, with an impact of action violations on action decisions only when the dimension of the action evaluated (correctness of grip or correctness of goal) was correct. This may suggest that regardless of the action dimension considered, correct representations have a greater weight than incorrect representations in the comprehension of observed actions. In our experiment, goal and grip similarity priming effects were not completely absent for no responses: the emergence of both goal and grip priming effects was slightly delayed and became visible from 220 ms of prime duration which contrast with priming effects for yes responses evidenced regardless of prime duration from 120 ms. Although a better understanding of the reasons underlying this delay is needed, the effect of response type on action priming effects does not impact the main result of Experiment 2, namely the appearance of grip similarity priming effect from 120 ms of prime processing, and the consistency of the whole data pattern between Experiments 1 and 2. At extremely short and long prime durations (Experiment 1), the effect of response type was not significant since at 66 ms prime duration was far too short to trigger grip similarity priming and at 300 ms grip similarity priming was already well established. Response type affected the prime durations at which grip similarity priming became obvious, between 120 and 220 ms (Experiment 2).

The main finding of Experiment 2, particularly clearly highlighted for yes responses, shows both grip and goal similarity priming effects from 120 ms of prime processing, with a greater magnitude of goal priming effects. In addition, as in Experiment 1, a cost for primes conveying incongruent grip and goal information (i.e., primes with partial violations) was observed from 120 ms regardless of response type, as reflected by significant interactions between grip and goal similarity. Since both grip and goal representations have to be activated in order to induce a cost when incongruent, this effect reinforces the idea that grip similarity priming emerges as early as 120 ms. Again, we can be fairly confident that differences in low-level perceptual similarity between prime-target pairs cannot account for the pattern of priming effects. The stronger masking procedure used in Experiment 2 minimised the possible influence of low-level similarity priming. More importantly, there was no correlation between the amplitude of grip/goal priming effects and perceptual similarity indices in the different conditions (Appendix 7).

Finally, the complementary analysis involving the neutral prime condition indicated that prime-target pairs sharing mere object similarity showed an advantage over prime-target pairs with dissimilar goals, but not dissimilar grips (in the yes response condition). This suggests that the advantage of prime-target pairs with similar goals over prime-target pairs with dissimilar goals (i.e., the goal priming effect in the main analysis) could be due to mere object priming. However, the magnitude of the goal priming effect (49 ms) was three times as big as the object priming effect (15 ms), indicating that object priming may only partially account for the goal similarity priming effects observed for yes responses. One interpretation of this additional result is that observers may use some information about object identity to activate a prediction about the goal of the action. We will go back later to this interpretation.

3. Experiment 3: Replication at 66 ms and 220 ms prime durations

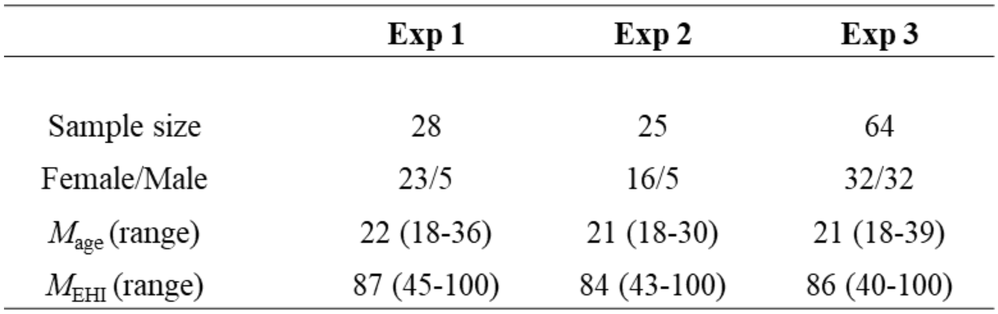
In the third experiment, we tried to replicate the results obtained in our previous experiments using a new set of stimuli, and with more participants. To keep the experimental context as similar as Experiment 1 and 2, only two prime durations were evaluated. We selected one prime duration in each of the two previous experiments. First, we selected 66 ms because it is the earliest time point, and as it is the timing at which the results could most likely vary, a replication will reinforce the previous results. Then, we selected 220 ms from the second experiment, because at this time point, the two dimensions clearly influenced the recognition of others' actions for both yes and no responses. Finally, to ensure that our results were reliable regardless of gender–we had more females than males in the previous samples –we recruited an equal number of female and male.

3.1. Methodology

3.1.1. Participants

Sixty-four healthy participants took part in the study (M_{age} 21, age range 18-39, 32 females). All were right-handed (M_{EHI} = 86 %, from 40 % to 100 %), reported normal or corrected-to-normal vision. Excepting the female/male balance, the mean and range of age and EHI were very similar to Experiment 1 and 2 (see Table 2). They provided written inform consent and received 10 euros for their participation. The protocol was approved by the Ethical Committee of the University of Lille and was in accordance with the declaration of Helsinki (1964, revised in 2013).

Table 2. Demographic data for Experiment 1, 2 and 3.



	Exp 1	**Exp 2**	**Exp 3**
Sample size	28	25	64
Female/Male	23/5	16/5	32/32
M_{age} (range)	22 (18-36)	21 (18-30)	21 (18-39)
M_{EHI} (range)	87 (45-100)	84 (43-100)	86 (40-100)

3.1.2. Stimuli

In Experiment 3, the second set of stimuli was used. As a remainder, this second set presented both the right hand of an actress and an object, without any other information (see General methodology section). As in Experiment 2, the prime-target action pairs included 5 conditions: "grip-similar, goal-similar"; "grip-similar, goal-dissimilar"; "grip-dissimilar, goal-dissimilar"; "grip-dissimilar, goal-dissimilar"; "Neutral". As in Experiment 2, there was a total of 5 prime-target pairs x 2 response types x 20 objects = 200 prime-target pairs. Trials were divided into 4 blocks of 100 trials, each block containing an equal number of trials of the 5 prime-target pairs x 2 response types x 2 prime duration conditions.

3.1.3. Procedure

Experiment 3 aimed at replicating the results obtained with 66 and 220 ms of visual presentation of the prime. The procedure and the rationale were the same as Experiment 2. The 4 blocks and the 100 experimental trials within each block were randomly presented. Short breaks were proposed between blocks. The experiment lasted ~1 hour.

3.1.4. Data analysis

As in Experiment 1 and 2, RTs were analysed as a function of grip similarity between prime and target (grip-similar, grip-dissimilar), goal similarity between prime and target (goal-similar, goal-dissimilar), prime duration (66 ms, 220 ms) and response type (yes, no); the four factors being repeated between participants and items. In addition, we also added the factor 'gender' (male, female) to account for a potential effect of this variable. Grip similarity and goal similarity should be modulated by the prime duration factor. At 66 ms, we should expect an effect from goal similarity but not from grip similarity, whereas at 220 ms, we should detect an effect from the two dimensions. As previously, the Grip similarity x Goal similarity interaction should be visible as soon as the two main effects are observed.

3.2. Results

3.2.1. Data preparation for response times analysis

Again, the error rate was relatively low (M_{accuracy} = 4%, from 0.32% to 19%). A chi-square test for independence revealed that errors were equally distributed between conditions

$\chi2(4) = .031$, $p > .99$. Errors and RT superior to 1500 ms and inferior to 150 ms were considered outliers and removed. Overall, 4.9 % of the data were removed.

3.2.2. Mixed-effect model analysis of correct action recognition RT

As detailed in the General Methodology section, mixed effects linear models were used to analyse the RT. The model included grip similarity (GRIP), goal similarity (GOAL), prime duration (DURATION), response type (RESP), gender (GENDER) and related first-order, second-order, third-order and fourth-order interactions as fixed effects. The random structure was selected using the approach describe in the General Methodology section. The final random structure included random intercepts and random slopes for RESP for both participants and items.

3.2.3. Effect of grip and goal similarity priming

The analysis revealed no GRIP x GOAL x RESP x DURATION x GENDER *F*(1,19287.0) = 2.46, *p* = .117, Westfall's *d* = 0.30 and no GRIP x GOAL x RESP x DURATION interaction either *F*(1,19287.4) = 0.096, *p* = .756, Westfall's *d* = 0.03, the two variables were no longer considered. The GRIP x GOAL x DURATION was significant *F*(1,19287.5) = 56.99, *p* < .001, Westfall's *d* = 0.37. The remaining analyses were carried out independently for 66 ms and 220 ms.

The model analysing the data for primes lasting 66 ms included random intercepts and random slopes for GOAL for participants, and random intercepts and random slopes for GRIP for items. At 66 ms of visual processing, the GRIP x GOAL interaction *F*(1,9624.5) = 11.17, *p* < .001, Westfall's *d* = 0.12, the GRIP main effect *F*(1,18.7) = 18.20, *p* < .001, Westfall's *d* = 0.08 and the GOAL main effect *F*(1,60.7) = 85.23, *p* < .001, Westfall's *d* = 0.19 were significant. In both cases, the similar dimension yielded faster response times than the dissimilar dimension (Goal-dissimilar minus Goal-similar = 31.92 ms, SE = 3.46; Grip-dissimilar minus

Grip-similar = 14.14 ms, *SE* = 3.31). In contrast, the GRIP x GOAL interaction was driven by the fact that priming situation in which the grip and goal dimensions were not congruent (Grip-similar, Goal-dissimilar; Grip-dissimilar, Goal-similar) yielded slower response times than priming conditions in which grip and goal dimensions were congruent (Grip-similar Goal-similar; Grip-dissimilar Goal-dissimilar; estimate of the interaction = -20.07 ms, *SE* = 6.01).

The model analysing the data for primes lasting 220 ms included random intercepts for participants, and random intercepts and random slopes for GRIP for items. At 220 ms of visual processing, the GRIP x GOAL interaction *F*(1,9590.5) = 188.21, *p* < .001, Westfall's *d* = 0.49, the GRIP main effect *F*(1,19) = 60.46, *p* < .001, Westfall's *d* = 0.21 and the GOAL main effect *F*(1,9590.9) = 127.41, *p* < .001, Westfall's *d* = 0.20 were significant. In both cases, the similar dimension yielded faster response times than the dissimilar dimension (Goal-dissimilar minus Goal-similar = 33.59 ms, *SE* = 2.98; Grip-dissimilar minus Grip-similar = 35.57 ms, *SE* = 4.57). In contrast, the GRIP x GOAL interaction was driven by the fact that priming situations in which grip and goal dimensions were not congruent (Grip-similar Goal-dissimilar; Grip-dissimilar Goal-similar) yielded slower response times than priming situations in which grip and goal dimensions were congruent (Grip-similar Goal-similar; Grip-dissimilar Goal-dissimilar; estimate of the interaction = -81.66 ms, *SE* = 5.95).

In sum, grip similarity and goal similarity effects were both detected as soon as 66 ms of prime duration, with the effect of goal similarity being twice as big as the grip similarity effect at 66 ms. As predicted, the interaction was significant as soon as both dimensions influenced independently action recognition, namely at 66 ms. When primes were available for 220 ms of visual processing, the influence of the two dimensions (and their interaction) on action recognition was even more important. Results are presented in Figure 23.

3.2.4. Complementary analysis: Effect of object priming

As we did in Experiment 2, the conditions in which grip and goal overlapped between prime and target were replaced with the neutral condition. In other words, neutral prime-target pairs (object similar) were compared, first, with prime-target pairs that differed on grip (grip dissimilar conditions) and, subsequently, with prime-target pairs that differed on goal (goal dissimilar conditions). This comparison was tested in situations where effects of grip similarity and goal similarity were observed in the main analysis of Experiment 3.

When primes were available for 66 ms of visual processing, object similarity did not show any advantage over grip dissimilarity $F(1,9696.3) = 0.13$, $p = .72$, Westfall's $d = 0.00$. This stands in contrast with the grip similarity priming effect observed in the main analysis and indicates that object repetition is not responsible for the grip similarity priming effect. Object similarity did show a 10 ms advantage over goal dissimilarity $F(1,9670.2) = 11.46$, $p < .001$, Westfall's $d = 0.06$ (Goal dissimilar minus Neutral = 9.95 ms, $SE = 2.94$). This suggests that object similarity priming could at least partially account for the effect of goal similarity in the main analysis, as the object similarity effect is 5 times less important than the goal similarity effect (Goal-dissimilar minus goal-similar = 48.96 ms, $SE = 5.63$).

When primes were available for 220 ms of visual processing, dissimilar grips $F(1,9579.4) = 19.39$, $p < .001$, Westfall's $d = 0.16$ (Grip dissimilar minus Neutral = -12.83, $SE = 2.91$) were more facilitative than similar objects. Thus, the mere presence of the object could not generate the effect of grip similarity for 220 ms of prime duration. Similar objects did present an advantage over dissimilar goals $F(1,9617.3) = 17.59$, $p < .001$, Westfall's $d = 0.07$ (Goal dissimilar minus Neutral = 11.83, $SE = 2.82$) but as in Experiment 2, this effect was smaller than the advantage of goal-similar over goal dissimilar (Goal-dissimilar minus Goal-similar = 34.05 ms, $SE = 4.23$). Therefore, the presence of the object could only partially account for the effect of Goal similarity observed in the main analysis.

3.2.5. Complementary analysis: influence of perceptual similarity

As in the Experiment 1 and 2, Spearman's rank correlations were computed between the goal and grip similarity priming effects obtained in each response condition and the corresponding perceptual similarity indices. There were no significant correlations (all *p* > .08, see Appendix 7).

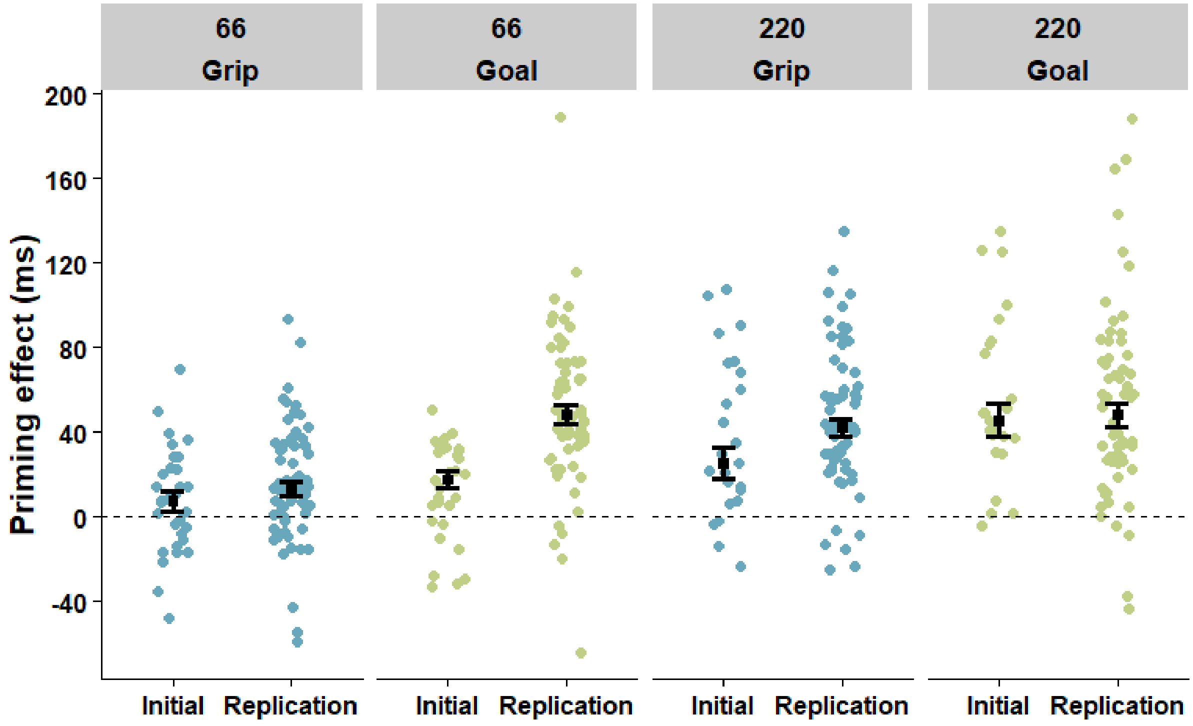


Figure 23. Individual mean priming effects in milliseconds according to grip, goal, prime duration and experiment type. For 66ms of prime duration, data from Experiment 1 (Initial) and Experiment 3 (Replication) are displayed. For 220ms of prime duration, data from Experiment 2 (Initial) and Experiment 3 (Replication) are displayed. Overall, the same pattern of data is observed. Small dots represent different individuals. Black dots represent the mean priming effects. Error bars represent standard error.

3.3. Interim discussion

In the third experiment, we sought to replicate the results we obtained in Experiment 1 and 2. The same priming paradigm was used with a different, more balanced, set of stimuli in a larger, gender-balanced, sample of participants. Grip and goal similarity both, independently, contributed to the recognition of the target action. Facilitative effects for primes sharing the same grip or the same goal were observed when primes were available for both 66 ms and 220 ms of visual processing. These results could not be fully explained by the presence of the object, though part of the effect of goal similarity could emerge from the processing of the object identity. As previously observed, when both grip and goal dimensions contributed to action recognition, an interaction was observed. Action recognition was easier when primes conveyed congruent information about the action (i.e., both grip and goal typical, or both grip and goal atypical) than when primes conveyed incongruent information about the action (i.e., grip typical but not goal or goal typical but not grip).

When primes were available for 220 ms of visual processing, the effect sizes obtained in Experiment 3 for grip and goal similarity were equivalent to those obtained in Experiment 2 (Westfall's *d* for Grip similarity: Experiment 2 for yes = 0.18, for no = 0.23; Experiment 3 = 0.21 ; Westfall's *d* for Goal similarity: Experiment 2 for yes = 0.30, for no = 0.34; Experiment 3 = 0.20). For 220 ms of prime duration, the replicated results clearly mirrored the original ones, which further support the independent contribution of grip and goal-related dimensions during action recognition. When primes were available for 66 ms of visual processing, results from Experiment 1 and 3 slightly differed. Overall, the effects reported in Experiment 3 were twice as big as the ones reported in Experiment 1 (Westfall's *d* for Grip similarity: Experiment 1 = 0.04, Experiment 3 = 0.08; Westfall's *d* for Goal similarity: Experiment 1 = 0.07, Experiment 3 = 0.19). In Experiment 1, we concluded that grip similarity was absent for 66 ms of prime duration, which contrasts with Experiment 3 in which grip similarity was already observed at 66 ms. Increased statistical power due to the larger sample size may partially account for this apparent discrepancy since in Experiment 1, grip similarity already approached significance at 66 ms. Furthermore, in the set of stimuli used in Experiment 3, the upper body of the actress was removed to balance the relative visibility of grips and visual goal. This could have favoured the emergence of an earlier contribution of the grip dimension. However it is important to note, that even in Experiment 3 the contribution of goal similarity was still twice as big as the one of grip similarity. We will go back to this issue later.

Contrasting with Experiment 2, no effect of response type was observed. The effects of response type in Experiment 2 were essentially related to primes available for 120 ms of visual processing: grip and visual goal dimensions already contributed for yes response, but not for no response while the results were similar for yes and no responses for 220 ms prime duration. Hence, we found no effect of response type in Experiment 3 for 220ms of prime duration. The effect of response type was already absent in Experiment 1, and unsurprisingly, when primes were available for 66 ms of visual processing, no modulation by response type was observed. Overall, this pattern is fairly congruent with the data obtained in Experiment 1 and 2.

Finally, the participant sample in Experiment 3 was gender balanced. Although there was no obvious a priori nor a posteriori reasons to expect gender effects, the possibility that gender could modulate action processing could not be firmly ruled out without testing. The gender did not interact with our effects, and, more importantly, the replication of the effects stands independently from the gender. This allows us to provide strong evidence that in our paradigm, gender has no major impact on the processing of target actions.

Overall, Experiment 3 nicely replicates the results of Experiment 1 and 2. Contrasting with our conclusions of Experiment 1, it may be wiser to conclude that both grip and visual goal dimensions contribute to action recognition as soon as 66 ms of action visual processing. Nonetheless, the overall pattern of results still indicates that grip and visual goal dimensions affect action processing with different time-course. Visual goal dimension affects action processing to a greater extent than grip dimension when primes are presented for 66 ms of visual processing, and the influence of visual goal dimension was maximal when primes were presented for 220 ms. In contrast, the influence of the grip dimension remains below the influence of the visual goal dimension at all durations of the prime but the last one, 300 ms.

When primes were presented for 300 ms of visual processing, the influence of the grip dimension was maximal and exceeded the one of the visual goal dimension.

4. Discussion on priming study

This first set of experiments assessed the timing of grip and visual goal activation during correctness judgements of photographs of object-directed actions. The selective activation of the grip or visual goal dimension was identified using action primes sharing the same grip or the same visual goal as target actions. Information about the activation time course was captured by varying the duration of the prime (66 ms and 300 ms in Experiment 1; 120 ms and 220 ms in Experiment 2; 66 ms and 220 ms in Experiment 3). Goal similarity and grip similarity affected the processing of the action target from 66 ms of prime visual processing. The influence of goal similarity was more important than that of grip similarity at all prime durations but the last one. This suggests that the influence of grip and visual goal during action recognition follows different time courses. Interestingly, action primes conveying incongruent grip and visual goal information interfered with action recognition when there was enough time for grip and goal representations to be co-activated. Overall, these results favour the idea that non-motor goal-related information is prioritised over visual kinematics and provide additional evidence for distinct but complementary roles of visual kinematics and non-motor goal-related information during the processing of others' actions. Before discussing the theoretical implications of these results, we first consider another experimental strategy to evaluate the time course of visual kinematics and non-motor goal-related information processing during action recognition.

PART 2 Visual Search Experiment

This work has been published in

Decroix, J., *& Kalénine, S. (2018). What first drives visual attention during the recognition of object-directed actions? The role of kinematics and goal information. Attention, Perception, & Psychophysics, 1–10. https://doi.org/ 10.3758/s13414-019-01784-7*

In our priming paradigm, action recognition was passively influenced by the processing of visual kinematics and non-motor goal-related information. In the present study, we wanted to evaluate the impact of visual kinematics and non-motor goal-related information during the active search of an action target. By measuring the evolution of gaze fixation during the visual search we were able to follow the relative attribution of attention over visual kinematics and non-motor goal-related information.

As aforementioned, goals have an important impact on action recognition (Hrkać, Wurm, & Schubotz, 2014; Novack, Wakefield, & Goldin-Meadow, 2016; Vallacher & Wegner, 1987, 2012; Zacks et al., 2001). Many pieces of evidence in this direction can also be found in the visual attention literature. During the observation of reach and grasp movements, both children and adults make proactive gaze movements towards the expected landing point of the action (Ambrosini et al., 2011; Flanagan & Johansson, 2003; Flanagan et al., 2013; Geangu et al., 2015), which suggests that observers do not simply follow the movement course as it unravels but predict and anticipate the goal of the action. This bias of interpreting actions as goal-directed seems to arise quite early in development. Infants are indeed able to track others' goals (Buresh & Woodward, 2007) and they show a renewal of attention when an actress stops her movement without achieving her goal (Baldwin et al., 2001). Although these data highlight the importance of goals during the perception and the recognition of actions performed by others (Ocampo & Kritikos, 2011), it remained to evaluate whether the importance was driven by the first processing of visual kinematics or by the first processing of non-motor goal-related information.

In the present study, the spontaneous orientation of visual attention towards visual kinematics or non-motor goal-related information was used to answer this question. Visual attention has indeed been found to impact the processes involved in the decoding of others' actions (D'Innocenzo, Gonzalez, Nowicky, Williams, & Bishop, 2017; Donaldson, Gurvich, Fielding, & Enticott, 2015; Leonetti et al., 2015; Muthukumaraswamy & Singh, 2008; Perry et al., 2010; Riach et al., 2018; Schuch et al., 2010; Woodruff & Klein, 2013; Wright et al., 2018) and to be affected by visual kinematics and non-motor goal-related information (Humphreys et al., 2013 for review). Yet the temporal dynamics of visual attention allocation on visual kinematics and non-motor goal-related information remains to be determined.

The present study aimed at investigating what captures attention first in an action discrimination task where observers are looking for correct actions among distractors that could have either the same grip or the same goal as the target action. The discrimination task is well-suited to directly and independently oppose grip and visual goal dimensions. In other words, is visual attention preferentially driven towards grip information or towards visual goal information that may help building a prediction about the actor's goal?

Here, we followed the repartition of eye movements during a visual search task to evaluate the influence of grip and visual goal information (e.g. orientation of the object) on the temporal allocation of visuospatial attention. Static photographs of actions were used, which allow displaying both grip and visual goal information at the exact same time. Grip configuration may not be as predictive of the outcome of the action as the full dynamic kinematic but significant changes in grip configuration can still be very informative of whether an action is correct or not. Moreover, grip configuration has been shown to be particularly important to identify what an actor is doing with an object (Naish et al., 2013). Therefore, visual kinematics have been manipulated through changes in grip configuration in our stimuli. Participants were then asked to find a photograph displaying a typical object-directed action among distractor action photographs. Distractor photographs displayed either a “similar visual goal but a dissimilar grip”, or a “similar grip but a dissimilar visual goal”, or both a “dissimilar visual goal and a dissimilar grip”. In case observers pay attention first towards the grip to derive the action goal, “similar grip but dissimilar action goal” distractors should capture visual attention earlier than “similar action goal but dissimilar grip” distractors. Alternatively, if observers first use visual goal information to orient the processing of kinematic information, then “similar action goal but dissimilar grip” distractors should capture visual attention earlier than “similar grip but dissimilar action goal”.

1. Methods

1.1. Participants

Twenty-two participants took part in the study[9]. Two participants were left-handed according to the Edinburgh handedness inventory (EHI; Oldfield, 1971) and were then excluded. One participant was excluded because of technical problems during the experimental session. Finally, two participants were excluded because of an atypical pattern of fixation in comparison to the remaining participants (see below). Eighteen participants (M_{age} 23, age range 18-27, 5 males) were then included in the final sample. All were right-handed (M_{EHI} 96%, from

---

[9] The sample size of ~20 participants was determined based on previous eye tracking studies using a similar paradigm (e.g. target search in 4-pictures display) in which differences of about 3% of fixation proportion were reported between two pictures in competition (e.g., Kalénine et al. (2012), Cohen *d* of 0.12 for the estimated difference of 3%). A power analysis indicated that twenty participants was sufficient to ensure a statistical power of 0.82 to detect difference in fixation proportion between two pictures (calculated from 50 simulations using the powerCurve function of the R package simr, Green & MacLeod, 2016).

63 to 100%), reported normal or corrected-to-normal vision. They provided written informed consent and were not paid for their participation. The study followed the ethical guidelines of the University of Lille and was in accordance with the declaration of Helsinki (1964, revised in 2013).

1.2. Stimuli and Design

Set 1 of stimuli was used in this experiment. Briefly, for each of the twenty objects, four 512 × 341 pixel coloured photographs of object-directed actions were designed by crossing the correctness of the grip and visual goal components of the action: the object-directed action could display a “typical grip and typical goal”, a “typical grip only”, a “typical goal only”, or an “atypical grip and atypical goal”. Detailed description can be found in the General Methodology section. An example of the stimuli can be found in Figure 24.

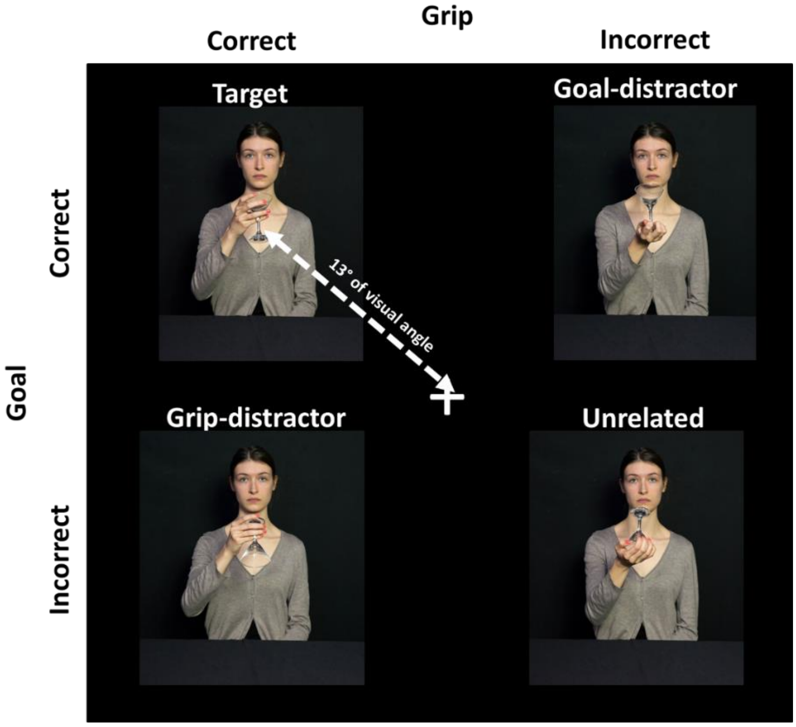


Figure 24. Design of the visual search experiment.

1.3. Procedure

Participants were comfortably seated in front of a 1024 x 768 computer screen in a quiet and darkened room. Head movements were restrained with a chin and forehead rest to reduce measurement errors. The vision was binocular but only the position of the left eye was recorded for all participants. Eye movements were measured continuously with an infrared video-based eye tracking system (EyeLink, SR Research), sampled at 500 Hz. Before each experimental session, the eye tracker was calibrated by asking participants to fixate a set of nine fixed locations distributed across the screen. After the calibration, instructions were given to each participant and a training session with feedback was provided. The training session included five representative trials with objects that were not in the experimental session. The experimental session was similar to the practice session, but without feedback. Each trial began with a fixation cross in the centre of the screen. Participants had to click on the fixation cross to make the display appear. For each reference object, pictures were randomly assigned to the different corners of the screen. The centre of each picture was at 13 degrees of visual angle of the centre of the screen. Participants were asked to click on the picture displaying the correct action according to the typical use of the object with the mouse. The "typical grip and typical goal" picture was defined as the "target", the "typical grip only" picture as the "grip-distractor", the "typical goal only" picture as the "goal-distractor", and the "atypical grip and atypical goal" picture as the "unrelated-distractor". Overall, there were 20 trials corresponding to each reference object. Eye movements were recorded from the beginning of each trial until the mouse-click response on the images.

1.4. Fixation proportion

Data analysis followed a procedure previously used in eye tracking studies to capture the evolution of eye movement distribution across time (Kalénine, Mirman, Middleton, &

Buxbaum, 2012; Lee, Middleton, Mirman, Kalénine, & Buxbaum, 2013; Lee, Mirman, & Buxbaum, 2014; Mirman, Dixon, & Magnuson, 2008; Mirman & Magnuson, 2009). Four areas of interest (AOI) associated with the displayed pictures were defined as the four 512 x 341 pixel quadrants of the 1024 x 643 pixel computer screen. We considered that participants fixated a given action type ("Target", "Grip-distractor", "Goal-distractor" and "Unrelated-distractor") when their gaze fell into the corresponding AOI. Fixation proportion on each action type was calculated over 50 ms time bins in order to reduce the noise in the fixation estimates and to facilitate statistical model fitting (see Data analysis section). For each time bin of each participant or each item, mean fixation proportion for each action type was computed by dividing the number of fixations on this action type by the total number of trials to avoid the selection bias introduced by varying trial-termination times (Mirman & Magnuson, 2009; Mirman, Strauss, Dixon, & Magnuson, 2010).

1.5. Saliency Maps

The experiment aimed at assessing which action dimension first drives visual attention when identifying a target action among distractors. Yet visual selective attention is largely influenced by the visual properties of the image to explore (e.g., colour, spatial orientation, intensity etc.). In order to partial out the effect of possible differences in low-level visual features between the four images on gaze behaviour during target action visual search, saliency maps were computed with the Saliency ToolBox for each stimuli (Walther & Koch, 2006). Subsequently, saliency values were extracted for each pixel and averaged across each area of interest (see "Fixation proportion" section). Therefore, a saliency index was available for each of the four pictures ("Target", "Grip-distractor", "Goal-distractor", "Unrelated-distractor") of each of the 20 displays. Paired comparisons showed a perceptual advantage for the "Goal-distractor” over the “Grip-distractor”, $t_{19} = -5$, $p < .001$. Thus, saliency indices were added as covariate in a complementary by-item analysis.

1.6. Data Analysis

The temporal dynamics of fixations on the two “Grip” and “Goal-distractor” pictures were compared in order to determine whether visual attention is first captured by grip or by visual goal information[10]. To capture the effect of time, fixation proportions over time were fitted as a function of fourth-order orthogonal polynomials. Orthogonal polynomials are well suited to characterise different behaviours of the fixation curves (see Mirman, 2014 for an introduction to Growth Curve Analysis). Fourth order polynomials were chosen since they have been proven successful to capture the rise and fall of the fixation curves of competing distractors during target identification (Mirman, 2014; Mirman et al., 2008). The intercept reflects differences in the overall height of the curve between conditions. In the present study, intercept differences between goal and grip distractors would not inform on which action dimension is processed first and was not of primary interest. Differences in timing between grip and goal processing would be particularly reflected by differences on the linear (1st order) and/or on the cubic (3rd order) time terms (Kalénine et al., 2012; Lee et al., 2013). If visual attention is first captured by grip information, then we should observe earlier fixations on the “Grip-distractor” in comparison to the “Goal-distractor”. This would be reflected by a more negative linear estimate (slope) or cubic estimate for the goal compared to the grip fixation curve. Conversely, we should observe earlier fixations on the “Goal-distractor” compared to the “Grip-distractor” if visual attention is first captured by goal information. This would be reflected by a more

---

[10] As the target fixation curve was incomparable to distractor fixation curves, the target was not included in the analysis (see Figure 25). The unrelated distractor was not added in the analysis because it has no influence on the test of our main hypothesis while increasing model complexity.

positive linear estimate or cubic estimate for the goal compared to the grip fixation curve. For example, the cubic time term has been shown to be sensitive to differences in early and late inflexions of the fixation curves (see Figure 3 of Kalénine et al., 2012 for an illustration). An early increase of fixation proportion on the "Goal-distractor" in comparison to the "Grip-distractor" would thus be statistically reflected by an interaction between the variable "distractor type" and the cubic (3rd polynomial order) time term.

In the main analysis, fixation proportions on the distractor pictures were averaged over items and analysed as a function of the fixed-effect factors of time (fourth-order orthogonal polynomials), distractor type ("Grip-distractor", "Goal-distractor") and the interaction between the two factors. The random structure includes random slopes for participants on each time term[11]. In a complementary analysis, fixation proportions on the distractor pictures were averaged over subjects and analysed as a function of the fixed-effect factors of time, distractor type, their interaction, and image saliency index and its interaction with time. By adding the saliency index covariate to the model, this complementary by-item analysis aimed at partialling out the influence of low-level visual features on the fixation curves. The random structure includes random slopes for items on each time term[12]. As detailed in the General methodology section, fixation proportions were analysed using mixed-effects linear models.

Overall main effects and interactions were evaluated with *F* statistics and *t* statistics individual parameter estimates were then used to evaluate the contrasts of interest between distractors.

---

[11] The R syntax of the model was: Fixation proportion = (intercept + linear + quadratic + cubic + quartic) * distractor type + (intercept + linear + quadratic + cubic + quartic | participant : distractor type)

[12] The R syntax of the model was: Fixation proportion = (intercept + linear + quadratic + cubic + quartic) * distractor type + (intercept + linear + quadratic + cubic + quartic) * saliency index + (intercept + linear + quadratic + cubic + quartic | participant : distractor type)

2. Results

2.1. Main analysis of fixation proportions

Overall, only trials on which the target image was correctly identified were included in the fixation analyses (M_{accuracy} 91% +/- 28%). As the task was to find the target action, two participants for whom fixations on the target never reached at least 50% of all fixations were considered performing the task correctly but with an atypical visual strategy and were excluded from the analysis. After visual inspection, the time-window of analysis was selected from display onset to 1500 ms after display onset, when the averaged target fixation curve reached a first plateau (see Figure 25 and Lee et al., 2013; Mirman et al., 2008 for similar procedure).

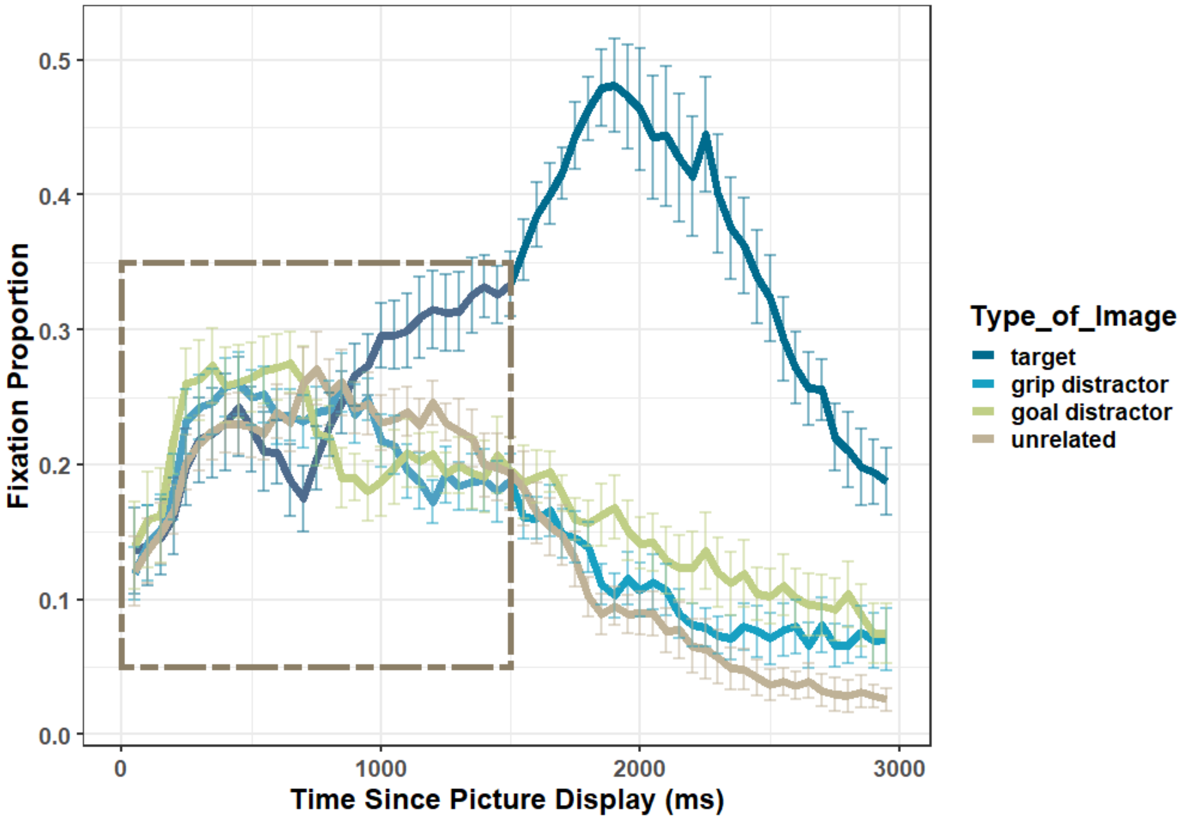


Figure 25. Mean fixation proportion and standard errors (error bars) over time as a function of image condition. The dashed-rectangle represents the window of interest for the analysis.

The analysis showed no main effect of distractor type, $F_{1,34} = 0.45$, $p = .506$. This indicates that overall, grip and goal distractors received an equivalent proportion of fixations over the whole 1500 ms time window ("Grip-distractor" *M* proportion 0.21; "Goal-distractor" *M* proportion 0.22). Importantly, however, a significant interaction was found between distractor type and the cubic (3rd order) time-term, $F_{1,34} = 4.77$, $p = .041$, reflecting an influence of distractor type on the time course of fixation proportion. The distractor type x cubic time term interaction was driven by an earlier increase of fixation proportion over the "Goal-distractor" in comparison to the "Grip-distractor" (estimate = - 0.13 ms, *SE* = 0.06), as visible on Figure 26. Distractor type did not interact with any other time terms (all $p > .157$).

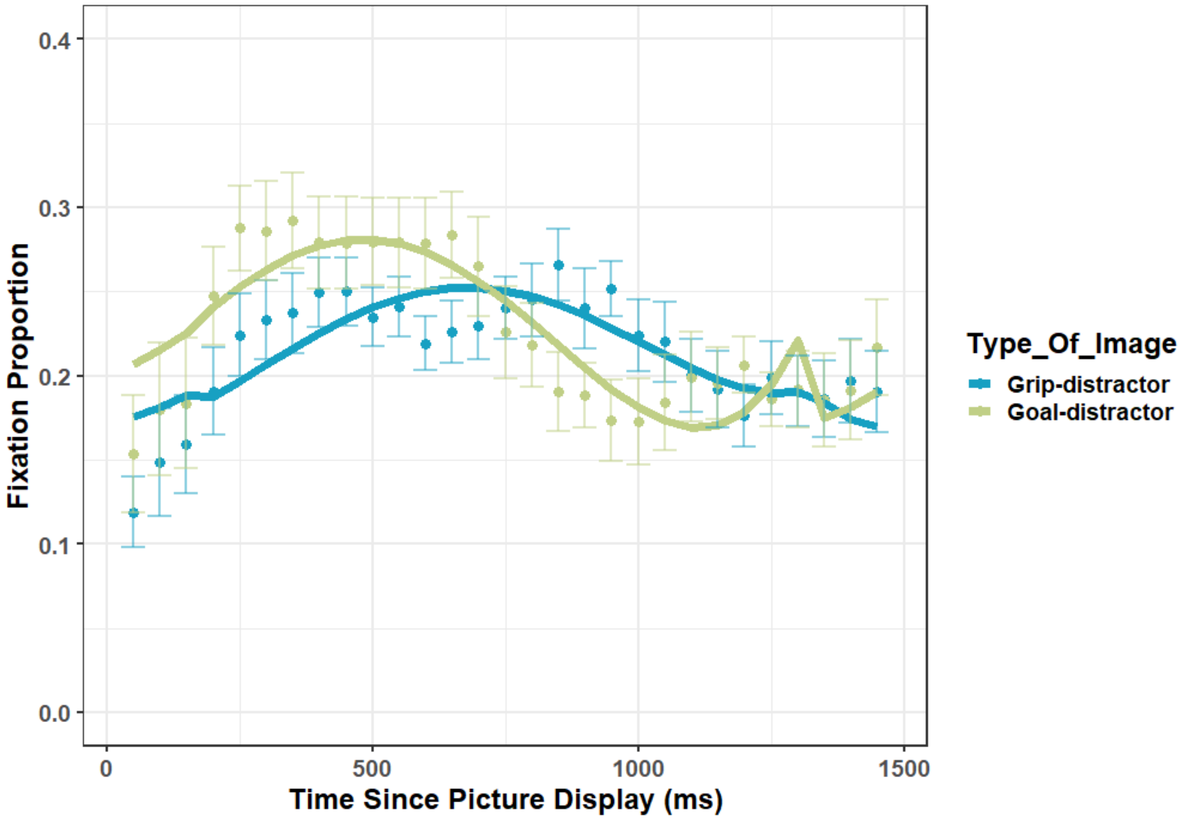


Figure 26. Model fit of the data (lines) and actual data (dots) for the Grip-distractor and Goal-distractor.

2.1.1. Complementary analysis of fixation proportions with saliency index as covariate

In the complementary by-item analysis including the saliency index, the interaction between distractor type and the cubic (3rd order) time-term was marginally significant, $F_{1,37} =$

3.76, $p = .060$, after taking the saliency index into account. As previously observed, there was an earlier rise in fixation proportion over the “Goal-distractor” in comparison to the “Grip-distractor” (estimate = - 0.13 ms, *SE* = 0.06). Importantly, there were no effects involving the saliency index on fixation proportions, neither in isolation (main effect: $F_{1,37} = 0.15$, $p = .706$), nor in interaction with the different time terms (all $p > .477$). In addition, at the item level, no correlation was found between the amplitude of grip and goal processing early in the time window (extracted from the random cubic estimates for items) and the saliency index (“Grip-distractor” condition: $r = 0.37$, $p = .107$; “Goal-distractor” condition: $r = -0.2$, $p = .399$). Overall, the complementary analysis indicates that we can be confident that the earlier fixations on goal-distractors cannot be fully explained by the greater visual saliency of the images in this condition.

3. Interim Discussion

The present study aimed at investigating the spontaneous capture of visual attention by grip and visual goal information. More specifically, we wanted to determine whether visual attention would be preferentially driven towards grip-related or goal-related information. In a visual search task, participants were asked to explore and select the photograph displaying the correct tool use action among action distractors. Gaze movements were used to evaluate to what extent grip-related (same grip as the target, but with a different action goal) and goal-related (same goal as the target but involving a different grip) distractors would capture participants’ visual attention before the identification of the target. Visual attention was found preferentially captured by goal-related distractors in comparison to grip-related distractors, but in a time-dependent manner. Visual attention over the goal-related distractors increased in the first part of the visual exploration (before 750 ms) but decreased in the second part (after 750 ms). Thus, observers do not only use non-motor goal-related information overall when decoding others’ actions, they rely on it first. They disengage their attention from it afterwards to use the other available information.

In the present study, we show that very early in the action recognition process, non-motor goal-related information is favoured over visual kinematics when the two dimensions are competing for attention (i.e. visual search of the correct actions). Donnarumma et al. (2017) observed that gaze movements behave in a predictive manner when predictions about the actor's goal were available, but in a reactive manner when predictions about the actor's goal were not available. Our study further suggests that when participants have no information about the actor's goal, they will still first look for non-motor goal-related information, probably to try to derive a prediction about the actor's goal. The visuo-attentional system would thus be affected by the predictive mechanisms at play in action recognition.

Although the gaze pattern corroborates the "goal first processing" hypothesis, visual attention during the action discrimination task was not solely captured by correct goal-related information but was also influenced by correct kinematic information. The disengagement of visual attention from the goal-related distractor in the second part of the visual exploration provides further evidence for the use of visual kinematics during action recognition. Recently, Koul et al. (2019) showed that actual visual kinematics are used to update the on-going motor simulation as a function of their informativeness. These data fits well with the overall pattern of fixations reported here, as visual kinematics became more relevant than goal-related information in the second part of the visual exploration.

Overall, the first capture of visual attention by non-motor goal-related information, and the subsequent withdrawal of visual attention from such information to favour a later attentional capture by visual kinematics is in line with predictive approaches of action recognition.

SYNTHESIS

Whereas the involvement of grip and visual goal dimensions during the processing of others' actions is well supported, their specific role and temporal organisation remain debated. In the previous chapters we dissociated two main groups of theories on the basis of the hypothesis they made on the place of visual kinematics during action recognition. Sensorimotor approaches hypothesise that action recognition is first guided by the processing of visual kinematics, whereas predictive approaches argue that predictions about others' goal are needed to make sense of the visual kinematics. On this basis, we hypothesised that if priority was given to visual kinematics, it would support sensorimotor approaches, whereas predictive approaches would be favoured if priority was given to non-motor goal-related information during the recognition of others' actions. These predictions were evaluated with a priming paradigm and a visual search paradigm. Three experiments investigated how much time of visual processing was necessary for visual kinematics and non-motor goal-related information to influence the recognition of others' actions. The contribution of the two dimensions was detected as soon as 66 ms of visual processing. Yet the pattern of results still suggested that non-motor goal-related information has a greater influence on action recognition than visual kinematics information early in the processing of others' actions, whereas the influence of visual kinematics was greater than the one of non-motor goal-related information later in the processing of others' actions. This pattern of results was further confirmed by the visual search task experiment, as non-motor goal-related information captured visual attention during the first hundred milliseconds of visual search, while visual kinematics captured visual attention later during visual search. Overall the results provided by the behavioural experiments re-affirm the contribution of visual kinematics and non-motor goal-related dimensions during the processing of others' object-directed actions, with priority first given to non-motor goal-related information.

The importance of goals and goal-related information have already been highlighted by several theoretical account (Bach et al., 2014; R. P. Cooper, Ruh, & Mareschal, 2014; van Elk et al., 2014a) and is supported by many experimental arguments (Flanagan & Johansson, 2003; Nicholson, Roser, & Bach, 2017; van Elk et al., 2008). Until now, the importance of goal and goal-related information have been supported at the level of the final recognition of the action. Yet the greater weight of non-motor goal-related information in action recognition is not sufficient to support the idea that non-motor goal-related information is first needed to make sense of the visual kinematics, as the strong influence of goal information could have been derived from the first analysis of visual kinematics (Kilner & Frith, 2008; Tidoni & Candidi, 2016). Some EEG studies previously argued in favour of the early importance of goal-related information during action processing (e.g., Ortigue et al., 2009) but they did not dissociate visual kinematics and non-motor goal-related information, it was then still possible that goal-related information could have been derived from visual kinematics. Our studies provide direct evidence that non-motor goal-related information is prioritised during action recognition, both when the task puts minimal requirement on the visuo-attentional system (i.e., in the priming paradigm, with central presentation of one action picture at a time), and with stronger loads on the visuo-attentional system (i.e., in the visual search task, when the two dimensions directly compete for visual attention).

Finally, the later involvement of visual kinematics during action recognition must not undermine its importance. Predictive approaches suggest that visual kinematics are used to test the goal prediction that has been derived from non-motor goal-related information (Donnarumma et al., 2017; Kilner, 2011; Kilner et al., 2007). Converging evidence suggests that visual kinematics are used to update predictions about the actor's action goal. Motor simulation has been shown to reflect the *expected* visual kinematics during the first steps of action observation but the *actual* visual kinematics during the last steps of action observation

(Cavallo et al., 2013; see Chapter 3). Recently, Koul et al. (2019) further showed that actual visual kinematics are used to update the on-going motor simulation as a function of their informativeness. The later influence of visual kinematics observed in our behavioural results further reinforce such interpretations.

In the next chapter, we will be interested in the neurophysiological basis of action recognition, and we will try to identify the nature of the mechanisms involved in the processing of visual kinematics and non-motor goal-related information.

CHAPTER 2:
NEUROPHYSIOLOGICAL CORRELATES OF GRIP AND VISUAL GOAL INFORMATION

In the previous chapter, we tried to identify whether the processing of grip and visual goal information could be reflected at the behavioural level. Our results provide additional evidence that both dimensions are involved during action recognition. More importantly, non-motor goal related information is found to be prioritised over visual kinematics during the first steps of action recognition, whereas visual kinematics are rather prioritised during the last steps of action recognition; when both types of information are used, they are integrated to one another. Nonetheless, behavioural methods are informative as long as the processes being targeted have an effective effect on behaviour. In other words, the brain could process some pieces of information, and yet not use them in the task of interest. In our case, grip configuration could have been processed at the same time as visual goal information without having any influence on the decision. Such information could have important consequences on models of action recognition. In particular, predictive approaches did claim that interpreting visual kinematics would require predictions about the actor’s goal. It may be possible that visual kinematics are first perceptually processed, but yet play no role in action recognition before the computation of goal predictions. This interpretation would still be consistent with the results found in our behavioural experiments. In contrast with behavioural methods, most of the neurophysiological methods have the double-edged characteristic to record the activity of the brain as a whole. Thus, processes that cannot be recordable at the behavioural level may be easier to capture with neurophysiological data. In addition, we have acknowledged that action processing was essentially the problem of the brain in Chapter 1 (see also Jeannerod, 1994, 2009), and thus, we are also interested in the neurophysiological basis of action recognition on its own.

In the first part of this chapter, we tackle the issue of the neurophysiological basis of action recognition through an event-related potential (ERP) study. ERP are recorded using electroencephalography (EEG) and allow the measurement of the electrical activity of the brain. Contrasting with methods such as *f*MRI, EEG benefits from its high temporal resolution and thus, is particularly relevant to investigate the neuronal dynamics underlying action recognition. The second part of this chapter will tackle the issue of the neurophysiological correlates of action recognition from a different perspective. We tried to establish the critical (i.e., causal) involvement of the frontoparietal network during action recognition while targeting a particular step of action recognition. One of the major limitations of most neurophysiological methods lies in the correlational nature of the evidence they provide. The brain activity is observed–recorded–along with some cognitive/perceptive/motor tasks, but it is not possible to conclude that the recorded activity is at the origin of the processes involved in these tasks. In contrast, transcranial magnetic stimulation (TMS) is used to apply a magnetic field next to a given brain area, which modifies the activity of the neurons in this stimulated area. If the neurons in the stimulated area are involved in the cognitive/perceptive/motor task participants are performing, their performance in the task should be different from conditions in which no TMS was applied. Therefore, the involvement of the frontoparietal network during action recognition will be investigated using TMS, coupled with a paradigm allowing to target particular moment in the process of action recognition.

PART 1: ERP correlates of grip and visual goal information decoding during the processing of others' actions with objects

This work was submitted for publication:

Decroix J., *Roger, C. & Kalénine S. (2019). Electrophysiological correlates of grip and goal decoding during the visual processing of others' object-directed actions. Manuscript submitted for publication.*

EEG is particularly salient to investigate the neural temporal dynamic of action processing. In particular, the technique of ERP has been used to deepen our understanding of the processes underlying the decoding of others' actions. ERP components not only provide information about the temporal course of action processing, but may also inform about the neuronal computation and the cognitive mechanisms[13] underlying the generation of the ERP components (see Luck, 2014 for a discussion of the different possible definitions of ERP components). Thereby, ERP are used here not only to establish the neurophysiological correlates of action recognition, but also to inform about the different processing steps at which visual kinematics and non-motor goal related information are involved. However, identifying an ERP component is not easy, and requires the conjunction of several conditions such as a particular topography, a particular polarity of the waveforms, or a particular experimental context (Kutas & Federmeier, 2011; Luck, 2014). Such conjunctions are rarely met in the literature of action recognition, and our aim is to provide further characterisations of the ERP components generated during action recognition.

---

[13] We discussed the difference between neuronal and cognitive mechanisms briefly in the first chapter, when we defined the notion of neural versus cognitive representations. We decided not to have any strong distinction between the two, as we believe that any cognitive representation is supported by some kind of neuronal mechanisms.

Several late ERP components have been related to the processing of visual actions. Modulations of the N400 component by the congruency between an action and the context in which it takes place (e.g., squeezing a lemon in the bathroom instead of the kitchen) have been repeatedly reported (Amoruso et al., 2013; Maffongelli et al., 2015; Mudrik, Lamy, & Deouell, 2010; Proverbio & Riva, 2009; Proverbio, Riva, & Zani, 2010; Sitnikova, Holcomb, Kiyonaga, & Kuperberg, 2008; Sitnikova, Kuperberg, & Holcomb, 2003). Such modulations have been interpreted as a marker of the integration between different pieces of action information (Amoruso et al., 2013). Yet, these studies usually involve semantic violations that encompass several pieces of action-relevant information at the same time (e.g., the object, the visual context, the action goal etc.). Accordingly, it can be a matter of difficulty to precisely identify the contribution of each piece of information in the observed N400 component modulations. Trying to dissociate the spatial relationships and the functional relationships between objects, Bach, Gunter, Knoblich, Prinz, and Friederici (2009) demonstrated similar N400 modulations when processing violations of the spatial relationships between objects (e.g., inserting a screwdriver in a screw with a matching versus mismatching orientation), and violations of the functional relationships (e.g., inserting a screwdriver in a keyhole versus a screw, both with a matching orientation). Spatial relationships corresponded to the motor interaction between the two objects, and may thus be related to motor act representations. Functional relationships can be related to the processing of non-motor goal-related components of the action, as they referred to the conceptual relevance between the two objects. Therefore, both visual kinematics and non-motor goal-related dimensions may independently contribute to modulations of the N400 component. This further suggests that the two dimensions can be dissociated and reflected in the components of visual ERPs, hence complementing previous behavioural results (see Chapter 4; van Elk et al., 2008). Focusing on the recognition of hand-object directed actions (i.e., does the action picture display a typical use of the object), Chang et al. (2018) reported modulations of the N300 component by the congruency between the hand grip and the object used in the action (e.g., using a precision grip on an upright pencil versus using a power grasp on an upside-down pencil). Nonetheless, it is unclear whether N300 modulations reflect the effect of grip congruency (related to the visual kinematics), or the impossibility to use the object for its typical function (related to non-motor goal-related information). In another EEG study (van Elk et al., 2012), the independent manipulation of visual kinematics and goal-object dimensions revealed an earlier modulation of the P300 component for goal-object violations (e.g., using a nail on a hammer) when compared to grip violations (e.g., grasping the hammer by its head instead of its handle). Importantly, the expected integration of non-motor goal-related information and visual kinematics (reflected by the statistical interaction between goal and grip violations) was not observed on this ERP component. However, it should be noted that participants were explicitly asked to independently judge whether the goal-object or the grip of the action was correct. The authors themselves suggested that the absence of integration between visual kinematics and non-motor goal-related information might have been explained by the specific task demands. Thus, P300 modulations may not be related to the spontaneous recognition of observed actions, which should require at some point the integration between the two action dimensions. Together, previous ERP findings indicate that P300, N300, and N400 components reflect the processing of different dimensions of observed actions; nevertheless, their selective sensitivity to visual kinematics, goal-related information, or the integration between the two dimensions remain to be identified.

Finally, the cognitive mechanisms underlying the modulation of the aforementioned ERPs unlikely reflect the initial (first) stages of action processing. Indeed, the previous ERP modulations have been related to the access to stored manipulation (i.e., related to the grip configuration) and functional knowledge (i.e., related to the visual goal information) relevant to the use of the object (e.g., N300 component; Chang et al., 2018), or associated to the integration of the two action dimensions (e.g., the N400 component; Amoruso et al., 2013). The processing of others' actions clearly begins much earlier. Different EEG techniques have for example found that discriminating between grasp-to-move and grasp-to-use actions modulated brain activity as early as 60 ms of action processing (Avanzini et al., 2013 with micro-state analysis; Ortigue et al., 2010 with micro-state analysis, 2009 with repetition-suppression paradigm). Whether such early modulations can be related to actual action processing or merely reflect perceptual differences in the study design remains debated (Catmur, 2015; Heyes, 2014). Moreover, early differences between visual actions performed with distinct goals have been, again, interpreted as evidence of early brain sensitivity to either visual kinematics (Ortigue et al., 2010) or action goal (Ortigue et al., 2009).

The present study aimed at characterising the ERP correlates of non-motor goal-related information and visual kinematics decoding at both early and late stages of action processing. The present paradigm used photographs of object-directed actions (e.g., writing with pencil) displaying a hand and a tool object. Actions could be typical or not according to the typical use of the object by the introduction of grip violations (e.g. upright pencil grasped with power grip), visual goal violations (e.g., upside-down pencil grasped with precision grip), or both grip and visual goal violations (e.g., upside-down pencil grasped with power grip). Grip violations did not prevent the performance of the typical goal of the action and vice versa, so that the two dimensions varied independently from one another. Importantly, object identity was kept constant across conditions, and object-related knowledge was equally diagnostic of grip and visual goal typicality. The concept of pencil is both associated to the typical functional goal of writing and to the typical precision grip for using the object. Therefore, any differences in processing actions with grip and visual goal violations would not merely reflect differences in the activation of object knowledge between conditions, but would rather be related to the activation of different action representations. In order to assess the spontaneous recognition of others' actions, participants were not explicitly asked to pay attention to one or the other dimension. They had to evaluate whether overall actions were correct or not while their brain activity was recorded using EEG. Differences in ERP amplitude as a function of grip typicality (grip-typical vs. grip-atypical, independently of goal typicality) and goal typicality (goal-typical vs. goal-atypical, independently of grip typicality) were assumed to reflect the decoding of the grip and the decoding of visual goal information, respectively. Differences in ERP amplitude as a function of grip and visual goal congruence (grip and visual goal dimensions congruent vs. incongruent, regardless of which dimension is correct and which is incorrect) were assumed to reflect the integration of grip and visual goal dimensions. We expected the integration between visual goal and grip dimensions to be visible on late ERP components (e.g., N300, N400), as the processing of incongruencies between different action dimensions has been notably detected in such time range (Bach et al., 2009; Chang et al., 2018; Giglio, Minati, & Boggio, 2013; Maffongelli et al., 2015). In addition, we wanted to evaluate whether independent grip and visual goal processing would be visible on earlier ERP components. Modulations of brain activity related to visual goals, but not to grip configuration, before the integration of the two dimensions would support predictive approaches of action recognition.

1. Methods

1.1. Participants

Thirty-one participants took part in the study. Three participants were excluded because of excessive noise in the EEG signal. The twenty-eight remaining participants (M_{age} = 21, age range: 18–29, 10 males) were all right-handed (M_{EHI} = 83%, range 27%-100%; Oldfield, 1971), and reported normal or corrected-to-normal vision. They provided written informed consent and received twenty euros for their participation. The protocol was approved by the Ethical

Committee of the University of Lille and was in accordance with the declaration of Helsinki (1964, revised in 2013).

1.2. Stimuli

The second set of stimuli was used in this experiment. Briefly, for each of the twenty reference objects, five coloured 1024 x 683 pixels photographs were taken, all displaying a hand and a tool object. Four out of the five pictures of the set presented hand-on-object actions. The remaining photograph corresponded to a no-action picture showing the hand and the object without any interaction between them. For each reference object, actions could be typical or not along the grip dimension and/or the visual goal dimension. Detailed description can be found in the General Methodology section. An example of the stimuli can be found in Figure 27.

Overall, 100 picture stimuli were divided in five conditions: “goal-typical grip-typical”, “goal-atypical grip-typical”, “goal-typical grip-atypical”, “goal-atypical grip-atypical”, “no-goal no-grip” (goal and grip were then neither typical nor atypical)[14].

1.3. Procedure

Participants first provided written informed consent. Afterwards, they were seated in a dimly illuminated room in front of a computer screen (1024 x 768, 60 Hz) and the EEG cap was positioned. Participants were carefully instructed to avoid eye and body movements during the recording session. They began the experiment with a training session involving 12 representative trials with three objects not included in the experimental session. Feedbacks were

---

[14] This “neutral” condition was first included as a control condition but was finally not analysed as being processed very differently from the other conditions.

provided during the training session to inform participants whether their answer was accurate or not. The experimental phase was equivalent to the training session but without any feedback.

Each trial started with a black screen of 1500 ms followed by the object-related action photograph for 1000 ms. In 12% of the trials, participants had to determine whether actions were correct or incorrect. They were explicitly told that a correct action corresponded to the typical use of the object. In those trials, the photograph was followed by a screen on which “correct” and “incorrect” were written on each side of the screen. The left/right position of the correct/incorrect responses was counterbalanced across trials so that participants could not prepare their motor response while processing the action photograph. This choice was made to avoid contamination of the relevant EEG signal by motor preparation. The response screen remained visible until the participant’s response (maximum 3000 ms). Participants responded on two separate keys of a response box with their left and right thumbs for correct vs. incorrect, respectively. After their response, a black screen with “ok” was displayed for 1000 ms. The “ok” screen was only for participants to know that their answer had been taken into account but was not informative in regard to the accuracy of their answer. The “ok” screen further allowed to avoid introducing variation on trial duration due to variation in participant’s response. Each trial was repeated six times, in six different blocks. Overall, there was 20 objects x 5 conditions x 6 repetitions = 600 trials. Consequently, there was a maximum of 120 trials per condition per participant. All conditions and objects were equally represented in each block. Trials inside blocks and blocks were randomly presented. The design is presented in Figure 27. Breaks were proposed between blocks. Blocks lasted about 7 mn each, and the overall experiment lasted ~two hours. The experiment was conducted with E-Prime V2.0.10.353 software (Psychology Software Tools, Pittsburgh, PA).

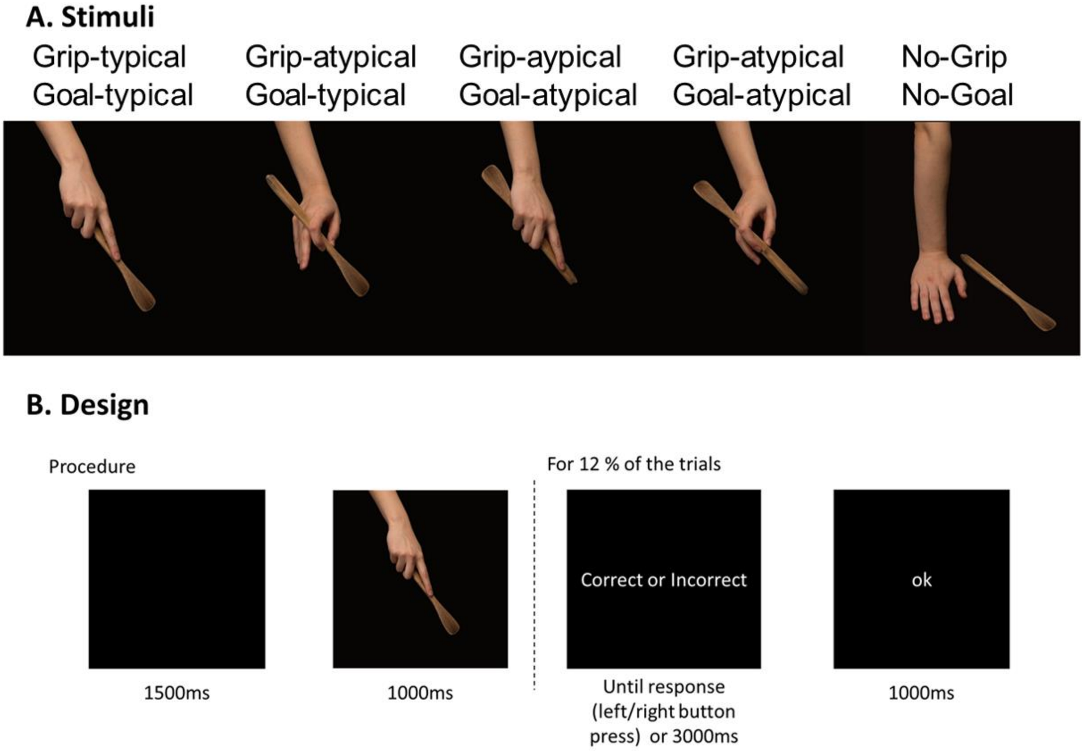


Figure 27. Design and procedure of the experiment. ***A.*** *Stimuli were divided in four experimental conditions by manipulating the typicality of the grip and the visual goal of the action. A fifth neutral condition was added as a control of object information.* ***B.*** *Procedure on a given trial. Responses were prompted for only 12% of the trials.*

1.4. EEG recording and analysis

EEG data were continuously collected from 128-channel Biosemi ActiveTwo (Biosemi B.V., Amsterdam, Netherlands) at a sampling rate of 1024 Hz by means of ActiView software. Electrode caps covering the whole head with equidistant layout were used. Electrode offset was kept below 20 µV. The offset values were the voltage difference between each electrode and the CMS-DRL reference. Electrooculographic (EOG) activities were bipolarly recorded using electrodes placed near both canthi (for measuring horizontal eye movements) and, below and above the left eye (for measuring vertical eye movements, i.e., blinks). Four additional electrodes were placed above the flexor pollicis brevis of each hand to monitor the electromyographic activity of the thumb (two on the right hand, two on the left hand). The electromyographic data were not directly related to the present study, and will not be discussed any further. A last electrode was placed on the left mastoid. Offline analysis was performed using BrainVision Analyzer 2.1 (Brain Products GmbH, Munich, Germany). One electrode (D7) did not register the brain activity for all subjects, and was interpolated[15]. Eye movement artefacts were first corrected using the Gratton and Coles' method (Gratton, Coles, & Donchin, 1983; Miller, Gration, & Yee, 1988). Remaining artefacts on the signal were marked manually by visual inspection on the continuous recorded EEG signal, regardless of the conditions. The raw signal was then filtered using a high pass filter at 0.1 Hz (zero-phase shift Butterworth filter, order 2) and a low pass filter at 100 Hz (zero-phase shift Butterworth filter, order 4). The continuous EEG signal was re-referenced on average reference[16]. The signal was then segmented into 1200 ms periods (200 ms before the action photograph onset, 1000 ms after action photograph onset). At this point, epochs contaminated by artefacts were not considered for subsequent analysis. About 15% of the trials were removed (*M* +/- *SD* remaining trials per participant: "goal-typical grip-typical", *M* = 103 +/-9 trials; "goal-atypical grip-typical", *M* = 104 +/- 8 trials; "goal-typical grip-atypical", *M* = 103 +/- 8 trials; "goal-atypical grip-atypical", *M* = 101 +/- 9 trials; "no-goal no-grip", 100 +/- 10 trials). Baseline correction was applied using the 200 ms time-window pre-action photograph onset. Finally, the EEG signal was averaged across trials for each condition.

ERPs were averaged across all participants and all conditions to define the analysis parameters (as recommanded by Luck & Gaspelin, 2017). Five ERP components were identified on the collapsed waveforms: P100 (90–140 ms), N170 (140–200 ms), P300 (200–260 ms), N300 (260–380 ms) and N400 (380–500 ms). Scalp map distributions were used to

---

[15] This electrode was not considered in the following analysis. Its interpolation could not affect the results.

[16] The left mastoid was considered as a reference but could not be used because of excessive noise in the mastoid signal.

gather neighbouring electrodes that show the greatest activity for each component (see Leek, Atherton, & Thierry, 2007; Wamain, Pluciennicka, & Kalénine, 2014, 2015 for similar procedure). ERPs were collapsed across B6–B7–B8–A28, and across A9–A10–A11–A15 to represent maximal posterior right and left activity, respectively, for P100, N170 and P300 (see Bledowski et al., 2004; Boehm, Dering, & Thierry, 2011; Bortoletto, Mattingley, & Cunnington, 2011; Kovács et al., 2006; Kumar, Yoon, & Humphreys, 2012; Leek et al., 2007; Peelen & Downing, 2007; Wamain et al., 2014, 2015 for similar timing and topography). ERPs were collapsed across C28–C27–C26–C18–C19–C20–C15–C14–C13 to represent maximal anterior central activity for N300 and N400 (see Amoruso et al., 2013; Bach et al., 2009; Chang et al., 2018; Renoult, Wang, Calcagno, Prévost, & Debruille, 2012; Wamain et al., 2014, 2015 for similar timing and topography). Scalp map distributions and corresponding grand average ERP for the collapsed electrodes are presented in Figure 28 for posterior site, and Figure 29 for anterior site.

Mean peak amplitudes and peak latencies (when available) were used as dependent variables. In order to best capture individual variability, mean peak amplitudes for the P100, N170, and P300 components were obtained for each participant by averaging the EEG activity on a +/- 10 ms time-window around each individual maximum peak for each component and condition (see Wamain et al., 2014 for similar procedure). Peak latencies were obtained using the timing of the maximum peak for the P100, N170, and P300 components, respectively, for each individual and each condition. The identification of individual peaks for the N300 and N400 components was not always apparent, as it is usually the case for late components. Thus, mean peak amplitudes for these components were obtained by averaging the activity over each time-window (260–380ms for the N300, and 380–500 for the N400). As a consequence, peak latencies were not analysed for the N300 and N400 components.

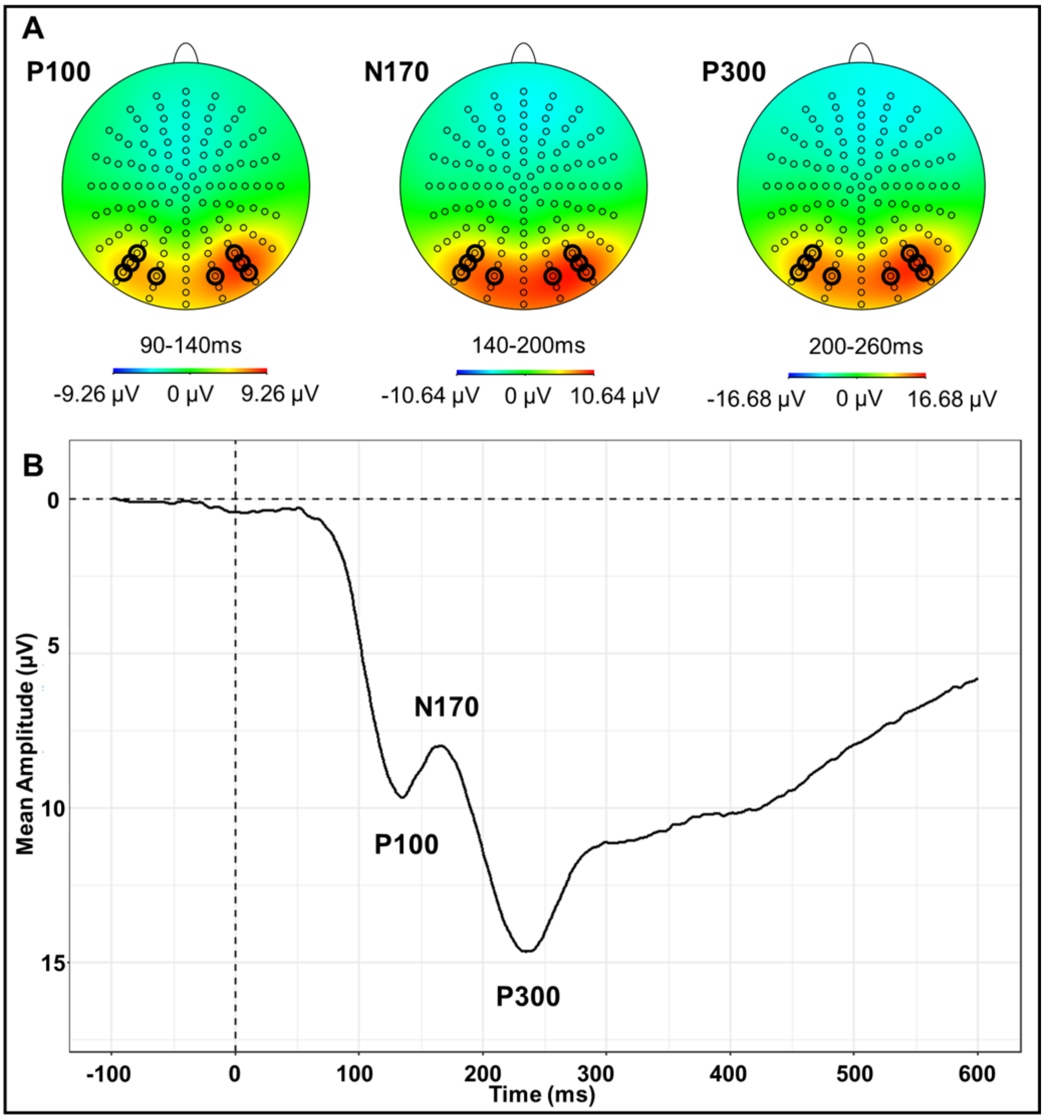


Figure 28. ***A.*** *Scalp map distribution corresponding to the two identified time-regions. Bold circles indicate the electrodes that have been averaged to obtain the mean amplitude of the P100, N170 and P300 respectively.* ***B.*** *Grand average ERP at the posterior site.*

1.5. Data analysis

Mean peak amplitudes / peak latencies were analysed as a function of grip typicality (grip-typical versus grip-atypical) and goal typicality (goal-typical versus goal-atypical), the two factors being repeated between participants. Grip activation was statistically tested through the main effect of grip typicality. Goal activation was statistically tested through the main effect of goal typicality. The integration of the two dimensions was statistically tested through the interaction between grip typicality and goal typicality factors.

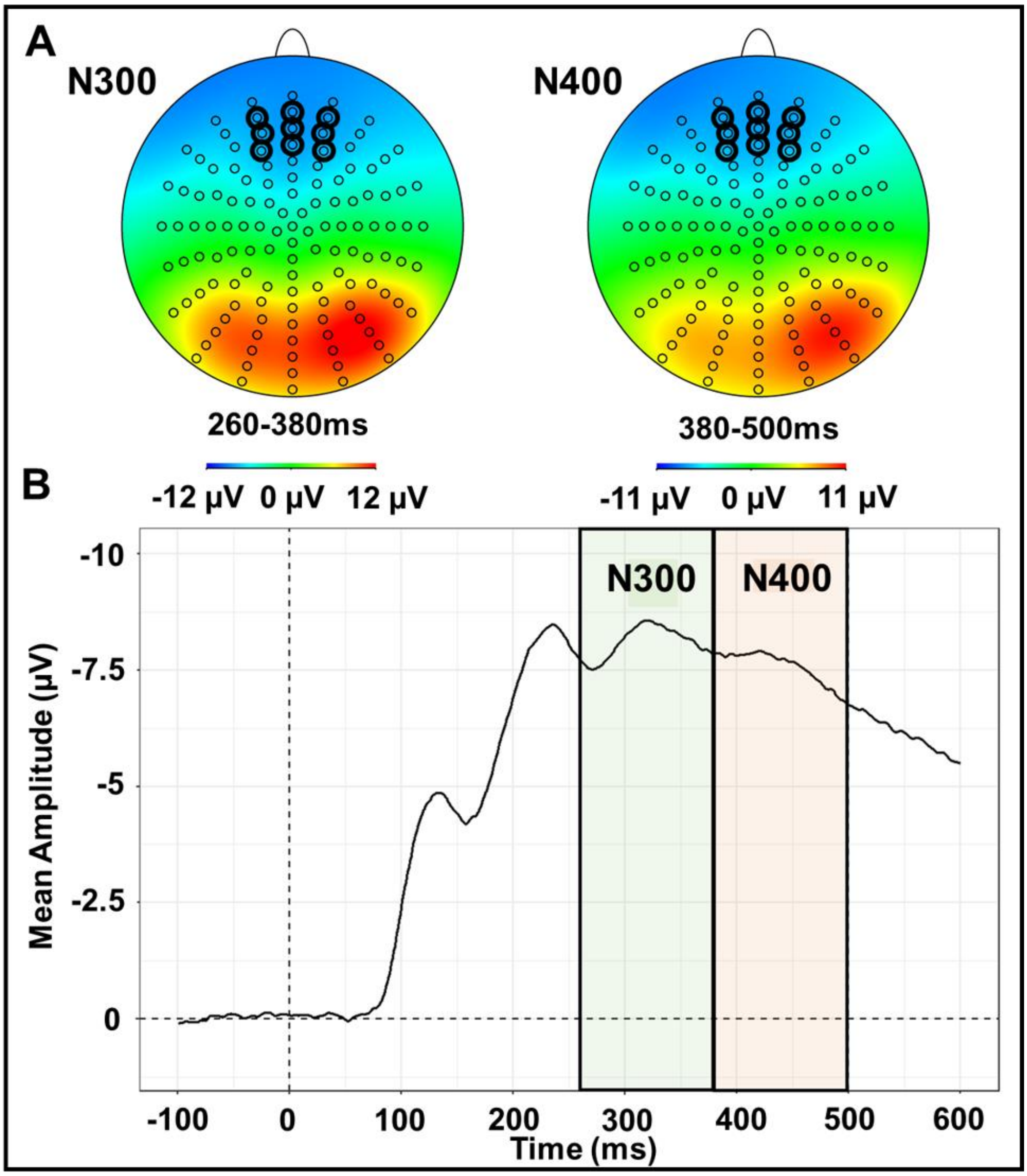


Figure 29. ***A.*** *Scalp map distribution corresponding to the two identified time-regions. Bold circles indicate the electrodes that have been averaged to obtain the mean amplitude of the N300 and N400 respectively.* ***B.*** *Grand average ERP at the anterior site.*

1.6. Statistical approach

Mean peak amplitudes / peak latencies were analysed using mixed-effect models to consider participants as a source of variation. Thereby, models included grip typicality, goal typicality, and the interaction between the two factors as fixed effects, and participants as random intercepts. Contrasting with the other experiments, only participants were included in the random structure; the EEG signal was averaged over the item that could not be considered as random parameters. Details can be found in the General Methodology section.

2. Results

Accuracy was not analysed in the EEG experiment considering the important working memory load involved in the response procedure. Overall mean accuracy on the 12% of trials associated with a response prompt was 73%. However, note that the identification of the different action pictures was verified during a pre-test (see General Methodology section). Participants were able to accurately identify an action as incorrect even when only one of the dimensions was atypical.

2.1. *P100, N170, and P300 on posterior site*

Analysis of mean peak amplitude of the P100 component revealed significant main effects of both grip typicality, $F(1,81) = 6.11$, $p_{\text{corrected}} = .046$, Westfall's $d = .09$, and goal-typicality, $F(1,81) = 15.27$, $p_{\text{corrected}} < .001$, Westfall’s $d = .15$. In both cases, the P100 was more positive for the typical dimension than for the atypical dimension (grip-atypical minus grip-typical = 0.36 µV, $SE = 0.14$; goal-atypical minus goal-typical = 0.56 µV, $SE = 0.14$). On the N170 component, main effects were significant for both grip typicality, $F(1,78.04) = 7.75$, $p_{\text{corrected}} = .020$, Westfall's $d = .08$, and goal typicality $F(1,79.01) = 10.32$, $p_{\text{corrected}} = .006$, Westfall's $d = .09$. In both cases, the N170 was more negative for the typical dimension than for the atypical dimension (grip-atypical minus grip-typical = -0.39 µV, $SE = 0.14$; goal-atypical minus goal-typical = -0.45 µV, $SE = 0.14$). The significance of the main effect of goal typicality on the P300 component did not survive the Bonferroni correction ($p_{\text{corrected}} = .077$). Results are displayed on Figure 30.

The analysis of the peak latencies did not reveal any significant effects for the P100, N170, or P300 components (all $p_{\text{uncorrected}} > .145$).

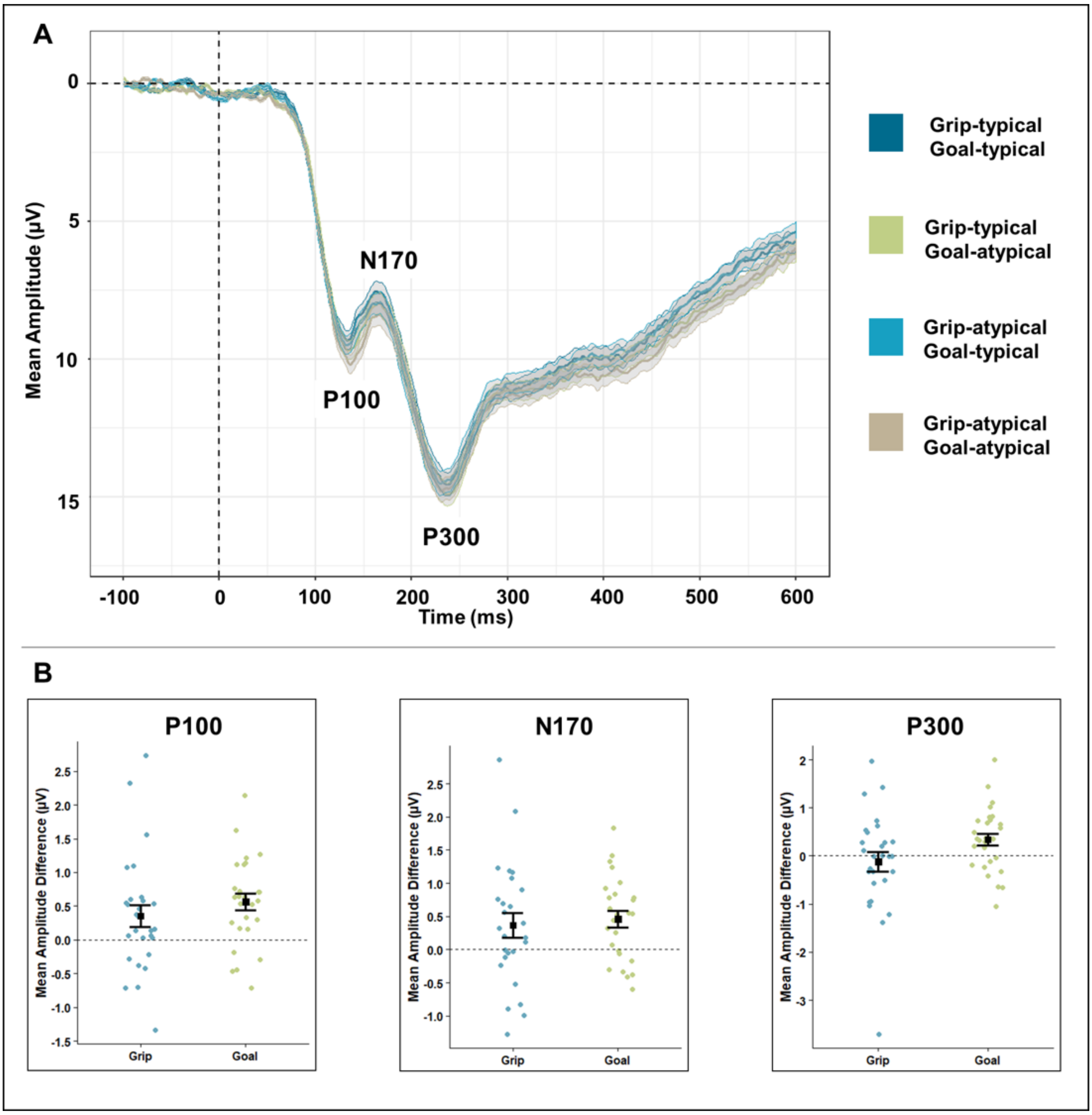


Figure 30. ***A.*** *ERP as a function of grip typicality and goal typicality at the posterior site. Ribbons represent standard errors.* ***B.*** *Individual mean estimates of the main effect of grip typicality (blue dots), main effect of goal typicality (green dots) for the P100, N170 and P300 components. Black dots represent overall mean and error bars represent standard errors.*

2.2. N300 and N400 on anterior site

Analysis of the mean peak amplitude of the N300 component revealed a main effect of goal typicality, $F(1,81) = 15.11$, $p_{\text{corrected}} < .001$, Westfall's $d = .15$, but no main effect of grip typicality, $F(1,81) = 0.93$, $p_{\text{uncorrected}} = .338$, Westfall's $d = .04$. Atypical goals were more negative than typical goals (goal-atypical minus goal-typical = 0.50 µV, *SE* = 0.13). Analysis of the mean peak amplitude of the N400 component revealed a main effect of grip typicality, $F(1,81) = 7.34$, $p_{\text{corrected}} = .016$, Westfall's $d = .11$, a main effect of goal typicality, $F(1,81) = 27.44$, $p_{\text{corrected}} < .001$, Westfall's $d = .21$, and a significant Grip typicality x Goal typicality interaction, $F(1,81) = 5.43$, $p_{\text{corrected}} = .045$, Westfall's $d = .18$. Interestingly, both main effects showed increased negativity for atypical conditions in comparison to typical conditions (grip-atypical minus grip-typical = 0.37 µV, *SE* = 0.14; goal-atypical minus goal-typical = 0.72 µV, *SE* = 0.14). Post-hoc tests indicated that the "grip-atypical goal-atypical" condition was more negative than the three other conditions, namely "grip-atypical goal-typical", $t(81) = -3.56$, $p = .002$, "grip-typical goal-atypical", $t(81) = -5.35$, $p < .001$, and "grip-typical goal-typical", $t(81) = -5.62$, $p < .001$, which were not significantly different from one another (all $p > .176$). Results are displayed on Figure 31.

3. Interim discussion

The current experiment examined the ERP components related to the identification of correct object-directed actions, while carefully dissociating the role of visual kinematics and non-motor goal-related information in action recognition. Object-directed actions could present grip and/or visual goal violations so that the two dimensions were manipulated independently. This design was appropriate to identify both the unique contribution of each dimension, and the integration of the two dimensions during the whole process of action recognition. Participants were asked to judge the overall correctness of object-directed actions such as pouring from a teapot or writing with a pencil, without any specific mention of the grip and visual goal dimensions.

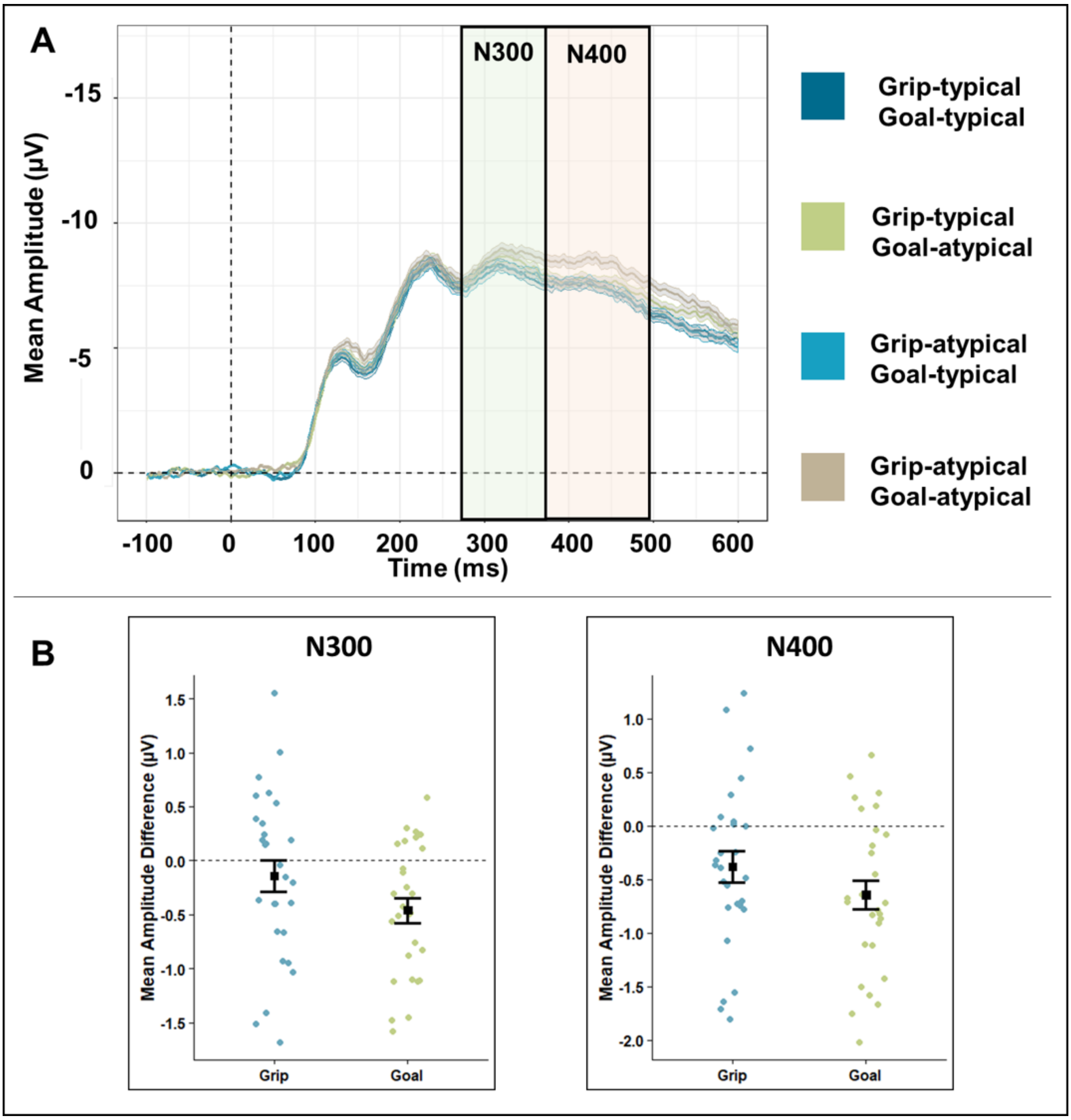


Figure 31. ***A.*** *ERP as a function of grip typicality and goal typicality at the anterior site. The green font represents the time-window of the N300 component. The orange font represents the time-window of the N400 component. Ribbons represent standard errors.* ***B.*** *Individual mean estimates of the main effect of grip typicality (blue dots), main effect of goal typicality (green dots) for the N300 and N400 components. Black dots represent overall mean and error bars represent standard errors.*

Overall, results showed that the identification of object-directed actions induced P100, N170, P300, N300, and N400 components. The polarity and topography of these components were very similar to those previously reported for the processing of visual stimuli (e.g., Bach et al., 2009; Bledowski et al., 2004; Chang et al., 2018; Peelen & Downing, 2007; Wamain et al., 2014 among others). The independent processing of both grip and visual goal was already visible on the P100/N170 components, reflecting the early independent perceptual processing of the two dimensions in the visual areas. Independent semantic processing of visual goals was later observed on the anterior N300 component, before the integration of grip and visual goal information on the N400 component. Reconciling inconsistent results from previous studies, we found that observers are able to decode both grip and visual goal information within the first hundred ms of action processing. By 400 ms of action processing, they consider both pieces of information in an integrated manner, as reflected by the expected N400 modulations according to grip and visual goal congruency. Importantly, immediately before this integration, there was a selective processing of visual goal visible on the N300, suggesting the prior activation of an action goal representation. The processing of grip configuration and visual goal was found to modulate both the P100 and the N170 components independently. The time windows of these two components are fairly congruent with previous “brain microstate” EEG studies on action processing. These studies identified periods of stability in the brain activity between roughly 0 and 120 ms over the visual cortex, and between roughly 120 and 200 ms over the posterior temporal and inferior parietal cortices (Avanzini et al., 2013; Ortigue et al., 2010). The P100 component is routinely used in ophthalmology to evaluate the integrity of the visual cortices (Kothari, Bokariya, Singh, & Singh, 2016; Sharma, Joshi, Singh, & Kumar, 2015), and the N170 has been shown to be particularly sensitive to inversion effects (comparison between upright / upside-down pictures) of body parts (Peelen & Downing, 2007; Rossion, Joyce, Cottrell, & Tarr, 2003). Both components have been linked to the activity of the visual cortices, the P100 being generated by the primary visual cortex, and the N170 by the associative visual cortices at the border of the temporal and parietal cortices (Peelen & Downing, 2007). They may then be considered as reflecting the perceptual processing of grip and visual goal information. Previous studies have mostly attributed such early brain modulations either to the processing of visual kinematics (Ortigue et al., 2010), or to the processing of non-motor goal-related information (Ortigue et al., 2009). The present experiment suggests that both dimensions contribute to these early modulations in an independent manner.

The N400 component was found to be modulated by the interaction between the grip configuration and the visual goal information. We revealed that the N400 generated when both the visual goal and the grip configuration were atypical was more negative than the one generated by any of the remaining combinations. In spite of very different design and stimuli, this pattern has been previously reported by Bach et al. (2009). In their study, participants had to evaluate whether two objects could be inserted to one another. Objects were sequentially presented and could have a correct functional relationship or not (screwdriver and screw vs. screwdriver and keyhole), or a correct motor relationship or not (horizontal screwdriver and horizontal screw). They found an N400 more negative when two semantic violations (i.e., in terms of functional relationship between two objects on the one hand, and in terms of motor relationship between two objects on the other hand) were present in the action (e.g., an horizontal screwdriver with a vertical orientation of a keyhole) than in any of the remaining combinations. In contrast, despite very similar design and stimuli, Chang et al. (2018) did not found the expected modulation of the N400 component as a function of action typicality. In their study, fully typical object-directed actions were compared to fully atypical object-directed actions (i.e., on both the non-motor goal-related and grip configuration dimension; e.g., an upside-down pencil with a power grasp). They argued that the absence of N400 sensitivity to semantic violations during action processing was due to participants' inability to "*rapidly match* *the semantic information conceptually*” (p. 7). Our results rule out such an interpretation, as both dimensions did modulate the N400 component. The action N400 component is thought to reflect the semantic integration between different action components. Accordingly, Amoruso et al. (2013) recently suggested that the action N400 could reflect, alongside with the linguistic N400, “*a common neurocognitive mechanism involved in the construction of meaning*” (p.1). Thus, our results indicate that, at such stages of action processing, visual kinematics and goal-related information are integrated and confronted to one another. The fact that partially typical actions were not different from fully typical actions may indicate that the cognitive system still consider those actions as being “plausible”, be it due to a typical grip (but an atypical goal), or to a typical goal (but an atypical grip).

The processing of the visual actions generated an N300 component which was modulated by the typicality of the visual goal information. What lies behind N300 components is unclear. It has sometimes been interpreted as an extension of the N400 components with similar sensitivity (Mudrik et al., 2010; Sitnikova et al., 2003), and occasionally as being a component clearly distinct from the N400 (West & Holcomb, 2002). Our results suggest that the N300 is, at least partially, independent from the N400 component. Chang et al. (2018) reported an N300 with a posterior distribution. Fully typical actions were more negative than fully atypical actions. Their N300 component was very similar (in terms of topography and functional sensitivity) to the “recognition potential” (Proverbio & Riva, 2009). Thus, these authors proposed that the increased negativity observed for fully typical actions reflected an easier access to visual semantic memory in comparison to fully atypical actions. However, this interpretation does not stand for our N300, as both the topography and functional sensitivity are not similar to their N300. One may argue that our N300 sensitivity to goal typicality could be simply driven by mere low-level perceptual differences between goal-typical and goal-atypical actions, as a similar sensitivity was found on the P100 / N170 components. We believe that this is relatively unlikely. If goal typicality effects had their roots in mere low-level perceptual differences, they should be observed on each component identified; but, they were not reliable on the P300 component. Hence, we argue that our N300 sensitivity to goal typicality is more likely to reflect the processing of visual goal information at a “semantic” stage of action processing, even though the exact cognitive mechanism remains to be identified.

Overall, these results are fairly congruent with the one obtained in the behavioural experiments: Visual kinematics and non-motor goal-related information are both processed early on during action recognition; non-motor goal-related information takes priority during the first steps of action recognition—reflected by the bigger effect sizes for the effect of non-motor goal related information here on the early ERP components, and by the isolated main effect of non-motor goal-related information on the N300 component—and the late reinvolvement of visual kinematics (reflected here on the N400 component). The present experiment further adds some information about the processes underlying these effects. Taken as a whole, the pattern of results observed is consistent with predictive approaches of action recognition (Donnarumma et al., 2017; Kilner, 2011; Kilner et al., 2007). An important challenge for predictive approaches is to specify the locus of the predictions observers make about action goals. From our results, it seems that the activation of action goal representations may arise at an early semantic stage after about 300 ms of action processing. First, visual kinematics and action goals may be perceptually processed by posterior areas (reflected by the P100 / N170 modulations). Then, non-motor goal-related information may be semantically processed and used to anticipate the visual kinematics the observer should expect from the actor (reflected by the N300 modulation). Finally, the expected and actual visual kinematics may be compared (reflected by the N400 modulation). The exact mechanisms underlying the generation of each component require further investigation, but our results raise ERP candidates that can directly address the place of visual kinematics and non-motor goal related information during action recognition.

PART 2: Frontal and parietal integration of visual kinematics and functional goals during object-directed action recognition: Evidence from TMS-priming?

This work was carried out in Alessio Avenanti's lab at the University of Bologna, Cesena Campus, benefiting from a funded 3-months internship. This work is currently in preparation for publication: ***Decroix, J.****, Borgomaneri, S., Kalénine, S. & Avenanti, A. (in prep). Frontal and parietal integration of visual kinematics and functional goals during object-directed action recognition: Evidence from TMS-priming?*

The EEG results provided some insight about the neural dynamics underlying action recognition. In particular, experimental evidence suggest that both visual kinematics and non-motor goal-related information are processed during the first hundred milliseconds of visual processing at the perceptual level. Subsequently, the processing of non-motor goal-related information may be deepened at a semantic stage of action visual processing. Importantly, this in-depth processing of non-motor goal-related information occurs independently from the processing of visual kinematics, and before the semantic integration of the two dimensions. One of the drawbacks of EEG lies in its poor spatial resolution, which makes difficult to draw any strong hypothesis about the brain areas involved in such processes. Yet, action recognition is thought to be underpinned by an extended brain network, usually referred to as the action observation network (AON; Avenanti, Candidi, et al., 2013; Caspers et al., 2010; Urgesi et al., 2014). Sneaking into the neural dynamics of the action observation network may be challenging, as *f*MRI provides data about the brain area involved but not about their dynamic involvement, whereas EEG provides data about the neural dynamics but not the brain areas involved[17]. One accessible strategy we selected to fulfil both spatial and temporal requirement

---

[17] This claim is quite clear cut, as some strategies do exist to compensate for the respective low temporal resolution and low spatial resolution of *f*MRI and EEG. Nonetheless,

is coupling transcranial magnetic stimulation (TMS) with our priming paradigm. TMS allows us to target a particular brain area, whereas priming allows us to set up action recognition at a particular stage of action visual processing. Coupling TMS and priming further allows to target particular neuronal population within the stimulated brain area, while knowing, to some extent, at what stage of action visual processing participants are. An additional critical advantage of TMS is that it allows to draw causal relationships between the stimulated brain area and the cognitive processes under investigation. EEG and *f*MRI merely record the activity of the brain while participants are performing some tasks; with TMS we modulate the activity of the brain and evaluate whether this modulation will affect the investigated cognitive processes. In summary, in EEG and *f*MRI, the modulations of the brain activity are observed; with TMS they are induced.

Within the AON, the frontoparietal "motor" network is one of the nodes that is of the greatest interest for researchers. Indeed, the frontoparietal network is thought to be the human brain regions where mirror neurons are located. As such, this network has been identified as an important node of the brain when visually recognising others' actions (Avenanti, Candidi, et al., 2013; Caspers et al., 2010; Molenberghs et al., 2012; Ocampo & Kritikos, 2011; Spunt & Lieberman, 2012; Thill et al., 2013; Urgesi et al., 2014; van Elk et al., 2014a; Van Overwalle & Baetens, 2009). Yet, the functions sub-served by this network are still debated. On the one hand, the inferior frontal cortex (IFG) has been causally involved in the fine-grained processing of visual kinematic parameters (Avenanti, Paracampo, Annella, Tidoni, & Aglioti, 2017; Candidi, Urgesi, Ionta, & Aglioti, 2008; L. Cattaneo, 2010; L. Cattaneo et al., 2011; Jacquet &

---

most of these strategies do not affect how the signal is recorded, but rather how the signal is statistically analysed. Although we believe that such strategies should be used as it increases the value of EEG and *f*MRI data, such strategies should be carefully interpreted, and may be better used to draw hypotheses that will be later evaluated with more suitable methods.

Avenanti, 2015; Koch et al., 2010; Michael et al., 2014; Pobric & Hamilton, 2006; Tidoni, Borgomaneri, di Pellegrino, & Avenanti, 2013; Urgesi, Candidi, Ionta, & Aglioti, 2007); on the other hand, the IFG has been advocated to play a key role in the processing of action goals independently from their underlying visual kinematics (L. Cattaneo, Sandrini, & Schwarzbach, 2010; Jacquet & Avenanti, 2015). Regarding the inferior parietal lobule (IPL) and despite the widely accepted idea of its involvement in action observation, only few studies have tried to directly establish the critical role of the IPL in action recognition (L. Cattaneo et al., 2010; Jacquet & Avenanti, 2015). The integrity of the IPL was found to be crucial for accurate action goal recognition by Cattaneo et al. (2010), but not by Jacquet and Avenanti (2015). Some studies did report a causal role of the IPL in action recognition, but through its connection with the primary motor cortex (Feurra et al., 2011; Koch et al., 2010), or through its connexion with the primary somatosensorial cortex (Valchev, Gazzola, Avenanti, & Keysers, 2016). Impaired IPL in brain-damaged patients seems to affect the processing of visual kinematics (Kalénine, Buxbaum, & Coslett, 2010; Kalénine et al., 2013) and the anticipation of others' actions (i.e., anticipating the moment at which one will start to move; Fontana et al., 2012) rather than the processing of action goals. Overall, the precise role of both the IFG and IPL in action understanding, and in particular their relative involvement in processing goal versus visual kinematics, remains strongly debated.

The seemingly contradictory results regarding the involvement of the frontoparietal network in action recognition are actually consistent with the assumptions of the predictive approaches. This network is supposed to be an integrative network where predicted visual kinematics arising from the predicted action goal are compared to the actual visual kinematics performed by the actor (Kilner, 2011; Lingnau & Downing, 2015). It should then not be surprising to find, within the same network, some neuronal populations dedicated to the processing of fine-grained visual kinematics, and others dedicated to the processing of more abstract action goals. Moreover, predictive approaches may also expect neuronal populations involved in the integration of the actual visual kinematics and non-motor goal-related information (e.g., prediction). Consistent with this hypothesis, a few *f*MRI studies have highlighted brain regions within the frontoparietal network that are sensitive to the congruency between non-motor goal-related information and visual kinematics (Schubotz, Wurm, Wittmann, & von Cramon, 2014; Wurm & Lingnau, 2015). For example, when observing object-directed actions, Schubotz et al. (2014) found that the IPL was involved in retrieving the motor information associated with objects (i.e., brain activity in the IPL increased with the number of possible actions associated with the object), whereas the IFG was associated with the integration between the actual visual kinematics and the object typical manipulation (i.e., brain activity in the IFG activity increased when the action was object-compatible, such as sharpening a pencil with a sharpener, compared to when the action was object-incompatible, such as pretending sharpening a pencil with an orange and a knife). Nonetheless, as far as we know, there is no causal evidence demonstrating the existence of neuronal populations within the frontoparietal network involved in the separate processing of different action components, or their integration during the recognition of others' actions.

Although they may represent different stages of action processing, the IFG and the IPL seem to be involved in both visual kinematics and non-motor goal-related information processing. As a matter of fact, the various results reported in the literature may be explained by the inability of most of the reviewed neuroanatomical studies to provide reliable temporal information. The visual recognition of others' actions can nonetheless be segregated into different stages of action processing that are temporally distributed (see Chapter 4; Amoruso et al., 2016, 2014; Ortigue et al., 2010, 2009), with back and forth communications between posterior and anterior brain areas (Avanzini et al., 2013). The theoretical importance of taking into account timing has also been largely acknowledged (Catmur, 2015; Grafton & Hamilton, 2007; Naish et al., 2014). It is then highly possible that the lack of temporal information may lead us to miss important phenomena at some stages of action processing. This proposition is particularly relevant for the IPL as this brain region is thought to be an intermediary brain structure between the IFG and the visual areas (including the superior temporal sulcus and the lateral occipitotemporal cortex; Kilner, 2011; Kilner et al., 2007; Lingnau & Downing, 2015). It may then be difficult to target this area when evaluating the overall performance of participants (i.e., the final result of action processing). Hence, the present study aimed at overcoming the aforementioned limitations by investigating whether neuronal populations dedicated to the processing of the visual kinematics, the processing of non-motor goal-related information, and the integration of the two dimensions could be detected in the IFG and in the IPL, while targeting a particular stage of action processing.

When applied at the onset of a primed target over a brain area involved in the generation of the priming effect, TMS has been found to diminish or even reverse the behavioural effects of priming (Silvanto & Cattaneo, 2017; Silvanto, Muggleton, & Walsh, 2008). Behaviourally, a priming effect reflects the modification of the processing of a target stimulus due to the prior presentation of another stimulus, the prime (see Chapter 3 & 4). Therefore, the principle of TMS-priming paradigms is to disturb the processing of the prime, and thus to reduce its influence on the processing of the target (see Figure 32).

Additionally, the effects of TMS on priming have been attributed to the differential effects that TMS engendered on neurons activated by the prime in comparison to neurons that are not. This phenomenon, known as the state-dependency effect of TMS, has shed new lights on the effect of TMS, and is thought to provide deeper insight on the results provided by TMS experiments in comparison to the more classical “virtual lesion” paradigms (see Lang et al., 2004; Siebner, Hartwigsen, Kassuba, & Rothwell, 2009; Siebner et al., 2004; Silvanto & Cattaneo, 2017; Silvanto et al., 2008). In the virtual-lesion paradigms, it is not possible to know which neuronal populations are stimulated within a brain area (Pascual-Leone, Bartres-Faz, & Keenan, 1999; Ziemann, 2010). In state-dependency paradigms, the brain is tuned to the cognitive operation of interest before the TMS stimulation (using priming for example); As a consequence, neurons involved in that cognitive operation are not in the same state that neurons that are not—there is “primed” neurons and “non-primed” neurons. Usually, tuning the brain into a cognitive operation has an effect on a subsequent task involving that cognitive operation—priming a photograph with the presentation of the same photograph before will facilitate its processing. If TMS is applied over a brain area in which such neurons are present (i.e., the “primed” neurons), then the induced behavioural effects will be affected by the TMS stimulation, decreasing, or even reversing, the facilitative effect of presenting the same photograph. This phenomenon allows the perturbation of specific neural populations within the stimulated areas, which makes TMS priming a very interesting tool to independently assess different parts of the largely connected frontoparietal network.

In the present study, pictures of object-directed actions were briefly primed (220 ms of prime duration, followed by 66 ms mask) by an action picture sharing (a) only the same visual goal, (b) only the same grip configuration, (c) both the same visual goal and the same grip, or (d) neither the same grip nor the same visual goal. TMS was applied over the IFG, the IPL or over the vertex (sham condition) at the target onset. In Chapter 4, we found that 220 ms of prime duration was sufficient to elicit priming effects from both repetition of the same grip and the same visual goal separately; this demonstrate the activation of both grip and visual goal dimensions at that stage of action visual processing. Importantly, behavioural results further established the integration of the two dimensions at 220 ms. Action recognition was faster following primes with both similar or both dissimilar grip and visual goal dimensions, than following primes with either similar grip or similar visual goal—a phenomenon known as the partial-repetition cost (Hommel, 2004). Thus, decrease or reversal of the grip and/or visual goal priming effects after single-pulse TMS would reveal the presence of two distinct neuronal populations dedicated to the grip and visual goal processing in the stimulated region. In addition, decrease or reversal of the partial-repetition cost would reveal the presence of neuronal populations dedicated to the integration of the two dimensions in the stimulated region. As discussed above, these effects could be detected in either the IFG or the IPL, as experimental evidence and theoretical accounts made no clear-cut hypothesis regarding the role of the two nodes of the frontoparietal network in processing grip and visual goal during action recognition.

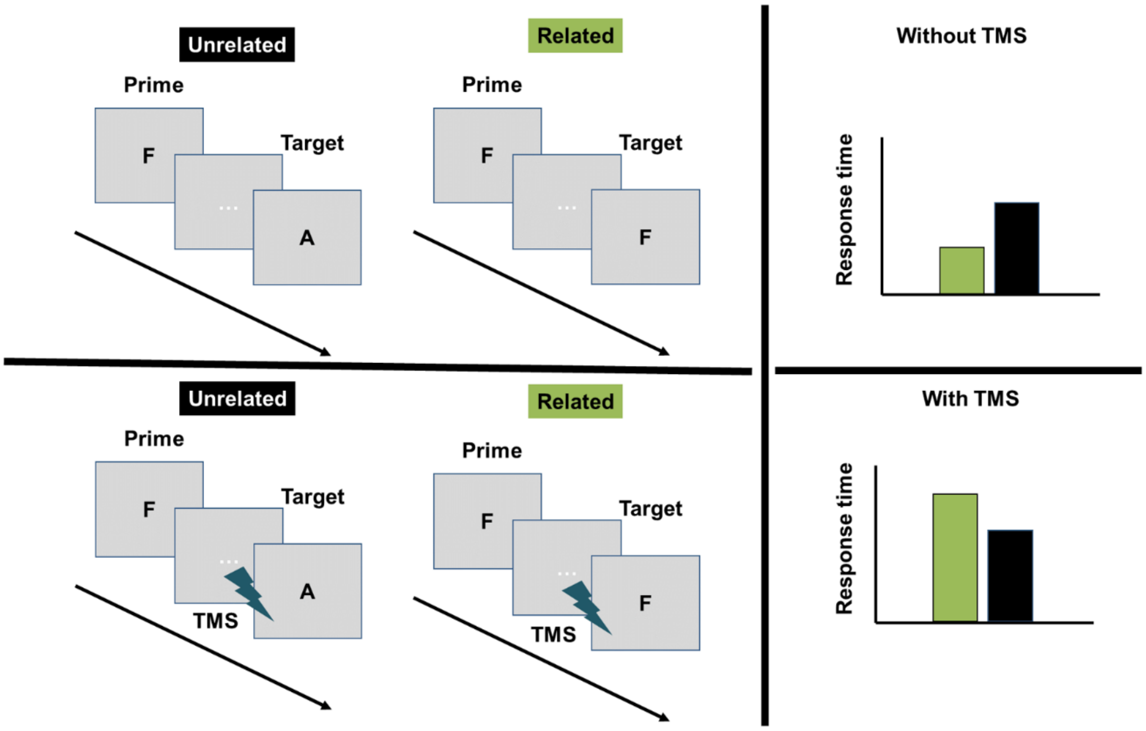


Figure 32. Schematic representation of a TMS-priming paradigm. Without TMS, related prime-target pairs (in green) will be associated with faster response times on target judgement (e.g., same or different as the prime?) than unrelated prime-target pairs (in black). When TMS is applied at the target onset, related prime-target pairs will be associated with slower response times than unrelated prime-target pairs. TMS disturb the information preactivated by the prime.

1. Methods

1.1. Participants

Eighteen participants (6 males, M_{age} = 24, age range: 21–29) were recruited in the present study. All were classified as right-handed according the Edinburgh Handedness Inventory (M_{EHI} = 0.83, range = 0.37–1; Oldfield, 1971), reported normal or corrected-to-normal vision and no history of neurological or psychiatric diseases. They all provided written informed consent. The experiment was carried out in accordance with the ethical guidelines of the Bologna University's Department of Psychology and with the declaration of Helsinki (1964, revised in 2013).

1.2. Stimuli

The same set of stimuli than in the EEG experiment was used (set 2). Briefly, for each of the twenty reference objects, five coloured 1024 x 683 pixels photographs were taken, all displaying a hand and a tool-object. Four out of the five pictures of the set presented hand-on-object actions. The remaining photograph corresponded to a no-action picture showing the hand and the object without any interaction between them. Here again, for each reference object, actions could be typical or not along the grip dimension and/or the visual goal dimension. Detailed description can be found in the General Methodology section.

1.3. Design and procedure

Photographs were implemented in a priming paradigm similar to the one used in the behavioural priming experiments (Chapter 4). The four types of photographs could be presented as prime. Only the full correct actions showing both correct grip and correct goal ("correct targets"), or the full incorrect actions showing both incorrect grip and incorrect goal ("incorrect targets”) could be presented as targets. This resulted in four prime-target relations for each reference object: “grip-similar, goal-similar”; “grip-similar, goal-dissimilar”; “grip-dissimilar, goal-dissimilar”; “grip-dissimilar, goal-dissimilar”.

There was a total of 2 grip similarity (grip-similar; grip-dissimilar) x 2 goal similarity (goal-similar; goal-dissimilar) x 2 response type (yes = “correct target”; no = “incorrect target”) x 20 objects = 160 trials. Each trial was repeated in three TMS blocks corresponding to the stimulated areas (IFG; IPL) and sham. Each TMS block was divided in two blocks of 80 trials, resulting in six blocks of eight minutes each. The order of blocks, as well as the order of trials within blocks were fully randomised. Overall there was 80 trials x 6 blocks = 480 trials. The experiment was conducted with E-Prime V2.0.10.353 software (Psychology Software Tools, Pittsburgh, PA).

Each trial started with a fixation cross for 5000 ms, followed by the prime for 220 ms, and by a pixelated black and white mask for 66 ms, and finally the target that was displayed until the participant’s response (see Figure 33). As in previous studies, TMS was delivered at target onset (L. Cattaneo, 2010; Z. Cattaneo, Rota, Vecchi, & Silvanto, 2008). The prime duration of 220 ms was chosen based on our previous priming study with a similar design, in which 220 ms prime + 66ms mask were sufficient to trigger both grip and goal priming effects (see Chapter 4). Participants were required to judge as fast and accurately as possible whether the target action was correct or not according to the typical use of the object (forced choice). They were required to press “c” or “b” on a keyboard using their left hand. The correct / incorrect pattern was counterbalanced between participants. Response times (RT) and correct responses (accuracy) were recorded. Participants first performed a training session with 12 representative trials on which they received feedback. The training session involved three additional objects that were not included in the experimental session. The experimental session was equivalent to the training session but without feedback. Participants could take breaks between the blocks.

At the beginning of the experimental session, participants were required to fulfil the consent form and the EHI. The TMS set-up was then calibrated (see below). The training session and the six experimental blocks were performed. At the end of the experiment, participants were debriefed. Overall the experiment lasted ~2h.

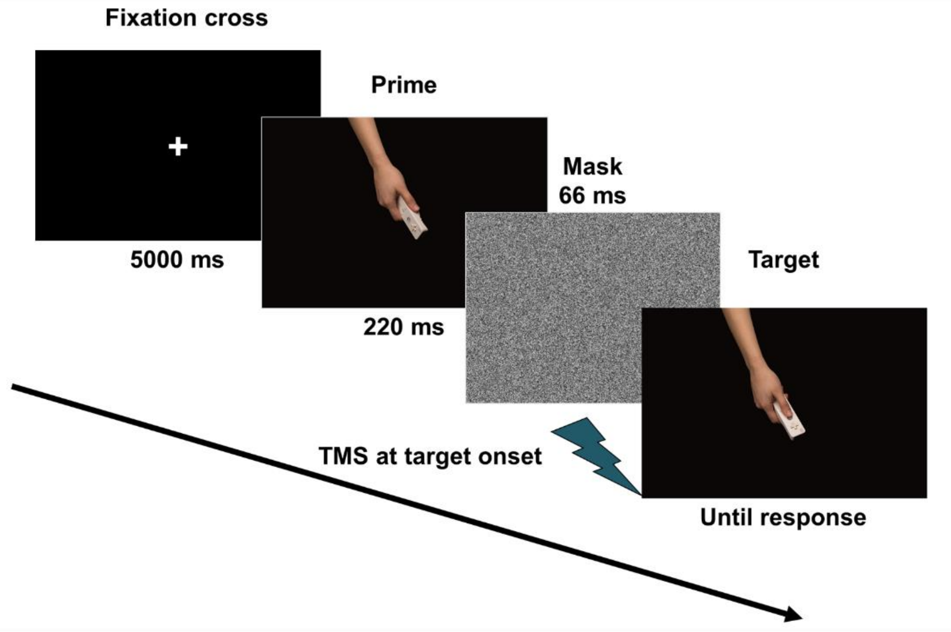


Figure 33. Trial procedure of the TMS-priming paradigm

1.4. Transcranial Magnetic Stimulation (TMS)

TMS pulses were delivered with a figure-of-eight coil (70 mm) and a Magstim Rapid2 stimulator (Magstim, Whitland, Dyfed, UK). The individual resting motor threshold (rMT) of each participant was identified as the minimal stimulation intensity producing motor evoked potentials (MEPs) of a minimum amplitude of 50 µV in the right first dorsal interosseous (FDI) muscle, with 50% probability (Rossini et al., 2015). MEPs were recorded by means of a Biopac MP 35 electromyograph (Biopac Systems, Inc., USA). EMG signals were band-pass filtered (30–1000 Hz) and digitised (sampling rate at 5 kHz). Pairs of silver / silver chloride surface electrodes were placed over the right FDI muscle using a belly / tendon montage. The intensity of stimulation used during the experiments was set at 110% of the individual rMT.

Prior to the experimental session, coil position was identified on each participant's scalp using the SofTaxic Navigator system (EMS, Bologna, Italy). In a first step, skull landmarks (nasion, inion, and two preauricular points) and ~60 points providing a uniform representation of the scalp were digitised by means of a Polaris Vicra Optical Tracking System (Northern Digital, Inc., Waterloo, ON, Canada). Coordinates in Talairach space were automatically estimated by the SofTaxic Navigator from an MRI-constructed stereotaxic template. Then, the scalp sites corresponding to IFG and IPL in the left hemisphere were selected (see Figure 34). Scalp sites corresponding to IFG and IPL were selected in the left hemisphere using the coordinates of the peaks of activation reported for the observation of hand-object pictures in the meta-analysis of Caspers et al. (2010). The IFG was targeted in the anterior–ventral part of the precentral gyrus (ventral premotor cortex) at the border with the pars opercularis of the inferior frontal cortex (coordinates: $x = -51$, $y = 7$, $z = 30$), corresponding to Brodmann's area 6/44 (Avenanti, Annela, & Serino, 2012; Avenanti, Annella, Candidi, Urgesi, & Aglioti, 2013; Caspers et al., 2010; Urgesi et al., 2007; Van Overwalle & Baetens, 2009). The selected frontal area is thought to be just one synapse away from the motor cortex as suggested by dual-coil TMS (L. Cattaneo & Barchiesi, 2011; Davare, Montague, Olivier, Rothwell, & Lemon, 2009), and thus can mainly be considered a premotor region. The IPL region was targeted in the anterior sector of the intraparietal sulcus ($x = -58$, $y = -24$, $z = 36$, corresponding to Brodmann's area 40; Avenanti et al., 2012; Caspers et al., 2010; Van Overwalle & Baetens, 2009). The IFG and IPL scalp sites were marked on the bathing cap with a pen. Then, the neuronavigation system was used to estimate the projections of the scalp sites on the brain surface (IFG M surface MNI coordinates ± SD: $x = -52 \pm 2$, $y = 6 \pm 1$, $z = 30 \pm 2$; IPL: $x = -58 \pm 3$, $y = -24 \pm 1$, $z = 35 \pm 1$). Stimulation of IFG and IPL was carried out by placing the coil tangentially over the marked scalp sites. Sham stimulation was performed by placing the coil tilted at 90° over the vertex, so that no electrical current was induced in the brain.

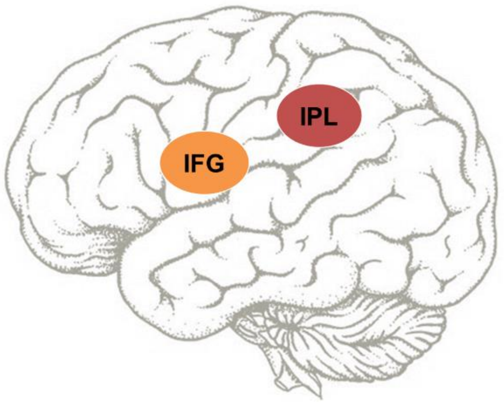


Figure 34. Brain areas stimulated. IFG: Inferior Frontal Gyrus; IPL: Inferior Parietal Lobule

1.5. *Data pre-processing*

No participant reported discomfort following the TMS experiment. The task was relatively easy to perform (M_{accuracy} = 94.5%, from 79.6% to 99.6%). Accuracy scores were equally distributed across condition, as evaluated by a chi-square test for independence, χ^2_{14} = 0.224, $p > .99$. First, errors and RT superior to 1500 ms and inferior to 150 ms were considered as conceptual outliers (i.e., data not related to the processes of interest), and were removed (5.65% of the data). Remaining RTs superior or inferior to five median absolute deviation of the median (Leys, Ley, Klein, Bernard, & Licata, 2013) computed for each condition and each participant were considered as statistical outliers (i.e., data not representative of the RT distribution), and were removed (1.24% of the data). Data is showed on Figure 35.

1.6. *Data analysis using mixed-effect models*

As detailed in the General Methodology, mixed-effect linear model approaches were used to take both individual and items as sources of variation and to consider their possible interaction with the factors of interest (e.g., a given participant or a given item, may be more sensitive to IFG stimulation than another one).

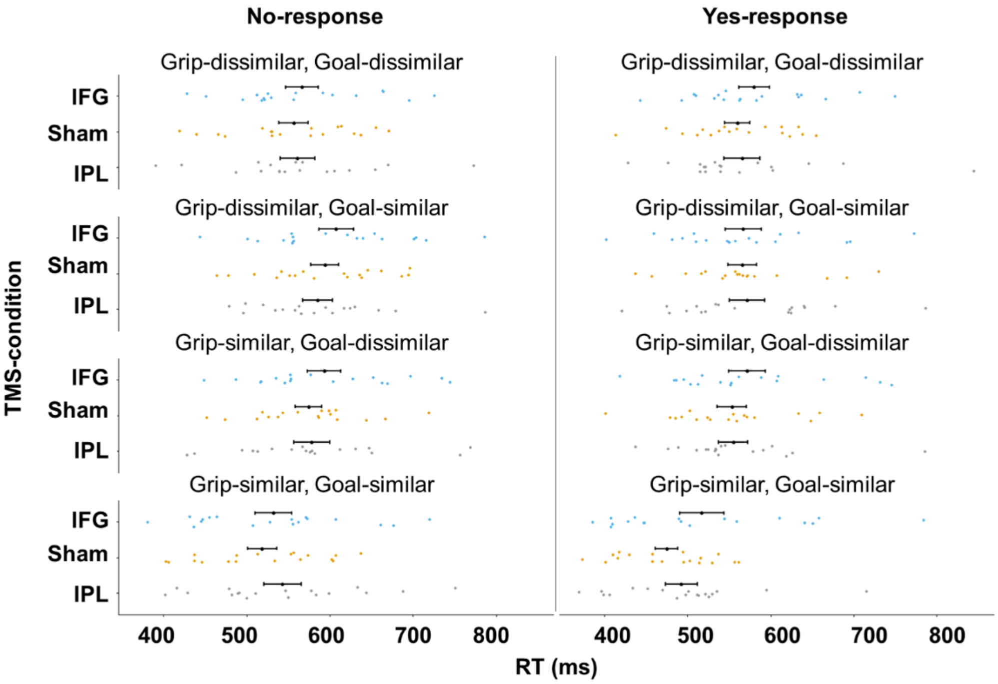


Figure 35. Individual mean RT in milliseconds as a function of TMS-condition (IPL, IFG or Sham), Response-type (yes or no), Grip-similarity ("Grip-similar" or "Grip-dissimilar") and Goal-similarity ("Goal-similar" or "Goal-dissimilar"). Error bars represent standard errors. Each dot represents an individual mean. Black dots represent the mean averaged over the participants.

The full model used to analyse the RTs included grip similarity ("grip-similar", "grip-dissimilar"), goal similarity ("goal-similar", "goal-dissimilar"), response type ("yes", "no"), TMS condition ("IFG", "IPL", "Sham"), and their respective interactions as fixed effects; participants and items as random intercepts; grip similarity, goal similarity, TMS condition, and response type as random slopes for participants; and finally, goal similarity and TMS condition as random slopes for items. For each yes and no models, grip similarity ("grip-similar", "grip-dissimilar"), goal similarity ("goal-similar", "goal-dissimilar"), response type ("yes", "no"), TMS condition ("IFG", "IPL", "Sham"), and their respective interactions were included as fixed effects; the random structures included both participants and items as random intercepts; and grip similarity, goal similarity, and TMS condition as random slopes for participants.

2. Results

RTs were analysed as a function of grip similarity (“grip-similar”, “grip-dissimilar”), goal similarity (“goal-similar”, “goal-dissimilar”), response type (“yes”, “no”), and TMS condition (“IFG”, “IPL”, “Sham”) using a mixed-model approach. Statistically, the main effects of grip similarity (“grip-similar” versus “grip-dissimilar”) and goal similarity (“goal-similar” versus “goal-dissimilar”) evaluated grip and goal priming effects, respectively. The Grip similarity x Goal similarity interaction evaluated the partial-repetition cost effect (i.e., the cost induced by sharing the same grip but not the same goal or sharing the same goal but not the same grip, in comparison to sharing both the same grip and the same goal and sharing neither the same grip nor the same goal). It was considered as a statistical marker of grip and goal integration. Therefore, we evaluated whether in comparison to sham, stimulating the IPL on the one hand, and the IFG on the other hand, would affect the Grip similarity x Goal similarity interaction. In the absence of influence on the interaction, the effect of IPL and IFG stimulation of grip similarity main effect and goal similarity main effect would be assessed. The factor response type was not of primary interest, but was first included in the analysis since yes / no response effects on action judgements have been previously reported (see Chapter 4; see also Yoon et al., 2010).

A first model reveals a significant Grip similarity x Goal similarity x TMS condition x Response type interaction, $F(2,7874.8) = 5.56$, $p = .004$. Analyses were then carried out separately for yes and no responses. The Grip similarity x Goal similarity x TMS condition interaction was significant for both yes response, $F(2,3904.9) = 3.63$, $p = .026$, and no response, $F(2,3937.3) = 4.45$, $p = .012$, respectively. For yes response, the cost induced by the repetition of one action dimension (i.e., “grip-similar, goal-dissimilar” or “grip-dissimilar, goal-similar”) in comparison to the repetition of either both or no action dimensions (i.e., “grip-similar, goal-similar" or "grip-dissimilar, goal-dissimilar") was reduced of 44 ms after IFG stimulation (estimate = -43.85 ms, *SE* = 16.66, *t*(3904.98) = -2.63, $p_{\text{corrected}}$ = .016, Westfall's *d* = .31) in comparison to sham. There were no significant differences between IPL stimulation and sham (estimate = -13.44 ms, *SE* = 16.63, *t*(3905.34) = -0.81, $p_{\text{uncorrected}}$ = .419, Westfall's *d* = .09). Interestingly, the inverse pattern was observed for no response: in comparison to sham, the partial-repetition cost was decreased after IPL stimulation (estimate = -37.12 ms, *SE* = 15.69, *t*(3938.35) = - 2.36, $p_{\text{corrected}}$ = .036, Westfall's *d* = .28), but not after IFG stimulation (estimate = 6.17 ms, SE = 15.63, *t*(3937.06) = 0.39, $p_{\text{uncorrected}}$ = .693, Westfall's *d* = .05). Results are displayed on Figure 36. No other effects related to TMS condition were significant.

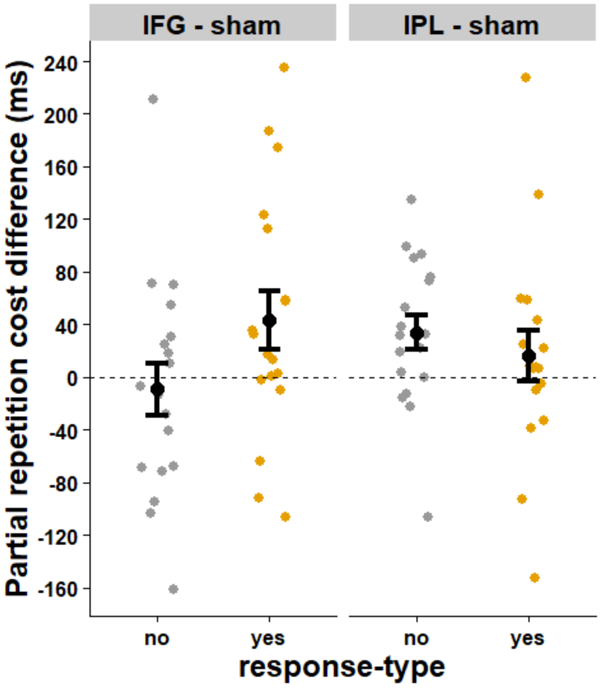


Figure 36. Individual mean difference in terms of partial repetition cost (i.e., difference between situations in which only one dimension is shared between the prime and the target and situations in which both grip and goal, or neither of them, are shared between the prime and the target) in milliseconds between IFG and sham on the right, and between IPL and sham on the left according to response type when both Grip and Goal were similar between the prime and the target versus when only one of the two dimension. Each dot represents an individual mean. Black dots represent the mean averaged over the participants. Error bars represent standard error.

3. Interim discussion

The present study aimed at investigating the critical role of the IFG and the IPL in the processing of grip configuration and visual goal information when visually recognising actions performed by others. More precisely, TMS priming was used to evaluate whether specialised neuronal populations related to the processing of (a) grip configuration, (b) visual goal information, or (c) the integration of the two dimensions, were critically involved in the left IFG and / or in the left IPL when visually identifying others' actions. Single-pulse TMS was delivered at the onset of a target picture displaying an object-directed action that was previously primed by an another action picture sharing (a) only the same grip configuration, (b) only the same visual goal, (c) both the same grip configuration and the same visual goal, or (d) neither the same grip configuration nor the same visual goal. The prime was displayed for 220 ms to ensure that both grip and visual goal dimensions had been activated (see Chapter 4). This paradigm allowed us to assess the critical involvement of IFG and IPL in processing two distinct important dimensions of actions, specifically the grip configuration and the visual goal dimension, while controlling the time course of action processing. Action recognition was found affected by both left IFG and left IPL stimulation in comparison to sham. This may suggest a non-specific effect of TMS, namely the fact that the stimulation may affect brain activity overall, independently from the experimental manipulations. Yet, IFG stimulation influenced the processing of correct action targets (i.e., yes response), whereas IPL stimulation impacted the processing of incorrect action target (i.e., no response), ruling out the possibility of a non-specific effect of TMS. Overall, results highlight the presence of specialised neuronal populations in both the IFG and the IPL dedicated to the integration of grip configuration and visual goal dimensions.

Although the exact effect of TMS remains debated (Silvanto & Cattaneo, 2017), reduction of priming effects after TMS stimulation at target onset has been attributed to the presence of neuronal populations dedicated to the generation of priming effects (Z. Cattaneo et al., 2008). TMS priming has been used previously to identify regions critically involved in the generation of facilitative priming effects (Ambrus, Dotzer, Schweinberger, & Kovács, 2017; Z. Cattaneo, Devlin, Salvini, Vecchi, & Silvanto, 2010; Gilaie-Dotan, Silvanto, Schwarzkopf, & Rees, 2010), including studies interested in the processing of others' actions (L. Cattaneo, 2010). In the present experiment, single-pulse TMS affected the relation between the prime and the target after IFG and IPL stimulation in comparison to sham. Interestingly, this effect of TMS did not modulate a repetitive, facilitative, priming effect, but a cost induced by the prime on the target processing, namely a reduction of the partial-repetition cost. The mechanisms underlying priming effects are complex, and the prior exposition of a given information does not always facilitate its subsequent processing (D'Angelo, Thomson, Tipper, & Milliken, 2016; Frings, Schneider, & Fox, 2015; Hommel, 2004). Several sources of evidence suggest that TMS has a different effect on neurons that are activated in comparison to neurons that are not. In particular, TMS is thought to increase the activity of neurons that are not activated (Silvanto & Cattaneo, 2017); in a priming paradigm, this leads to either a facilitation of the non-primed dimension (Silvanto & Cattaneo, 2017), or an overall increase in the noise level (Miniussi, Harris, & Ruzzoli, 2013), which abolished the advantage of the primed dimension. If the advantage of the primed dimension at the neural level is detrimental at the behavioural level, the diminution of this advantage (regardless of the specific mechanism involved) would lead to an improvement of the behavioural performance. This possibility has been already acknowledged from a theoretical point of view (Silvanto & Cattaneo, 2017; Silvanto et al., 2008), and supported by a few TMS experiments (Kehrer et al., 2015). The present data provide additional empirical evidence of cost reduction in TMS priming. By showing that the partial-repetition cost between action dimensions is reduced after IFG and IPL stimulation in comparison to sham, this experiment demonstrates the presence of neuronal populations in the IFG and in the IPL involved in the integration of visual kinematics and non-motor goal-related information, as hypothesised by predictive approaches of action recognition (Bach et al., 2014; Kilner, 2011).

The fact that the effect of IFG and IPL stimulation on the partial-repetition cost was different for correct targets (i.e., yes response) and incorrect targets (i.e., no response) should not lead to the conclusion that grip configurations and visual goal information are integrated in the IFG for correct targets and in the IPL for incorrect targets. Instead, this modulation is most likely reflecting temporal differences in the processing of correct and incorrect actions. This interpretation is first supported by faster response times for correct actions in contrast to incorrect actions. Therefore, correct and incorrect action targets are necessarily differently affected by the prime. Furthermore, in Chapter 4, we have found that the emergence of priming effects required a shorter prime duration for correct targets compared to incorrect targets. Different time courses may then underlie the processing of correct versus incorrect actions. Effects of response type on action processing are inconsistent and not always reported. The lack of both theoretical consideration and experimental data on such effects makes their interpretation hazardous. Nonetheless, we suggest that the possibility of different time courses may be an interesting hypothesis to start to uncover these effects.

SYNTHESIS

The neurophysiological bases of action recognition have been extensively investigated, especially by the *f*MRI literature (Caspers et al., 2010; Molenberghs et al., 2012; Van Overwalle & Baetens, 2009). These studies have led to the identification of an extensive network known as the AON, which includes the frontoparietal areas and the lateral occipitotemporal areas. Yet, the organisation of this network, and more precisely when and how each of the brain areas of the AON are involved during action recognition, remained largely discussed (Kilner & Frith, 2008; Thioux et al., 2008; Tidoni & Candidi, 2016). In this chapter, we wanted to contribute to the debate through methodologies that are able to provide information about the temporal dynamics of the neurophysiological operations involved in action recognition.

The ERP experiment directly mirrors the results of the behavioural experiments and provides important refinement: Both visual kinematics and non-motor goal-related information are perceptually processed during the first steps of action recognition, but the semantic interpretation of the non-motor goal-related information seems to occur earlier than the semantic interpretation of the visual kinematics. Such pattern may explain why, in the priming experiments, the influence of non-motor goal-related information is more important for 220 ms of prime duration, whereas the influence of visual kinematics surpasses the one of the non-motor goal-related information later, when primes are available for 300 ms of visual processing. Overall, as we discussed above, these results fit thoroughly with the neurocognitive model proposed by the predictive approaches: The earlier semantic processing of goal-related information (reflected by N300 modulations) may suggest the generation of a prediction about the visual kinematics, before the semantic integration of the two dimensions, reflecting the comparison between expected and observed visual kinematics (reflected by N400 modulations).

The TMS-priming paradigm provides additional evidence in favour of predictive approaches. In this study, we investigated the critical involvement of the frontoparietal network at an action step where both visual kinematics and non-motor goal-related information are processed and integrated. Sensorimotor approaches are not very definitive on the prediction they make about the involvement of the frontoparietal network: Indeed, depending of the theories discussed in Chapter 2, motor acts or action goals, or none of them, may be encoded in the network. Predictive approaches are quite clear on the involvement of the frontoparietal network, as it is supposed to be the area where visual kinematics and non-motor goal-related information are integrated. Thus, predictive approaches strongly emphasise the integrative role of the frontoparietal network, whereas the sensorimotor approaches do not. The results obtained in our TMS-priming experiment supports the existing data showing the involvement of the frontoparietal network in processing visual kinematics and non-motor goal-related information, and provide the first direct evidence for its involvement in integrating the two dimensions.

In Chapter 2, we stated that the account developed by Csibra (2008) could accommodate a lot of different pattern of results, as the main point of this assertion was that the motor simulation performed by the frontoparietal network was unnecessary for action recognition. Visual kinematics could then be processed first or not, and the frontoparietal network could be incidentally activated without any particular role during action recognition per se. TMS is particularly useful here as this technique directly evaluates whether the frontoparietal network is critical for action recognition or not. Previous “virtual lesions” studies already provided some evidence for its critical involvement during action recognition, but one may object that it may be due to the modulation of distant brain areas not directly stimulated but part of the network in which the frontoparietal node is embedded in. By targeting specific neuronal population at a particular step of action processing, TMS priming is less impacted by such criticisms. Our study is, as far as we know, the first TMS-priming paradigm that evaluates action recognition, and hence argues in disfavour of Csibra’s account, as disturbing the frontoparietal network has an impact on action recognition.

Overall, the neurophysiological experiments along with the behavioural experiments have addressed most of the important features of action recognition: The temporal dynamics at both the behavioural and neurophysiological level, along with some information about the neuronal substrate involving in processing visual kinematics and non-motor goal-related information. One remaining issue regards the format of the representations involved during the recognition of others' actions. Such information could be obtained using time-frequency analysis of the EEG data. In particular, the mu-rhythm and beta-rhythm desynchronisations[18] have been linked to the activity of the sensorimotor areas. These rhythms have been found to be specifically suppressed during movement executions, and similar desynchronisations have been observed during action recognition as well. Linking the processing of visual kinematics or the processing of non-motor goal-related information, or the integration of both to mu / beta rhythms desynchronisation would orient towards the sensorimotor format of the representations of these two dimensions within the frontoparietal network. Although much work remains to do, the results provided by the two neurophysiological experiments are fully compatible, and complete the results provided by the behavioural experiments. Non-motor goal-related information seems to guide action recognition, although visual kinematics seems to be processed simultaneously to (but independently from) non-motor goal-related information during the perceptual stages of action processing.

---

[18] The suppression or diminution of a particular frequency-band of oscillation of the brain activity recorded over the central areas using electroencephalography (EEG).

CHAPTER 3: INDIVIDUAL CHARACTERISTICS AND ACTION RECOGNITION: EXPLAINING THE INDIVIDUAL DIFFERENCES IN PRIORITISING GOAL OVER GRIP DURING THE FIRST STEPS OF ACTION RECOGNITION

This work is part of a larger project carried out in collaboration with Nicolas Morgado from Paris-Descartes University. This study has several components; only parts of it are reported here. This work is currently in preparation for publication:

Decroix, J., *Morgado, N. & Kalénine, S. (in prep). Preference for visual goal over grip explained by individual characteristics during the recognition of object-directed actions.*

In the previous chapters, we gathered evidence strongly suggesting that the recognition of object-directed actions was driven by the first processing of non-motor goal-related information. Although perceptually processed at the same time than non-motor goal-related information, visual kinematics seem to influence action recognition later than non-motor goal-related information. We concluded that these results fitted more with the predictive approaches than with the sensorimotor approaches.

Despite arguing in favour of the predictive approaches, we do not think that action recognition is an inflexible process. Instead, we acknowledge the existence of different strategies to recognise others' actions (Lautrey, 2003; Siegler & Shipley, 1995; Springer, Parkinson, & Prinz, 2013; Uithol & Paulus, 2014) and hypothesise that the strategy used will depend on individual characteristics and on situations. Consistent with this general pluralist view, the social context and the social characteristics of individuals have been shown to influence action recognition. For example, Lewkowicz et al. (2015) found that the ability to predict whether a "reach and grasp" movement aimed at grasping an object to take it for oneself or give it to someone else was positively correlated to the ability to identify the mental states of others. In the same vein, Pfeifer, Iacoboni, Mazziotta and Dapretto (2008) found positive correlations between the activity of the frontoparietal network during action recognition and social abilities in children (see also DiGirolamo, Simon, Hubley, Kopulsky, & Gutsell, 2019 for similar results in adults). Moreover, the activity of the frontoparietal network during action recognition was negatively correlated with impairment of social abilities in children with autism spectrum disorders (Dapretto et al., 2006). In the same vein, various indices of motor simulation have been positively associated with the ability to take the psychological perspective of others (Borgomaneri, Gazzola, & Avenanti, 2015; Y. Cheng, Yang, Lin, Lee, & Decety, 2008; Gazzola, Aziz-Zadeh, & Keysers, 2006; Keysers & Gazzola, 2006). Thus, social abilities seem to play a role in recognising others' actions. Note that the aforementioned indices considered the activity of the frontoparietal network as an index of motor simulation, and assumed that motor simulation was involved in action recognition. Yet, as discussed in the previous chapters, if the frontoparietal network is critically involved in action recognition, it is unclear whether this involvement should be related to the processing of visual kinematics or whether it should be related to the processing non-motor goal-related information.

Predictive approaches hypothesised that the recognition of others' actions is possible through the computation of a prediction about the action goal, and that this prediction guides the processing of visual kinematics (Kilner, 2011; Kilner et al., 2007). Recent accounts have suggested that the knowledge we have about others may serve as a source of information to derive predictions about the action goal, and that non-visual information may be involved in action recognition through predictive mechanisms (Bach & Schenke, 2017; Westra, 2019). Some data already suggest that our knowledge and representations about the world can bias the way we process others' actions. It is, for example, assumed that stereotypes may bias the recognition of others' actions through predictive mechanisms (Westra, 2019). For example, object-directed actions that are unfamiliar to the observer tend to require more activation of the frontoparietal network when compared to familiar object-directed actions (Nicholson et al., 2017). Similarly, recent evidence indicates that our appreciation of the actors (e.g., do I like the actor or not; Kozak, Marsh, & Wegner, 2006; Marsh et al., 2010) or our knowledge about the actors (e.g., is the object liked by the actor or not; Schenke, Wyer, & Bach, 2016) also influences the way we recognise the actions they perform. What remains to establish is whether indicators related to social abilities and individual characteristics could influence the way we process others' actions, and in particular the priority given to goal-related information during action processing.

Social power, as a particular individual characteristic, may be an interesting factor to consider, as it has been related to our social abilities such as the tendency to take the psychological perspective of others. Social power usually refers to the influence one has on others and the ability of a person to control and maintain this influence (Anderson, John, & Keltner, 2012; Berger, 2008; Blader & Chen, 2014; Galinsky, Magee, Inesi, & Gruenfeld, 2006). Accordingly, participants reporting a higher sense of social power (i.e., the feeling one has on their ability to influence others; see Anderson et al., 2012) also tend to report lower perspective taking abilities (Galinsky et al., 2006), which can be explain as obtaining and maintaining social power requires to sometimes ignore others' needs and desires. Furthermore, experimental manipulations to transiently increase or decrease the feeling of social power have an effect on the tendency to take the perspective of others': participants probed with high social power–they were asked to recall and write an essay about a situation in which they had a high-level of social power–take less the perspective of others' than participants probed with low social power (Galinsky et al., 2006; but see Cesario, Jonas, & Carney, 2017; Gronau, Erp, Heck, Cesario, & Jonas, 2017 for discussions on difficulties to replicate such subtle effects). Consequently, individual characteristics are related to our social abilities, which in turn are related to our ways to process others' actions.

In the present study, we wanted to explore the relationships between strategies employed by the observer to recognise others' actions, his/her individual characteristics, and his/her social behaviours. We considered the relative strength of goal similarity over the grip similarity effects on action recognition as an indicator of the engagement of the observer in predictive strategies to recognise others' actions. In other words, the higher the impact of goal similarity on target recognition relative to the impact of grip similarity, the higher participants rely on non-motor goal-related information, the more they are engaged in predictive strategies. We considered perspective taking as an interesting candidate of social abilities, as it has been frequently associated with the engagement of motor simulation during action recognition (e.g., Borgomaneri et al., 2015; Gazzola et al., 2006). Sense of social power was considered as a valuable candidate of the individual characteristics of the observer that could influence the individual strategies employed during action recognition (Anderson et al., 2012). Following the predictions of Galinsky et al. (2006), participants with higher sense of social power should be less engaged in using predictive strategies during action recognition. This prediction is particularly relevant to evaluate with priming paradigms, as these participants should be the ones who benefit the less from the information conveyed by the prime. In addition, preliminary work of Nicolas Morgado, expert in social power and collaborator on the project, suggests that the relationship between social power and perspective taking may not be as reliable as previously thought, possibly because social power is confounded with dominance (i.e., the explicit reliance on aggressive strategies to acquire and maintain social power). Thus, an explicit (i.e., using questionnaires) and implicit (i.e., using an implicit association test; IAT) evaluation of dominance was added to evaluate whether the relationship between social power and perspective taking could be related to the strategy one employs to obtain and maintain social power (i.e., dominance), rather than social power per se.

1. Methodology

1.1. Participants

Sixty-four healthy participants took part in the study[19] (M_{age} = 21, from 18 to 39; 32 females). All were right-handed (M_{EHI} = 86 %, from 40 % to 100 %), reported normal or corrected-to-normal vision. One participant was removed because of missing data in the report of the questionnaires. They provided written informed consent and received 10 euros for their participation. The protocol was approved by the Ethical Committee of the University of Lille and was in accordance with the declaration of Helsinki (1964, revised in 2013).

1.2. Procedure

The present study was divided in three main phases: (1) First, participants performed the priming paradigm presented in Chapter 4; (2) Then, they performed an Implicit Association Test (hereafter IAT, see below) developed by Nicolas Morgado; (3) And finally, they fulfilled a series of questionnaires (see below). The Behavioural Identification Form (hereafter, BIF) was always administered first, followed by the Interpersonal Reactivity Index (hereafter, IRI). Right after the two first questionnaires, half of the participants fulfilled the sense of power questionnaire then the dominance questionnaire, the other half fulfilled these two questionnaires in the reverse order. The overall order of the questionnaires was not fully randomised because there were 24 orders to represent, which would require an equivalent number of participants for each possible order (e.g., ten participants per possible order would require a total sample of 240 participants). We decided that it was not optimal as the ordering

---

[19] A sample size of about 50 participants was chosen to ensure sufficient statistical power (.80) for anticipated moderate effect sizes (Cohen *d* = .40 for the critical paired comparisons).

effect of questionnaires was not our primary interest. Thus, we decided to order the questionnaires from the more implicit to the less implicit. Questionnaires about social power and dominance were the most explicit, and the most susceptible to influence one another. As such, we only counterbalanced these two questionnaires. The overall experiment lasted ~1 hour and a half.

1.2.1. Priming paradigm

The priming paradigm used has already been reported as the 3rd Priming Experiment in Chapter 4. Briefly, action target photographs were primed by action photographs sharing the same grip and/or the same goal as the target. Primes were visually available for 66 or 220 ms, and were separated from the target by a pixelated mask lasting 66 ms. Participants had to judge as fast and accurate as possible whether the action target was typical or not according to the typical use of the object. Details can be found in the General Methodology section, and in Chapter 4, Experiment 3. As reported in Chapter 4, sharing the same action goal or the same grip as the target facilitates the action judgement even when primes were available for 66 ms of visual processing.

In the present study, we were interested in the relative strength of goal over grip influence during the very first steps of action recognition, namely, when primes were available for 66 ms of visual processing. Errors and response times over 1500 ms and below 150 ms were considered outliers and removed. We focused only on yes responses, as they showed greater sensitivity for the discrimination between goal and grip. First, we selected the median of the RTs of each participant in the four grip x goal similarity conditions for yes response and 66 ms of prime duration. Second, grip similarity effects and goal similarity effects were obtained for each participant by subtracting the mean of the median of the similar dimensions from the mean of the median of dissimilar dimensions, and dividing this difference by the overall mean of the median response times of each participant. Third, the grip similarity effect was subtracted from the goal similarity effect to obtain an individual goal priority index. The greater this final value, the more the individual prioritises visual goal information over grip information.

1.2.2. Implicit association test (IAT)

The IAT was used to evaluate the strength of association between the concept of dominance and the concept of self. The IAT was composed of three phases: two training phases for each concept, and the third phase to test the strength of association between the two concepts.

Phase 1:

"Dominant" and "submissive" were displayed at the top of a computer screen, one to the right, the other to the left. Half of the participants had "dominant" on the right and "submissive" on the left, the other half had the reverse pattern. One word related to one of the two categories appears at the centre of the screen, and participants were asked to categorise the word as fast and accurate as possible. The word remained on the screen until participants' response. Each trial started by a fixation cross for 500 ms, followed by the word to categorise. There were four words belonging to the category "dominant" ("dominant"; "autoritaire"; "supérieur"; "dominateur") and four words belonging to the category "submissive" ("soumis"; "assujetti"; "inférieur"; "asservi"). The eight words were presented twice, resulting in 16 trials.

Phase 2:

"Self" and "others" were displayed at the top of a computer screen. In Phase 1, the two categories were randomised *between* participants, but for a given participant the position of the two categories was fixed. In Phase 2, the position of the two categories was randomised from one trial to another (randomisation *within* participants). As previously, one word related to one of the two categories appears at the centre of the screen, and participants had to categorise it as fast and accurate as possible. Each trial started by a fixation cross for 500 ms, followed by the word to categorise. There were four words belonging to the category “self” (“moi”; “mon”; “soi”; “moi-même”) and four words belonging to the category “others” (“eux-mêmes”; “leurs”; “autrui”; “autres”). The eight words were presented twice, resulting in 16 trials.

Phase 3:

“Dominant” and “submissive” were displayed on the top of the screen and their position on each side of the screen was fixed and the same as Phase 1. “Self” and “others” were displayed right below the “dominant” and “submissive” categories. As in Phase 2, the position of the two categories could switch from one trial to another. Consequently, there was two possible configurations: “dominant” paired with “self” and “submissive” paired with “others” or “dominant” paired with “others” and “submissive” paired with “self”. The word appearing in the centre of the screen could belong to one of the four categories. Participants had to categorise the word as fast and accurate as possible. The eight words related to “dominant” / “submissive” and the eight words related to “self” / “others” were presented in both possible configurations (i.e., dominant-self/submissive-others and dominant-others/submissive-self), which resulted in 16 words x 2 possible configurations = 32 trials. The 32 trials were repeated five times, resulting in 160 trials overall.

The IAT index was extracted from the response times obtained in Phase 3 (Phase 1 and Phase 2 are considered as training phases). Errors in categorisation and response times inferior to 150 ms were not considered. The index was computed by subtracting the averaged response times of the configuration “dominant-self / submissive-others” from the averaged response times of the configuration “dominant-others / submissive-self”. The higher the index, the stronger the association dominant-self.

1.2.3. Interpersonal Reactivity Index (IRI)

The IRI was developed by Davis (1983). This questionnaire provides four sub-scores representing four dimensions of empathy. We used the French version of the IRI developed by Gilet, Mella, Studer, Grühn and Labouvie-Vief (2013). The IRI is composed of 28 items. Participants had to select from a 5-point Likert scale whether each item fits with their personality. The scale goes from “I completely disagree” (0; “Fortement en désaccord”) to “I completely agree” (4; “Fortement d’accord”). We will only focus on the sub-score generally reported in previous studies on action recognition: The perspective taking (PT) sub-score. See Appendix 8.

Scores were obtained by adding the individual score of each item. The higher the score the better the self-reported perspective-taking ability of the participants.

1.2.4. Personal sense of social power

The subjective sense of social power was developed by Anderson, John and Keltner (2012). The English version was first translated into French by French speakers experts in Psychology and fluent in English and then back-translated from French to English by bilingual individuals to ensure that the translated items remained similar to the original ones (Morgado, François & Palluel-Germain, personal communication). The personal sense of social power is composed of eight items. Participants had to select from a 7-point Likert scale whether each item fits with their personality. The scale goes from “I completely disagree” (1; “Fortement en désaccord”) to “I completely agree” (7; “Fortement d’accord”). See Appendix 9.

Scores were obtained by averaging the individual scores of the different items. The higher the score, the more participants considered having a high social power.

1.2.5. Behavioral Identification Form (BIF)

The BIF was developed by Vallacher and Wegner (1989). We obtained the French version by translating the English version into French by two French speakers, experts in Psychology and fluent in English. The obtained version was then back-translated from French to English by bilingual individuals to ensure that the translated items remained similar to the original ones. Each of the 25 items presents an action (e.g., “Attending class”). For each item, two alternatives were proposed (e.g., “sitting in a chair” or “looking at the blackboard”), and participants were asked to select the one they prefer. Among the two alternatives, one was always more abstract than the other (e.g., “looking at the blackboard”). See Appendix 10.

Scores were obtained by adding one point each time the “most abstract” alternative was selected. The higher the score, the higher the tendency of the participants to represent actions in an abstract manner.

1.2.6. Subjective ratings of dominance

The questionnaire of subjective ratings of dominance was developed by Cheng, Tracy and Henrich (2010). The French version was custom-developed for the purpose of the present study. There were seven items in the questionnaire. Participants had to select from a 5-point Likert scale whether each item fits with their personality. The scale goes from “Not at all” (1; “Pas du tout”) to “Totally” (5; “Totalement”). See Appendix 11.

Scores were obtained by adding the individual score of each item. The higher the score, the more participants consider themselves as adopting a dominant behaviour.

1.3. Data analysis

In the present study, we were interested in the effect of different personality dimensions on the relative strength of goal influence in comparison to the grip influence. Thus, we will try to evaluate the ability of the different scores evaluating the perspective taking (IRI), the level of action identification (BIF), the sense of social power and the dominance to predict the goal priority index.

2. Results

Given the exploratory nature of the study, we first looked at the paired correlations between each of the variables. In a second phase, we ran a multiple linear regression to explain the goal priority index by different scores. Finally, crossing the results of the correlations and the results of the multiple linear regression led us to further explore the relationship between the goal priority index, the score of dominance and the sense of social power.

2.1. Cross-correlation

The cross-correlation of Pearson between the variables grip similarity effect, goal similarity effect, goal priority index and BIF, PT, personal sense of social power, dominance and IAT scores can be found in Appendix 12.

The goal priority index significantly correlated with grip, r = -.53, p < .001 and goal similarity effects, r = .77, p < .001, which is expected because the two effects are used to compute the index. The score of dominance significantly correlated with the score of sense of power, r = .44, p < .001, which means that the higher the subjective rating of dominance, the higher the feeling of sense of power. Contrary to what we could have expected, there was no correlation between the perspective taking score and the score of sense of power, r = -.13, p =.324. Perspective taking did however correlate with the score of dominance, r = -.32, p = .012, which means the higher the subjective rating of dominance, the lower the perspective taking score. No other correlations were significant.

2.2. *Multiple Linear Regressions*

To further explore the relationships between the goal priority index and the scores of the different questionnaires (BIF, Perspective Taking, sense of power and dominance), a multiple linear regression was performed. The model was:

Goal priority index ~ BIF + Perspective Taking + Sense of Power + Dominance

Following the advice of Field, Miles and Field (2012), the standardised residuals, the hat value and Cook's *d* were computed to verify whether the results of the model could have been driven by outliers. We followed an iterative procedure and kept removing participants from the model until the three indices did not detect any outliers anymore. Following this procedure, three models were fitted and five participants removed. The last regression model on the 58 remaining participants was significant, $R^2 = 0.20$, $F(4,53) = 3.33$, $p = .016$. Sense of social power, $\beta = 0.020$, $t(53) = 5.49$, $p = .019$, and self-rating of dominance, $\beta = -0.005$, $t(53) = -2.27$, $p = .027$, were significant predictors of the model (see Figure 38 and 39). These results indicate that a higher sense of social power predicts a higher goal priority index whereas a higher self-rating of dominance predicts a lower goal priority index.

Interestingly, on the remaining 58 participants there was no simple Pearson correlation neither between goal priority index and dominance, $r = -.24$, $p = .07$, nor between goal priority index and sense of social power, $r = .17$, $p = .21$. This absence of simple Pearson correlations in spite of being significant predictors in the model may reflect the presence of moderation effects between the sense of social power and the self-rating of dominance (Hayes, 2009; Memon, Hwa, Ramayah, & Ting, 2018). To further explore this possibility, we ran a Sobel test (Hayes, 2009; Sobel, 1982, 1986). Sobel tests are used to evaluate whether the relationship between an independent variable and a dependent variable is modulated by a third variable. Here, our dependent variable is the index of goal priority. As the independent variable, we selected the sense of social power, and as the third variable, the level of dominance. As dominance is one of the strategies used to increase our social power (J. T. Cheng et al., 2010), we believed that it would make more sense to orient our analysis this way. Thus, we evaluated whether dominance was modulating the effect of social power in explaining the variations of goal priority. This analysis first requires to run a regression explaining sense of social power by dominance, then a regression explaining the index of goal priority by sense of social power and dominance (see Figure 37). The Sobel test was significant, Sobel $z = -2.14$, $p = .032$, which means that the score of dominance significantly affected the relationship between the sense of social power and the goal priority index.

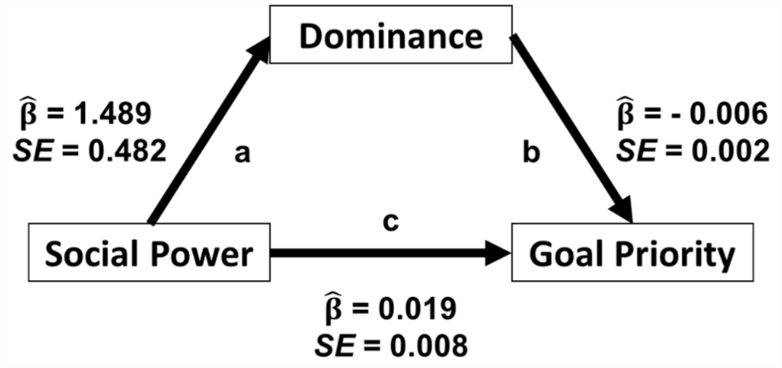


Figure 37. Impact of the sense of social power on goal priority via dominance. Bêta represents the estimate of the model for social power (path c) and dominance (path b) predicting goal priority and the estimate for social power predicting dominance (path a).

3. Discussion

In the reported study, we were wondering whether the engagement into predictive mechanisms could be related to individual characteristics and social abilities. Previous theoretical accounts stressed out the need to investigate the relationships between action recognition and social abilities (e.g., Bach & Schenke, 2017). The relative strength of goal similarity over grip similarity on action recognition was taken as an index of engagement in predictive mechanisms. The computation of this index directly stemmed from the position we took in the previous chapters, and the way we interpreted our results, namely that reliance on non-motor goal-related information instead of visual kinematics could be best interpreted in the predictive approaches of action recognition. We found that the sense of social power and self-rating of dominance predicted the evolution of this index. In particular, the sense of social power was positively associated with the index of goal priority: the higher the sense of social power, the more observers are engaged in predictive mechanisms. Self-rating of dominance had the opposite effect: the higher the dominance, the less observers were engaged in predictive mechanisms. Importantly, the effect of social power on the goal-priority index was only revealed when dominance is taken into account.

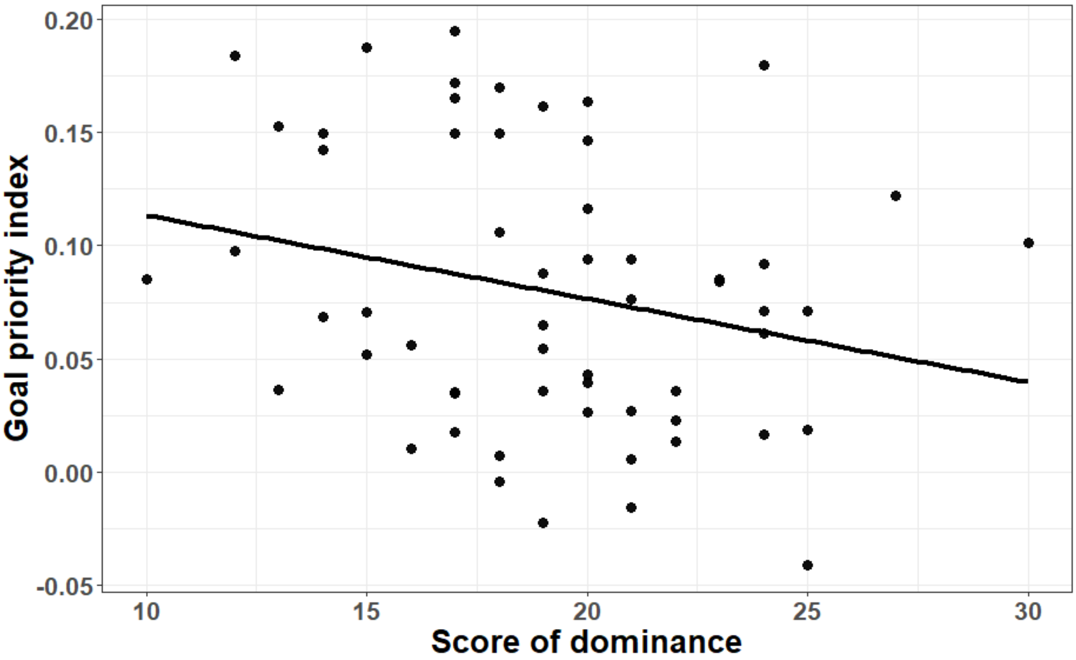


Figure 38. Correlation between the goal-priority index and the score of dominance. The higher the score of dominance, the lower the goal priority index.

Social power and dominance were considered as interesting candidates for predicting individual variability in terms of relying on non-motor goal-related information instead of visual kinematics. This choice was driven by hypotheses suggesting an influence of the sense of social power on social abilities (e.g., Galinsky et al., 2006). In particular, Galinsky et al. (2006) hypothesised opposite effects between social power and the ability to take the psychological perspective of others: the higher the sense of social power, the lower the perspective taking ability. They based their hypotheses on the negative relationship found between the two dimensions. In our experiment, we failed to replicate this result, which reinforces some preliminary works conducted by our collaborator Nicolas Morgado, who could not replicate this result either. In anticipation of this issue, we included a measure of self-rating of dominance: we indeed hypothesised that the relationship between perspective taking abilities and social power could rather be due to the strategy used to obtain and maintain social power. Dominance consists in using intimidation and fear to obtain social power (J. T. Cheng et al., 2010), which requires, to some extent, to not empathise with others. Accordingly, dominance was negatively associated with the ability to take the perspective of others. Thus, it is possible that the results obtained by Galinsky et al. (2006) may have been driven by a sur-representation of participants using dominance as a strategy to obtain social power in their sample.

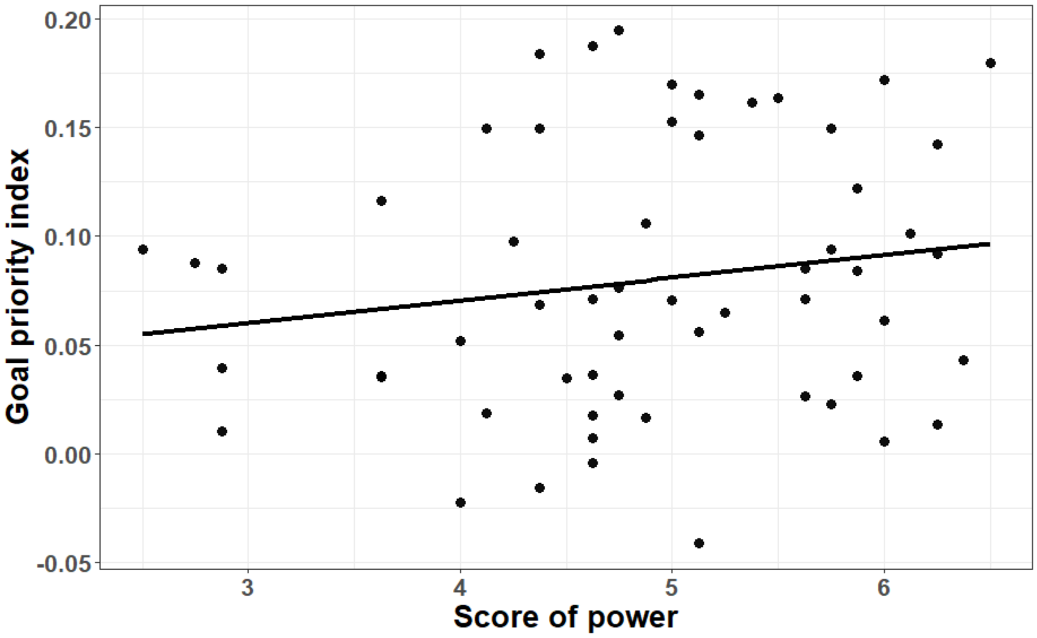


Figure 39. Correlation between the goal priority index and the score of social power. The higher the score of social power, the higher the goal priority index.

Following the hypotheses of Galinsky et al. (2006) on social power and perspective taking abilities, we predicted that people with a higher sense of social power and dominance would have more difficulties to recognise the actions of others, and would less engage in predictive strategies of action recognition. Our results are partially congruent with this hypothesis. Higher self-reported dominance did predict lower engagement in predictive mechanisms. Higher sense of social power however predicts the opposite relation with higher engagement in predictive mechanisms. Furthermore, dominance was modulating the relationship between social power and the engagement in predictive mechanisms. This pattern is fully congruent with our interpretation of the relationships between social power and dominance. The use of a dominant strategy to obtain and maintain social power has a detrimental effect on the use of predictive strategies during action recognition. Different strategies to obtain and maintain social power may be, in contrast, associated positively with the use of predictive strategies during action recognition. For example, it has been suggested that social power could be obtained through strategies based on attentiveness and compassion (e.g., J. T. Cheng et al., 2010), which requires to empathise with others. It is possible that people who make use of a dominant strategy to obtain and maintain social power rely less on non-motor goal-related information because they are less inclined to proactively predict others' actions and instead, process observed actions as they unfold. Further studies should explore in more details these relations.

Overall, our results suggest that our individual tendency to rely on visual kinematics or non-motor goal-related information may vary as a function of different individual characteristics. Uncovering how individual characteristics may affect action recognition could be an important step to improve models of action recognition. In particular, our study acknowledges the possibility that action recognition may consist in different strategies, and the selection of some strategies over others may depend of individual preferences (Lautrey, 2003; Siegler & Shipley, 1995; Springer, Brandstädter, & Prinz, 2013; Uithol & Paulus, 2014).

General Discussion

1. Overall summary

Recognising actions performed by others is a core ability of humans, and is essential in our social everyday life. When recognising others' actions, we do not only process their actions as they unfold, but we anticipate the course of the action as well as why they are performed for, that is, we anticipate their goals (see Chapter 1). The anticipation of the unfolding movement, and the recognition of others' goals have become a central question in the action recognition literature, as goals are thought to be in the actor's mind. Then, how are we able to access to the actor's goal? The action recognition literature has considered this question from two sides (see Chapter 2). On the one hand, sensorimotor approaches have considered that goals are, in fact, reflected in the actor's visual kinematics. The observer can extract the goal-related information in the visual kinematics, and thus, understand the actor's action goal. This ability is thought to rely on motor simulation by which the observer relies on its own motor system to anticipate the actor's goal. The same mechanisms are hypothesised in the motor control literature: motor programs are turned into the perceptual consequences they aim to achieve in order to evaluate whether these motor programs are the best to fulfil the actor's goal. The relationship between goals and motor programs is nonetheless not straightforward, and predictive approaches have argued that observers could not make sense of visual kinematics on their own. Instead, predictive approaches propose that the processing of visual kinematics needs to be guided by a prediction about the actor's goal. Predicting the actor's goal allows to predict the visual kinematics one needs to expect. Finally, the predicted kinematics can be compared to the actual visual kinematics. The actual visual kinematics will, in turn, be used to adjust and optimise the predictions. Consequently, predictive approaches hypothesised that observers need to first rely on action components that are not the visual kinematics. Both approaches have been powered up by the mirror neuron literature in which the frontoparietal motor system is thought to be a key node for action recognition. Yet, in most of the situations, the two approaches discussed the same experimental evidence provided by the mirror neuron literature. Thus, evidence coming from the mirror neuron literature were not decisive, and further sources of experimental evidence were needed. In this thesis, we tried to evaluate whether observers first rely on visual kinematics or whether they first rely on non-motor goal-related information when recognising others' actions.

Predictive approaches in their initial formulation did not make any strong claim about the kind of contextual information that could be used to derive these goal predictions, but some authors later proposed that objects could be good candidates for deriving goal-related information (see Chapter 2, section "The affordance-matching hypothesis"). Therefore, we reduced our scope to object-directed actions. In object-directed actions, goal-related information does not only rely on the visual kinematics, but also on the proper way to handle the object. Consequently, we designed action photographs in which visual kinematics and non-motor goal-related information could vary independently from one another. Visual kinematics were manipulated through variations of the grip configuration applied on the object, and non-motor goal-related information was essentially manipulated through the object position relative to the hand. This design was later implemented in various experimental situations to more precisely evaluate the temporal dynamics of the influence of visual kinematics and non-motor goal-related information on action recognition.

In Chapter 4, we found that the observers gave greater importance to non-motor goal-related information compared to visual kinematics during the first steps of action recognition, but that this tendency was reversed during the later steps of action recognition. In Chapter 5, we further found that the first steps of action recognition involved perceptual processing of both visual kinematics and non-motor goal-related information. Interestingly, our data suggest that post-perceptual stages of action recognition were driven by semantic processing of non-motor goal-related information, which was first visible after ~300ms of visual processing. The integration of the two sources of information occurs even later, after ~400ms of visual processing (Chapter 5, EEG). We finally provided evidence for a causal role of both frontal and parietal nodes of the action observation network in integrating the two sources of information (Chapter 5, TMS). In Chapter 6, we reported that relying on non-motor goal-related information or visual kinematics could be affected by individual preferences. We further showed that these individual preferences in relying on non-motor goal-related information versus visual kinematics could be related to some social characteristics of the observers.

2. Visual kinematics and non-motor goal-related information during the recognition of object-directed actions

Overall, our data suggest that non-motor goal-related information is prioritised over visual kinematics during the recognition of object-directed actions. The influence of non-motor goal-related information exceeds that of visual kinematics already during perceptual stages of action recognition. This influence becomes even clearer during post-perceptual stages of action recognition where non-motor goal-related information may drive the interpretation of the visual kinematics. The influence of visual kinematics keeps rising all along the processing of others' actions, to finally overcome the influence of non-motor goal-related information during the late post-perceptual stages of action recognition. In the following discussion, we will first discuss if the presence of the object could have led the observers to first rely on non-motor goal-related information. Then we will consider the present findings in the light of other findings orienting towards a late involvement of visual kinematics during action recognition.

2.1. Object identity, an unlikely predictor

In order to create a situation in which goal-related information would not rely on visual kinematics only, we decided to investigate the particular case of object-directed actions. By varying the position of the object relative to the hand, we could vary the possibility that the action could end with the typical functional goal we could expect from the object identity. The visual kinematics were varied through the grip configuration applied on the object, which could correspond to the typical manipulation we could expect from the object or not. In both cases, the two pieces of information were neither typical nor atypical overall (e.g., the grips were always biomechanically correct). It is their relationship vis-à-vis object identity that made them correct or incorrect. An important consequence of this manipulation is that the object is the cornerstone of the experimental strategy. The first processing of object identity is necessary for the task. It is legitimate to wonder whether the first processing of object identity could explain why participants first relied on non-motor goal-related information instead of visual kinematics.

First, we must highlight that the object is always the same across the different conditions, so the object knowledge is shared between the action violation conditions. Thus, the object always had to be identified. If any bias can be found, it may be first through the identification of the object. A natural objection would be that the access to object identity could be easier in some situations compared to others. For example, an object shifted from left to right on the vertical axis may be more difficult to identify than the same object shifted from right to left on the vertical axis. Classical studies on visual object discrimination have long since demonstrated that the visual presentation of the object has no impact on retrieving object identity–a phenomenon known as mirror invariance (E. E. Cooper, Biederman, & Hummel, 1992; Dilks, Julian, Kubilius, Spelke, & Kanwisher, 2011). The EEG literature further reinforces these data by showing that the visual perceptual processing of faces and bodies is strongly impacted by the inversion of the visual stimuli (inversion effect), whereas the visual perceptual processing of objects is not (Peelen & Downing, 2007; Rossion, Gauthier, et al., 2003; Rossion & Jacques, 2008). Therefore, it is unlikely that the position of the object relative to the hand could have affected the access to object identity.

Another potential bias comes from the influence of functional and manipulation knowledge involved in object semantics and derived from object identify. In the General Methodology section, we established that the two kinds of knowledge were required to perform the task: Functional knowledge to identify the functional goal of the object, and manipulation knowledge to identify the hand posture associated with the use of the object. Object identity gives equally access to both functional and manipulation knowledge independently (Buxbaum & Kalénine, 2010; van Elk et al., 2014a). One may still argue that functional knowledge could be more easily accessed than manipulation knowledge or that the weight of functional knowledge in object semantics could be greater than that of manipulation knowledge. Yet manipulation knowledge is a semantic feature of objects to the same extent that functional knowledge is (Campanella & Shallice, 2011). Moreover, the literature on object semantics does not provide clear response on whether functional knowledge is accessed before, after or simultaneously as manipulation knowledge (Buxbaum & Kalénine, 2010; van Elk et al., 2014a). It is sure that the two types of knowledge may be more or less involved depending on the task requirement, but our task does not clearly address one or the other type of knowledge. At best, our results may indicate that functional knowledge may be prioritised over manipulation knowledge during the identification of object-directed actions. This would still provide valuable information on the interplay between functional and manipulation knowledge during action recognition, but would remain to be tested more directly.

2.2. The late involvement of visual kinematics in action recognition

In the action recognition literature, visual kinematics have been the main and first visual information of interest when trying to make sense of others' actions. By introducing the idea that visual kinematics information could not be processed without being guided by a prediction about the actor's goal, predictive approaches also introduced the idea that different sources of information should be first used to compute this prediction. As we established in Chapter 3, sensorimotor proponents have extensively demonstrated that visual kinematics could provide sufficient information to allow the observer to recognise action goals. Thereby, visual kinematics clearly fuels decisions about action goals (Ansuini et al., 2014; Cavallo et al., 2016; Lewkowicz et al., 2015; Quesque & Coello, 2015). Yet observers seem to benefit the most from visual kinematics when other sources of information are absent. Indeed, reliance on kinematic information is increased when the object is not visible (Thioux & Keysers, 2015), when the action is difficult to recognise (Nicholson et al., 2017) or when the recognition of the action directly addresses and requires the processing of fine-grained kinematic information (Pobric & Hamilton, 2006; Tidoni et al., 2013). In all of the above studies, the increased reliance on visual kinematics goes along with an increased activity within the frontoparietal network, which suggest that this network may be sensitive to the relative place of visual kinematics relative to other sources of goal-related information. We also previously mentioned studies showing that motor simulation first reflects the kinematics the observer can expect from the context, and only later the kinematics actually performed by the actor (Cavallo et al., 2013; Koul et al., 2019). In a similar vein, proactive gaze movements–often considered as an evidence that observers anticipate the end-goal of the action–have been found when observers have sufficient information to predict the action goal. When no contextual information allows the computation of a prediction (Donnarumma et al., 2017), or when the observer has no sufficient motor expertise in regards to the observed action (Ambrosini et al., 2013; Geangu et al., 2015), proactive gaze movements are not observed anymore. The greater involvement of visual kinematics during the late stages of action recognition in our experiments (Chapter 4 and 5) is clearly in line with the aforementioned literature. Following sensorimotor approaches, it is not clear why visual kinematics should influence action recognition that late. Predictive approaches, in contrast, do hypothesise that visual kinematics are used to optimise the top-down prediction about the actor's goal. One specificity of our experimental strategy is that both visual kinematics and non-motor goal-related information were made available at the very same time. This was achieved by using static action stimuli. Although one may wonder to what extent the results may be generalisable to dynamic stimuli (including the postural but also the amplitude/timing aspects of visual kinematics), this procedure was necessary to control for the timing of availability of the different sources of visual information. We could still have expected visual kinematics to be processed first in the absence of contextual cues to build predictions on. Thus, our data may suggest that even when no priors are provided to the observers, their cognitive system first attempts to compute a prediction, and that visual kinematics (at least the postural aspects) are not the most appropriate information to do so in comparison to other sources of goal-related information.

2.3. Synthesis

Overall, we do not believe that object knowledge could have biased our results in any predictable ways. In addition, the present findings corroborate and extend previous results by showing that visual kinematics are secondly used during action recognition, even when no extra source of information has been provided upstream. In the framework of predictive approaches, the results reported in this thesis would be interpreted as follows: participants first use non-motor goal-related information (i.e., the object position relative to the hand) to compute a prediction about the actor's goal (Chapter 4, Priming, Visual Search), most likely during the immediate post-perceptual stages of action processing (Chapter 5, EEG). The prediction about the actor's goal allows participants to anticipate what should be the visual kinematics of the observed action. The expected visual kinematics are then compared to the actual visual kinematics within 400 ms (Chapter 5, EEG), most likely within the frontoparietal network (Chapter 5, TMS). The aforementioned literature already pointed out in that direction by showing that when presented prior to the action, various sources of information can be used to recognise others' actions. Our results further demonstrate that when two sources of goal-related information are presented at the same time, visual kinematics are still not prioritised.

3. Versatile action recognition: various strategies and mechanisms

By arguing that non-motor goal-related information is prioritised during action recognition, we do not mean that visual kinematics cannot be used to make sense of others' actions. Some data even demonstrate that the observer can use the sole information provided by visual kinematics to predict the upcoming event of an action (e.g., Avenanti et al., 2017; Lewkowicz et al., 2015). Manera, Becchio, Schouten, Bara and Verfaillie (2011) have, for example, found that the pure visual kinematics of communicative actions of one actor could be used as pieces of information to predict the actions of another actor, even though there was no direct contact between the two actors. In line with the sensorimotor approaches, we still found that visual kinematics are processed from the first perceptual stages of action recognition. What we argue is that when confronted to other sources of goal-related information, visual kinematics are not prioritised and not the most useful information.

The literature on action recognition increasingly acknowledges the plurality of the strategies and mechanisms underpinning the processing of others' actions. Although we do not think that object semantics could have biased the pattern of results obtained in any predictable ways, it may be possible that relying first on non-motor goal-related information in object-directed actions could be specific to that category of actions. In the following section, we will first argue that object-directed actions are, indeed, a particular category of action, then we will discuss a pluralist perspective of action recognition.

3.1. The specificity of object-directed actions

In the action recognition literature, objects have sometimes been claimed as being part of the context of the action, and could act as a mere confound in the study of action recognition (Hommel, 2014; Ruggiero & Catmur, 2018). Although we value these insights, and that objects should be carefully considered when studying action recognition, we also believe that object-directed actions are an action category on their own (Bach et al., 2014; van Elk, van Schie, & Bekkering, 2014b). We further argue that objects are constitutive parts of object-directed actions and by no means a mere context in them. While designing an experiment in which visual kinematics were to be used to prime the subsequent denomination of an object, Sim, Helbig, Graf and Kiefer (2014) asked participants to guess which objects were used in action video clips where only the visual kinematics were available (objects were removed during the video editing): participants were unable to determine which objects were used. Using EEG, Wamain, Pluciennicka and Kalénine (2014) also found that point-light displays presenting object-directed actions were processed differently from point-light displays presenting meaningful actions non-directed towards objects. Some recent theoretical arguments in this direction can also be found in the action recognition literature. Incidentally, by arguing that communicative gestures should be considered as a separate category of actions, Novack and Goldin-Meadow (2016) also support the idea that object-directed actions constitute a category on their own. Even more substantial, if isolating object-directed actions as a separate category seems to be quite audacious in the field of action recognition, this distinction in the literature on apraxia is part of common knowledge (Bartolo et al., 2019; Bartolo & Ham, 2016; Buxbaum, 2001;

Buxbaum & Kalénine, 2010; H. L. Gallagher & Frith, 2004). Object-directed actions (i.e., transitive actions) are dissociated from actions miming the use of an object (i.e., pantomime), which are themselves dissociated from actions without objects, but used to communicate (i.e., intransitive actions). If object-directed actions are indeed a separate category of action, it does not seem right to exclude objects from models of action recognition. It most likely implies that some processes are specific to object-directed actions, but these processes should be investigated on their own (Bach, Nicholson, & Hudson, 2015; van Elk et al., 2014b). If we may not extend the results obtained on object-directed actions to all types of actions, the reverse is also true: results on actions non-directed towards objects should not carelessly be extended to the recognition of object-directed actions.

Recognising object-directed actions as a particular type of action opens an interesting path at the border of action recognition and object semantics. The case of pretence may be an interesting way to investigate this relationship. In this case, a functionally irrelevant object can be used to mime an action normally performed with another functionally relevant object (Schubotz & von Cramon, 2009; Schubotz et al., 2014). Using a pencil to suggest a hammering action is possible and does not depend on the object-information: here the whole goal-related information relies on the visual kinematics. At best, object information may interfere with the processing of the action. Schubotz et al. (2014) found that object semantics were related to the activity of the inferior parietal lobule whereas the actual action (e.g., hammering with a pencil) was related to the activity of the inferior frontal gyrus. Through pretence, it may be possible to dissociate the information conveyed by object knowledge from the information conveyed by the visual kinematics. An extended perspective would be to investigate the development of this ability. Indeed, pretend-play (or symbolic play) is thought to be particularly important in the development of various cognitive abilities in children, including the development of social abilities such as Theory-of-Mind (Lillard, 2017; Quinn, Donnelly, & Kidd, 2018; Smith, Englander, Lillard, & Morris, 2013). Another orientation that the literature on object semantics has started to explore is the interplay between visual kinematics and object recognition. When participants are asked to perform perceptual judgements on object, visual kinematics spontaneously influence task performance, even if they are irrelevant (Borghi, Flumini, Natraj, & Wheaton, 2012; Kumar, Riddoch, & Humphreys, 2013; Kumar et al., 2012; Natraj, Pella, Borghi, & Wheaton, 2015; Natraj et al., 2013). These data indicates that objects may not only inform about visual kinematics, but visual kinematics also inform about objects (Bach et al., 2014). In any case, investigating the interplay between objects and visual kinematics will be an important step to improve our understanding of the recognition of object-directed actions.

3.2. A pluralist view of action recognition

Recognising object-directed actions as a particular category of actions and further recognising that different mechanisms may underpin the processing of this category go along with recognising that action recognition may not be a unitary process. Some theoretical accounts have started to argue that action recognition would be better considered as a toolbox of several processes and strategies rather than a unitary process (Bach et al., 2014; Springer, Parkinson, et al., 2013; Uithol & Paulus, 2014). Already in the mirror neuron literature, some positions acknowledged the plurality of the mechanisms in action recognition: the involvement of motor simulation when we observe actions from a third-person perspective may work differently from when we are engaged in an interaction, from a second-person perspective (Quesque & Coello, 2015; Schilbach, 2010; Sebanz & Knoblich, 2009). In the General Methodology section, we already emphasised the plurality of the strategies and mechanisms involved during action recognition. Indeed, we wanted a task that did not explicitly address one of the two visual sources of information. Studies from Flanagan et al. (2013) nicely demonstrated that the visual strategy used to process an observed action varies as a function of the participants' task. Evaluating the weight of the object currently used by the actor induces a gaze pattern very similar to the one the participant would use to perform the task himself (Flanagan & Johansson, 2003; Flanagan et al., 2013), which was not observed when participants had to predict which of two objects would be grasped (Flanagan et al., 2013). These data demonstrate that the use of visual kinematics varies depending on the participant's task. Similarly, a growing body of evidence now suggests that the involvement of the frontoparietal network during action recognition is highly sensitive to various factors (D'Innocenzo et al., 2017; Donaldson et al., 2015; Leonetti et al., 2015; Muthukumaraswamy & Singh, 2008; Perry et al., 2010; Pomiechowska & Csibra, 2017; Riach et al., 2018; Schuch et al., 2010; Woodruff & Klein, 2013; Wright et al., 2018). For instance, when the visuo-attentional system focuses less on the visual kinematics, the sensorimotor activity fades away evenly (Muthukumaraswamy & Singh, 2008; Perry & Bentin, 2009; Perry et al., 2010). In *f*MRI studies, the activity of the frontoparietal network decreases when non-motor goal-related information are available (Nicholson et al., 2017; Thioux & Keysers, 2015). In the same vein, Pomiechowska and Cisbra (2017) recently reported a decrease of sensorimotor engagement during action observation when the observed action was preceded by language information when compared to observed action preceded by pure noise. Language might be considered here as a non-motor source of information that decreases the reliance on visual kinematics, and thus reduces the frontoparietal activity. Together, such results illustrate that contextual and situational factors may influence the mechanisms of action recognition. In our own data, we found that both visual kinematics and non-motor goal-related information were considered during action recognition, but that their relative priority during the processing of observed actions varies. Chapter 6 was built as an extension of Chapter 4 and 5 as it shows that the individual preferences in relying of non-motor goal-related information relative to visual kinematics during the first steps of action processing could be predicted by individual characteristics. Overall, the aforementioned literature supports a pluralist view of action recognition, in which the mechanisms and strategies to process others' actions are affected by the characteristics of the individuals and the situations.

Uithol and Paulus (2014) identified four types of strategies that may be particularly at play in certain action recognition situations and that may be underpinned by different cognitive and neuronal mechanisms. Action classification refers to tasks requiring the recognition of others' actions as belonging to a certain category of action: Infants that are surprised when an action video is interrupted in the middle of an action instead of between two actions need to recognise that the sequence of the interrupted action is in fact only a part of this action (Baldwin et al., 2001). Target prediction refers to tasks requiring to anticipate the outcome of an ongoing action before its end. This kind of task has been massively used to support the use of visual kinematics during action recognition (e.g., Ansuini et al., 2014; Quesque & Coello, 2015 for review; see also Chapter 3). Super-ordinate action recognition requires participants to understand to what kind of higher goal an isolated action goal could belong (e.g., cutting tomato is part of making a sandwich). This form of action recognition is mainly involved in studies evaluating the ability of participants to make sense of an action sequence (Braukmann et al., 2017; Giglio et al., 2013; Hrkać et al., 2014; Maffongelli et al., 2015; Maffongelli, D'Ausilio, Fadiga, & Daum, 2019). Finally, response selection refers to the ability of an observer to produce an appropriate response given a social cue. For example, if you meet someone for the first time, and a hand is moved towards you, you spontaneously move your contralateral hand because you recognise the actor's movement as being part of a shaking hand movement. Response selection is massively at play in joint action, as observers need to coordinate their own actions with the actions of the actor (Quesque & Coello, 2015; Sebanz & Knoblich, 2009). Overall, Uithol and Paulus (2014) admitted that the different forms of action recognition most likely coexist altogether, although some particular instances may help to reveal some particular forms. For example, our task (i.e., judging whether an object-directed action would be correct or not according to the typical use of object) may be related to the action classification form of action recognition: Indeed, we need to recognise that grasping an upright pencil with a precision grip is related to the action category of “writing”. Nonetheless, our task may also be related to the super-ordinate form of action recognition: our participants had to acknowledge that the action photographs displayed only part of the action sequence they belong to (e.g., grasping a pencil is the first step to later write on a sheet of paper). To some extent, our stimuli could also be related to the target prediction form of action recognition, as identifying a correct action photograph of a pencil in the hand may ease the later recognition of a sheet of paper because of the expected development of the action. A first step towards a more comprehensive view of pluralist conceptualisations of action recognition could be to investigate how the different strategies would use different sources of action-related information.

4. Concluding remarks

The last decades have massively contributed to the emergence of action understanding as a major theme in the scope of social cognition in Psychology and Neuroscience. As an emerging theme, various tasks and paradigms have been used in an attempt to understand how humans are making sense of others’ behaviours. During action recognition, it seems clear that visual kinematics are involved, even if the mechanisms underlying their processing remain to be established. Recently, various sources of information such as the visual context, objects, the knowledge we have about the actor or its forthcoming action have all been found to influence action recognition. The overall neuronal network underlying the visual processing of actions has also been identified. Various TMS evidence further directly and causally involve the frontoparietal network in the processing of others’ actions. Finally, it seems quite established that action recognition involves not only the online decoding of actions, but also a semantic structure on which previous knowledge is used to make sense of others' actions. All along this thesis, we have tried as much as possible to question whether the data we obtained could be generalised or not, and to identify how much of our effects could be limited to the particular situations under investigation. This attempt was mainly motivated by the rising awareness of the methodological challenge social cognition is facing. Uithol and Paulus (2014) can be quoted as a first attempt to grasp the complexity of action recognition, but some parts of social cognition, particularly the theory-of-mind literature, have already started their transition towards a more organised field (Kennedy & Adolphs, 2012; Schaafsma, Pfaff, Spunt, & Adolphs, 2015). After a wide extension in various creative ways, we believe that the action recognition literature should now try to develop a concrete conceptual framework with clear hypotheses, and propositions to prove or disprove these hypotheses. Now that the field has data, it needs theories.

References

1. Allen, M., & Friston, K. J. (2018). From cognitivism to autopoiesis: towards a computational framework for the embodied mind. *Synthese*, *195*(6), 2459–2482. https://doi.org/10.1007/s11229-016-1288-5
2. Ambrosini, E., Costantini, M., & Sinigaglia, C. (2011). Grasping with the eyes. *Journal of Neurophysiology*, *106*(3), 1437–1442. https://doi.org/10.1152/jn.00118.2011
3. Ambrosini, E., Reddy, V., de Looper, A., Costantini, M., Lopez, B., & Sinigaglia, C. (2013). Looking Ahead: Anticipatory Gaze and Motor Ability in Infancy. *PLoS ONE*, *8*(7), 1–9. https://doi.org/10.1371/journal.pone.0067916
4. Ambrus, G. G., Dotzer, M., Schweinberger, S. R., & Kovács, G. (2017). The occipital face area is causally involved in the formation of identity-specific face representations. *Brain Structure and Function*, *222*(9), 4271–4282. https://doi.org/10.1007/s00429-017-1467-2
5. Amoruso, L., Finisguerra, A., & Urgesi, C. (2016). Tracking the time course of top-down contextual effects on motor responses during action comprehension. *Journal of Neuroscience*, *36*(46), 11590–11600. https://doi.org/10.1523/JNEUROSCI.4340-15.2016
6. Amoruso, L., Gelormini, C., Aboitiz, F., Alvarez González, M., Manes, F., Cardona, J. F., & Ibanez, A. (2013). N400 ERPs for actions: building meaning in context. *Frontiers in Human Neuroscience*, *7*(March), 57. https://doi.org/10.3389/fnhum.2013.00057
7. Amoruso, L., Sedeño, L., Huepe, D., Tomio, A., Kamienkowski, J., Hurtado, E., … Ibanez, A. (2014). Time to Tango: Expertise and contextual anticipation during action observation. *NeuroImage*, *98*, 366–385. https://doi.org/10.1016/j.neuroimage.2014.05.005
8. Amoruso, L., & Urgesi, C. (2016). Familiarity modulates motor activation while other species' actions are observed: A magnetic stimulation study. *European Journal of Neuroscience*, *43*(6), 765–772. https://doi.org/10.1111/ejn.13154
9. Anderson, C., John, O. P., & Keltner, D. (2012). The Personal Sense of Power. *Journal of Personality*, *80*(2), 313–344. https://doi.org/10.1111/j.1467-6494.2011.00734.x
10. Ansuini, C., Cavallo, A., Bertone, C., & Becchio, C. (2014). The visible face of intention: Why kinematics matters. *Frontiers in Psychology*, *5*(JUL), 1–6. https://doi.org/10.3389/fpsyg.2014.00815
11. Ansuini, C., Santello, M., Massaccesi, S., & Castiello, U. (2005). Effects of End-Goal on Hand Shaping. *Journal of Neurophysiology*, *95*(4), 2456–2465. https://doi.org/10.1152/jn.01107.2005
12. Avanzini, P., Fabbri-Destro, M., Campi, C., Pascarella, A., Barchiesi, G., Cattaneo, L., & Rizzolatti, G. (2013). Spatiotemporal dynamics in understanding hand-object interactions. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(40), 15878–15885. https://doi.org/10.1073/pnas.1314420110
13. Avenanti, A., Annela, L., & Serino, A. (2012). Suppression of premotor cortex disrupts motor coding of peripersonal space. *NeuroImage*, *63*(1), 281–288. https://doi.org/10.1016/j.neuroimage.2012.06.063
14. Avenanti, A., Annella, L., Candidi, M., Urgesi, C., & Aglioti, S. M. (2013). Compensatory plasticity in the action observation network: Virtual lesions of STS enhance anticipatory simulation of seen actions. *Cerebral Cortex*, *23*(3), 570–580. https://doi.org/10.1093/cercor/bhs040
15. Avenanti, A., Candidi, M., & Urgesi, C. (2013). Vicarious motor activation during action perception: beyond correlational evidence. *Frontiers in Human Neuroscience*, *7*(May), 1–8. https://doi.org/10.3389/fnhum.2013.00185
16. Avenanti, A., Paracampo, R., Annella, L., Tidoni, E., & Aglioti, S. M. (2017). Boosting and decreasing action prediction abilities through excitatory and inhibitory tDCS of inferior frontal cortex. *Cerebral Cortex*, 1–15. https://doi.org/10.1093/cercor/bhx041
17. Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–

412. https://doi.org/10.1016/j.jml.2007.12.005

18. Bach, P., Gunter, T. C., Friederici, A. D., Knoblich, G., & Prinz, W. (2005). Action comprehension: Deriving spatial and functional relations. *Journal of Experimental Psychology: Human Perception and Performance, 31*(3), 465–479. https://doi.org/10.1037/0096-1523.31.3.465

19. Bach, P., Gunter, T. C., Knoblich, G., Prinz, W., & Friederici, A. D. (2009). N400-like negativities in action perception reflect the activation of two components of an action representation. *Social Neuroscience, 4*(3), 212–232. https://doi.org/10.1080/17470910802362546

20. Bach, P., Nicholson, T., & Hudson, M. (2014). The affordance-matching hypothesis: how objects guide action understanding and prediction. *Frontiers in Human Neuroscience, 8*(May), 254. https://doi.org/10.3389/fnhum.2014.00254

21. Bach, P., Nicholson, T., & Hudson, M. (2015). Response: No need to match: a comment on Bach, Nicholson, and Hudson's "Affordance-Matching Hypothesis." *Frontiers in Human Neuroscience, 9*(December), 2015–2017. https://doi.org/10.3389/fnhum.2015.00685

22. Bach, P., & Schenke, K. C. (2017). Predictive social perception: Towards a unifying framework from action observation to person knowledge. *Social and Personality Psychology Compass*, *11*(7), e12312. https://doi.org/10.1111/spc3.12312

23. Baldwin, D. A., Baird, J. A., Saylor, M. M., & Clark, M. A. (2001). Infants parse dynamic action. *Child Development, 72*(3), 708–717. https://doi.org/10.1111/1467-8624.00310

24. 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

25. Barsalou, L. W., Simmons, W. K., Barbey, A. K., & Wilson, C. D. (2003). Grounding conceptual knowledge in modality-specific systems. *Trends in Cognitive Sciences, 7*(2), 84–91. https://doi.org/10.1016/S1364-6613(02)00029-3

26. Bartolo, A., Claisse, C., Gallo, F., Ott, L., Sampaio, A., & Nandrino, J.-L. (2019). Gestures convey different physiological responses when performed toward and away from the body. *Scientific Reports, 9*(1), 12862. https://doi.org/10.1038/s41598-019-49318-3

27. Bartolo, A., & Ham, H. S. (2016). A Cognitive Overview of Limb Apraxia. *Current Neurology and Neuroscience Reports*, *16*(8), 75. https://doi.org/10.1007/s11910-016-0675-0

28. Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015). Parsimonious mixed models. *ArXiv Preprint ArXiv:1506.04967*, 1–27. https://doi.org/arXiv:1506.04967

29. Becchio, C., Koul, A., Ansuini, C., Bertone, C., & Cavallo, A. (2018a). Seeing mental states: An experimental strategy for measuring the observability of other minds. *Physics of Life Reviews*, *24*, 67–80. https://doi.org/10.1016/j.plrev.2017.10.002

30. Becchio, C., Koul, A., Ansuini, C., Bertone, C., & Cavallo, A. (2018b). The observability principle and beyond: Reply to comments on "Seeing mental states: An experimental strategy for measuring the observability of other minds" by Cristina Becchio et al. *Physics of Life Reviews*, *24*, 114–117. https://doi.org/10.1016/j.plrev.2018.01.006

31. Berger, C. R. (2008). Power, Dominance, and Social Interaction. In *The International Encyclopedia of Communication*. https://doi.org/10.1002/9781405186407.wbiecp093

32. Binkofski, F., & Buxbaum, L. J. (2013). Two action systems in the human brain. *Brain and Language*, *127*(2), 222–229. https://doi.org/10.1016/j.bandl.2012.07.007

33. Blader, S. L., & Chen, Y. (2014). What's in a Name? Status, Power, and Other Forms of Social Hierarchy. In *The Psychology of Social Status* (pp. 71–95). https://doi.org/10.1007/978-1-4939-0867-7

34. Blakemore, S.-J., & Decety, J. (2001). From the perception of action to the understanding of intention. *Nature Reviews Neuroscience, 2*(8), 561–567. Retrieved from www.nature.com/reviews/neuro

35. Bledowski, C., Prvulovic, D., Hoechstetter, K., Scherg, M., Wibral, M., Goebel, R., & Linden, D. (2004). Localizing P300 Generators in Visual Target and Distractor Processing: A Combined Event-Related Potential and Functional Magnetic Resonance Imaging Study. *Journal of Neuroscience*, *24*(42), 9353–9360. https://doi.org/10.1523/JNEUROSCI.1897-04.2004

36. Boehm, S. G., Dering, B., & Thierry, G. (2011). Category-sensitivity in the N170 range: A question of topography and inversion, not one of amplitude. *Neuropsychologia*, *49*(7), 2082–2089. https://doi.org/10.1016/j.neuropsychologia.2011.03.039

37. Borghi, A. M., Flumini, A., Natraj, N., & Wheaton, L. a. (2012). One hand, two objects: emergence of affordance in contexts. *Brain and Cognition*, *80*(1), 64–73. https://doi.org/10.1016/j.bandc.2012.04.007

38. Borgomaneri, S., Gazzola, V., & Avenanti, A. (2015). Transcranial magnetic stimulation reveals two functionally distinct stages of motor cortex involvement during perception of emotional body language. *Brain Structure and Function*, *220*(5), 2765–2781. https://doi.org/10.1007/s00429-014-0825-6

39. Borroni, P., Gorini, A., Riva, G., Bouchard, S., & Cerri, G. (2011). Mirroring avatars: Dissociation of action and intention in human motor resonance. *European Journal of Neuroscience*, *34*(4), 662–669. https://doi.org/10.1111/j.1460-9568.2011.07779.x

40. Bortoletto, M., Mattingley, J. B., & Cunnington, R. (2011). Action intentions modulate visual processing during action perception. *Neuropsychologia*, *49*(7), 2097–2104. https://doi.org/10.1016/j.neuropsychologia.2011.04.004

41. Botvinick, M. M. (2008). Hierarchical models of behavior and prefrontal function. *Trends in Cognitive Sciences*, *12*(5), 201–208. https://doi.org/10.1016/j.tics.2008.02.009

42. Brauer, M., & Curtin, J. J. (2018). Linear mixed-effects models and the analysis of nonindependent data: A unified framework to analyze categorical and continuous independent variables that vary within-subjects and/or within-items. *Psychological Methods*, *23*(3), 389–411. https://doi.org/10.1037/met0000159

43. Braukmann, R., Bekkering, H., Hidding, M., Poljac, E., Buitelaar, J. K., & Hunnius, S. (2017). Predictability of action sub-steps modulates motor system activation during the observation of goal-directed actions. *Neuropsychologia*, *103*(May), 44–53. https://doi.org/10.1016/j.neuropsychologia.2017.07.009

44. Brysbaert, M., & Stevens, M. (2018). Power analysis and effect size in mixed effects models: A tutorial. *Journal of Cognition*, *1*(1), 1–20. https://doi.org/10.5334/joc.10

45. Buresh, J. S., & Woodward, A. L. (2007). Infants track action goals within and across agents. *Cognition*, *104*(2), 287–314. https://doi.org/10.1016/j.cognition.2006.07.001

46. Buxbaum, L. J. (2001). Ideomotor Apraxia: a Call to Action. *Neurocase*, *7*(6), 445–458. https://doi.org/10.1093/neucas/7.6.445

47. Buxbaum, L. J. (2017). Learning, remembering, and predicting how to use tools: Distributed neurocognitive mechanisms. *Psychological Review*, *124*(3), 346–360. https://doi.org/10.1037/rev0000051

48. Buxbaum, L. J., & Kalénine, S. (2010). Action knowledge, visuomotor activation, and embodiment in the two action systems. *Annals of the New York Academy of Sciences*, *1191*, 201–218. https://doi.org/10.1111/j.1749-6632.2010.05447.x

49. Buxbaum, L. J., & Kalénine, S. (2020). Apraxia: A disorder at the cognitive-motor interface. In M. S. Gazzaniga (Ed.), *Handbook of Cognitive Neuroscience*.

50. Campanella, F., & Shallice, T. (2011). Manipulability and object recognition: is manipulability a semantic feature? *Experimental Brain Research*, *208*(3), 369–383. https://doi.org/10.1007/s00221-010-2489-7

51. Candidi, M., Urgesi, C., Ionta, S., & Aglioti, S. M. (2008). Virtual lesion of ventral premotor cortex impairs visual perception of biomechanically possible but not impossible

actions. *Social Neuroscience*, *3*(3–4), 388–400. https://doi.org/10.1080/17470910701676269
52. Caspers, S., Zilles, K., Laird, A. R., & Eickhoff, S. B. (2010). ALE meta-analysis of action observation and imitation in the human brain. *NeuroImage*, *50*(3), 1148–1167. https://doi.org/10.1016/j.neuroimage.2009.12.112
53. Catmur, C. (2015). Understanding intentions from actions: Direct perception, inference, and the roles of mirror and mentalizing systems. *Consciousness and Cognition*, 1–8. https://doi.org/10.1016/j.concog.2015.03.012
54. Catmur, C., Press, C., & Heyes, C. (2016). *Mirror Neurons from Associative Learning*. 1–45. https://doi.org/10.1002/9781118650813.ch20
55. Cattaneo, L. (2010). Tuning of ventral premotor cortex neurons to distinct observed grasp types: A TMS-priming study. *Experimental Brain Research*, *207*(3–4), 165–172. https://doi.org/10.1007/s00221-010-2454-5
56. Cattaneo, L., & Barchiesi, G. (2011). Transcranial Magnetic Mapping of the Short-Latency Modulations of Corticospinal Activity from the Ipsilateral Hemisphere during Rest. *Frontiers in Neural Circuits*, *5*(October), 1–13. https://doi.org/10.3389/fncir.2011.00014
57. Cattaneo, L., Barchiesi, G., Tabarelli, D., Arfeller, C., Sato, M., & Glenberg, A. M. (2011). One's motor performance predictably modulates the understanding of others' actions through adaptation of premotor visuo-motor neurons. *Social Cognitive and Affective Neuroscience*, *6*(3), 301–310. https://doi.org/10.1093/scan/nsq099
58. Cattaneo, L., Sandrini, M., & Schwarzbach, J. (2010). State-dependent TMS reveals a hierarchical representation of observed acts in the temporal, parietal, and premotor cortices. *Cerebral Cortex*, *20*(9), 2252–2258. https://doi.org/10.1093/cercor/bhp291
59. Cattaneo, Z., Devlin, J. T., Salvini, F., Vecchi, T., & Silvanto, J. (2010). The causal role of category-specific neuronal representations in the left ventral premotor cortex (PMv) in semantic processing. *NeuroImage*, *49*(3), 2728–2734. https://doi.org/10.1016/j.neuroimage.2009.10.048
60. Cattaneo, Z., Rota, F., Vecchi, T., & Silvanto, J. (2008). Using state-dependency of transcranial magnetic stimulation (TMS) to investigate letter selectivity in the left posterior parietal cortex: A comparison of TMS-priming and TMS-adaptation paradigms. *European Journal of Neuroscience*, *28*(9), 1924–1929. https://doi.org/10.1111/j.1460-9568.2008.06466.x
61. Cavallo, A., Bucchioni, G., Castiello, U., & Becchio, C. (2013). Goal or movement? Action representation within the primary motor cortex. *European Journal of Neuroscience*, *38*(10), 3507–3512. https://doi.org/10.1111/ejn.12343
62. Cavallo, A., Heyes, C., Becchio, C., Bird, G., & Catmur, C. (2014). Timecourse of mirror and counter-mirror effects measured with transcranial magnetic stimulation. *Social Cognitive and Affective Neuroscience*, *9*(8), 1082–1088. https://doi.org/10.1093/scan/nst085
63. Cavallo, A., Koul, A., Ansuini, C., Capozzi, F., & Becchio, C. (2016). Decoding intentions from movement kinematics. *Scientific Reports*, *6*(November), 37036. https://doi.org/10.1038/srep37036
64. Cesario, J., Jonas, K. J., & Carney, D. R. (2017). CRSP special issue on power poses: what was the point and what did we learn? *Comprehensive Results in Social Psychology*, *2*(1), 1–5. https://doi.org/10.1080/23743603.2017.1309876
65. Chang, Y.-T., Chen, H.-Y., Huang, Y.-C., Shih, W.-Y., Chan, H.-L., Wu, P.-Y., … Wang, C.-I. (2018). Left centro-parieto-temporal response to tool–gesture incongruity: an ERP study. *Behavioral and Brain Functions*, *14*(1), 6. https://doi.org/10.1186/s12993-018-0138-7
66. Changeux, J.-P. (1983). *L'homme neuronal* (Arthème Fa). Retrieved from http://www.amazon.fr/Lhomme-neuronal-Jean-Pierre-Changeux/dp/2818502888/ref=sr_1_1?s=books&ie=UTF8&qid=1395414543&sr=1-1&keywords=l%27homme+neuronal
67. Cheng, J. T., Tracy, J. L., & Henrich, J. (2010). Pride, personality, and the evolutionary

foundations of human social status. *Evolution and Human Behavior*, *31*(5), 334–347. https://doi.org/10.1016/j.evolhumbehav.2010.02.004

68. Cheng, Y., Yang, C. Y., Lin, C. P., Lee, P. L., & Decety, J. (2008). The perception of pain in others suppresses somatosensory oscillations: A magnetoencephalography study. *NeuroImage*, *40*(4), 1833–1840. https://doi.org/10.1016/j.neuroimage.2008.01.064

69. Chiavarino, C., Apperly, I. A., & Humphreys, G. W. (2012). Understanding Intentions: Distinct Processes for Mirroring, Representing, and Conceptualizing. *Current Directions in Psychological Science*, *21*(5), 284–289. https://doi.org/10.1177/0963721412452727

70. Coello, Y., & Delevoye-Turrell, Y. (2007). Embodiment, spatial categorisation and action. *Consciousness and Cognition*, *16*(3), 667–683. https://doi.org/10.1016/j.concog.2007.07.003

71. Cohen, M. X. (2017). Where Does EEG Come From and What Does It Mean? *Trends in Neurosciences*, *40*(4), 208–218. https://doi.org/10.1016/j.tins.2017.02.004

72. Cooper, E. E., Biederman, I., & Hummel, J. E. (1992). Metric invariance in object recognition: A review and further evidence. *Canadian Journal of Psychology/Revue Canadienne de Psychologie*, *46*(2), 191–214. https://doi.org/10.1037/h0084317

73. Cooper, R. P., Ruh, N., & Mareschal, D. (2014). The goal circuit model: A hierarchical multi-route model of the acquisition and control of routine sequential action in humans. *Cognitive Science*, *38*(2), 244–274. https://doi.org/10.1111/cogs.12067

74. Cooper, R. P., & Shallice, T. (2000). Contention scheduling and the control of routine activities. *Cognitive Neuropsychology*, *17*(4), 297–338.

75. Cooper, R. P., & Shallice, T. (2006). Hierarchical schemas and goals in the control of sequential behavior. *Psychological Review*, *113*(4), 887–916. https://doi.org/10.1037/0033-295X.113.4.887

76. Costantini, M., Committeri, G., & Galati, G. (2008). Effector- and target-independent representation of observed actions: Evidence from incidental repetition priming. *Experimental Brain Research*, *188*(3), 341–351. https://doi.org/10.1007/s00221-008-1369-x

77. Csibra, G. (2008). Action mirroring and action understanding: an alternative account. In P. Haggard, Y. Rossetti, & M. Kawato (Eds.), *Sensorimotor Foundations of Higher Cognition. Attention and Performance XXII.* (Oxford Uni). Oxford: Oxford University Press.

78. D'Angelo, M. C., Thomson, D. R., Tipper, S. P., & Milliken, B. (2016). Negative priming 1985 to 2015: A measure of inhibition, the emergence of alternative accounts, and the multiple process challenge. *The Quarterly Journal of Experimental Psychology*, *0218*(April), 1–47. https://doi.org/10.1080/17470218.2016.1173077

79. D'Innocenzo, G., Gonzalez, C. C., Nowicky, A. V., Williams, A. M., & Bishop, D. T. (2017). Motor resonance during action observation is gaze-contingent: A TMS study. *Neuropsychologia*, *103*(November 2016), 77–86. https://doi.org/10.1016/j.neuropsychologia.2017.07.017

80. Dapretto, M., Davies, M. S., Pfeifer, J. H., Scott, A. A., Sigman, M., Bookheimer, S. Y., & Iacoboni, M. (2006). Understanding emotions in others: Mirror neuron dysfunction in children with autism spectrum disorders. *Nature Neuroscience*, *9*(1), 28–30. https://doi.org/10.1038/nn1611

81. Davare, M., Montague, K., Olivier, E., Rothwell, J. C., & Lemon, R. N. (2009). Ventral premotor to primary motor cortical interactions during object-driven grasp in humans. *Cortex*, *45*(9), 1050–1057. https://doi.org/10.1016/j.cortex.2009.02.011

82. Davis, M. H. (1983). A Mulitdimensional Approach to Individual Differences in Empathy. *Journal of Personality and Social Psychology*, *44*(1), 113–126. https://doi.org/10.1037/0022-3514.44.1.113

83. de Lange, F. P., Spronk, M., Willems, R. M., Toni, I., & Bekkering, H. (2008). Complementary Systems for Understanding Action Intentions. *Current Biology*, *18*(6), 454–

457. https://doi.org/10.1016/j.cub.2008.02.057

84. Decety, J., & Grèzes, J. (2006). The power of simulation: Imagining one's own and other's behavior. *Brain Research*, *1079*(1), 4–14. https://doi.org/10.1016/j.brainres.2005.12.115

85. Decroix, J., & Kalénine, S. (2018). Timing of grip and goal activation during action perception: a priming study. *Experimental Brain Research*, *236*(8), 2411–2426. https://doi.org/10.1007/s00221-018-5309-0

86. Di Pellegrino, G., Fadiga, L., Fogassi, L., Gallese, V., & Rizzolatti, G. (1992). Understanding motor events: a neurophysiological study. *Experimental Brain Research*, 176–180.

87. DiGirolamo, M. A., Simon, J. C., Hubley, K. M., Kopulsky, A., & Gutsell, J. N. (2019). Clarifying the relationship between trait empathy and action-based resonance indexed by EEG mu-rhythm suppression. *Neuropsychologia*, *133*(August), 107172. https://doi.org/10.1016/j.neuropsychologia.2019.107172

88. Dilks, D. D., Julian, J. B., Kubilius, J., Spelke, E. S., & Kanwisher, N. (2011). Mirror-Image Sensitivity and Invariance in Object and Scene Processing Pathways. *Journal of Neuroscience*, *31*(31), 11305–11312. https://doi.org/10.1523/JNEUROSCI.1935-11.2011

89. Dinstein, I., Thomas, C., Behrmann, M., & Heeger, D. J. (2008). Essay A mirror up to nature. *Current Biology*, *18*(1), 13–18.

90. Donaldson, P. H., Gurvich, C., Fielding, J., & Enticott, P. G. (2015). Exploring associations between gaze patterns and putative human mirror neuron system activity. *Frontiers in Human Neuroscience*, *9*(9), 53–58. https://doi.org/10.3389/fnhum.2015.00396

91. Donnarumma, F., Costantini, M., Ambrosini, E., Friston, K., & Pezzulo, G. (2017). Action perception as hypothesis testing. *Cortex*, *89*, 45–60. https://doi.org/10.1016/j.cortex.2017.01.016

92. Enticott, P. G. (2015). Toward a functional account of the human mirror system. Comment on "Grasping synergies: A motor-control approach to the mirror neuron mechanism" by A. D'Ausilio et al. *Physics of Life Reviews*, *12*, 104–105. https://doi.org/10.1016/j.plrev.2015.01.018

93. Evans, J. S. B. T. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annual Review of Psychology*, *59*, 255–278. https://doi.org/10.1146/annurev.psych.59.103006.093629

94. Evans, J. S. B. T., & Stanovich, K. E. (2013). Dual-Process Theories of Higher Cognition: Advancing the Debate. *Perspectives on Psychological Science*, *8*(3), 223–241. https://doi.org/10.1177/1745691612460685

95. Fadiga, L., Fogassi, L., Pavesi, G., & Rizzolatti, G. (1995). Motor facilitation during action observation: a magnetic stimulation study. *Journal of Neurophysiology*, *73*(6), 2608–2611. https://doi.org/10.1152/jn.1995.73.6.2608

96. Feurra, M., Bianco, G., Polizzotto, N. R., Innocenti, I., Rossi, A., & Rossi, S. (2011). Cortico-Cortical Connectivity between Right Parietal and Bilateral Primary Motor Cortices during Imagined and Observed Actions: A Combined TMS/tDCS Study. *Frontiers in Neural Circuits*, *5*(August), 1–9. https://doi.org/10.3389/fncir.2011.00010

97. Field, A., Miles, J., & Field, Z. (2012). *Discovering statistics using R* (Sage publi).

98. Flanagan, J. R., & Johansson, R. S. (2003). Action plans used in action observation. *Nature*, *424*(6950), 769–771. https://doi.org/10.1038/nature01861

99. Flanagan, J. R., Rotman, G., Reichelt, A. F., & Johansson, R. S. (2013). The role of observers' gaze behaviour when watching object manipulation tasks: predicting and evaluating the consequences of action. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *368*(1628), 20130063. https://doi.org/10.1098/rstb.2013.0063

100. Fogassi, L., Ferrari, P. F., Gesierich, B., Rozzi, S., Chersi, F., & Rizzolatti, G. (2005).

Parietal Lobe: From Action Organization to Intention Understanding. *Science*, *308*(5722), 662–667. https://doi.org/10.1126/science.1106138

101. Fontana, A. P., Kilner, J. M., Rodrigues, E. C., Joffily, M., Nighoghossian, N., Vargas, C. D., & Sirigu, A. (2012). Role of the parietal cortex in predicting incoming actions. *NeuroImage*, *59*(1), 556–564. https://doi.org/10.1016/j.neuroimage.2011.07.046

102. Fox, N. A., Bakermans-Kranenburg, M. J., Yoo, K. H., Bowman, L. C., Cannon, E. N., Vanderwert, R. E., … Van Ijzendoorn, M. H. (2016). Assessing Human Mirror Activity With EEG Mu Rhythm: A Meta-Analysis. *Psychological Bulletin*, *142*(3), 291–313. https://doi.org/10.1037/bul0000031

103. Frings, C., Schneider, K. K., & Fox, E. (2015). The negative priming paradigm: An update and implications for selective attention. *Psychonomic Bulletin & Review*, (APRIL), 1–21. https://doi.org/10.3758/s13423-015-0841-4

104. Galinsky, A. D., Magee, J. C., Inesi, M. E., & Gruenfeld, D. H. (2006). Power and perspectives not taken. *Psychological Science*, *17*(12), 1068–1074. https://doi.org/10.1111/j.1467-9280.2006.01824.x

105. Gallagher, H. L., & Frith, C. D. (2004). Dissociable neural pathways for the perception and recognition of expressive and instrumental gestures. *Neuropsychologia*, *42*(13), 1725–1736. https://doi.org/10.1016/j.neuropsychologia.2004.05.006

106. Gallagher, S. (2008). Direct perception in the intersubjective context. *Consciousness and Cognition*, *17*(2), 535–543. https://doi.org/10.1016/j.concog.2008.03.003

107. Gallese, V. (2005). Embodied simulation: From neurons to phenomenal experience. *Phenomenology and the Cognitive Sciences*, *4*(1), 23–48. https://doi.org/10.1007/s11097-005-4737-z

108. Gallese, V., Fadiga, L., Fogassi, L., & Rizzolatti, G. (1996). Action Recognition in the premotor cortex. *Brain*, (119), 593–609.

109. Gallese, V., Rochat, M., Cossu, G., & Sinigaglia, C. (2009). Motor cognition and its role in the phylogeny and ontogeny of action understanding. *Developmental Psychology*, *45*(1), 103–113. https://doi.org/10.1037/a0014436

110. Gangitano, M., Mottaghy, F. M., & Pascual-Leone, A. (2001). Phase-specific modulation of cortical motor output during movement observation. *Neuroreport*, *12*(7), 1489–1492. https://doi.org/10.1097/00001756-200105250-00038

111. Gazzola, V., Aziz-Zadeh, L., & Keysers, C. (2006). Empathy and the Somatotopic Auditory Mirror System in Humans. *Current Biology*, *16*(18), 1824–1829. https://doi.org/10.1016/j.cub.2006.07.072

112. Geangu, E., Senna, I., Croci, E., & Turati, C. (2015). The effect of biomechanical properties of motion on infants' perception of goal-directed grasping actions. *Journal of Experimental Child Psychology*, *129*, 55–67. https://doi.org/10.1016/j.jecp.2014.08.005

113. Giglio, A. C. A., Minati, L., & Boggio, P. S. (2013). Throwing the banana away and keeping the peel: Neuroelectric responses to unexpected but physically feasible action endings. *Brain Research*, *1532*, 56–62. https://doi.org/10.1016/j.brainres.2013.08.017

114. Gilaie-Dotan, S., Silvanto, J., Schwarzkopf, D., & Rees, G. (2010). Investigating representations of facial identity in human ventral visual cortex with transcranial magnetic stimulation. *Frontiers in Human Neuroscience*, *4*(June), 1–11. https://doi.org/10.3389/fnhum.2010.00050

115. Gilet, A. L., Mella, N., Studer, J., Griihn, D., & Labouvie-Vief, G. (2013). Assessing dispositional empathy in Adults: A french validation of the interpersonal reactivity index (IRI). *Canadian Journal of Behavioural Science*, *45*(1), 42–48. https://doi.org/10.1U37/a0030425

116. Godard, M., Wamain, Y., & Kalénine, S. (2019). Do manufactured and natural objects evoke similar motor information? The case of action priming. *Quarterly Journal of Experimental Psychology*, 174702181986221. https://doi.org/10.1177/1747021819862210

117. Gollwitzer, P. M. (2012). Mindset theory of action phases. *Handbook of Theories of Social Psychology: Volume 1*, 526–546. https://doi.org/10.4135/9781446249215.n26

118. Gonzalez Rothi, L. J., Ochipa, C., & Heilman, K. M. (1991). A Cognitive Neuropsychological Model of Limb Praxis. *Cognitive Neuropsychology*, *8*(6), 443–458. https://doi.org/10.1080/02643299108253382

119. Grafton, S. T., & Hamilton, A. F. D. C. (2007). Evidence for a distributed hierarchy of action representation in the brain. *Human Movement Science*, *26*(4), 590–616. https://doi.org/10.1016/j.humov.2007.05.009

120. Gratton, G., Coles, M. G. H., & Donchin, E. (1983). A new method for off-line removal of ocular artifact. *Electroencephalography and Clinical Neurophysiology*, *55*(4), 468–484. https://doi.org/10.1016/0013-4694(83)90135-9

121. 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

122. Gronau, Q. F., Erp, S. Van, Heck, D. W., Cesario, J., & Jonas, K. (2017). A Bayesian Model-Averaged Meta-Analysis of the Power Pose Effect with Informed and Default Priors : The Case of Felt Power. *Comprehensive Results in Social Psychology*, (283876), 1–20.

123. Hamilton, A., & Grafton, S. (2006). Goal Representation in Human Anterior Intraparietal Sulcus. *Journal of Neuroscience*, *26*(4), 1133–1137. https://doi.org/10.1523/JNEUROSCI.4551-05.2006

124. Hamilton, A., & Grafton, S. (2008). Action Outcomes Are Represented in Human Inferior Frontoparietal Cortex. *Cerebral Cortex*, *18*(5), 1160–1168. https://doi.org/10.1093/cercor/bhm150

125. Hayes, A. F. (2009). Beyond Baron and Kenny: Statistical mediation analysis in the new millennium. *Communication Monographs*, *76*(4), 408–420. https://doi.org/10.1080/03637750903310360

126. Heider, F., & Simmel, M. (1944). An Experimental Study of Apparent Behavior. *The American Journal of Psychology*, *57*(2), 243–259.

127. Heyes, C. (2010). Where do mirror neurons come from? *Neuroscience and Biobehavioral Reviews*, *34*(4), 575–583. https://doi.org/10.1016/j.neubiorev.2009.11.007

128. Heyes, C. (2014). Submentalizing: I Am Not Really Reading Your Mind. *Perspectives on Psychological Science*, *9*(2), 131–143. https://doi.org/10.1177/1745691613518076

129. Hickok, G. (2009). Eight Problems for the Mirror Neuron Theory of Action Understanding in Monkeys and Humans. *Journal of Cognitive Neuroscience*, *21*(7), 1229–1243. https://doi.org/10.1162/jocn.2009.21189

130. Hommel, B. (2004). Event files: Feature binding in and across perception and action. *Trends in Cognitive Sciences*, *8*(11), 494–500. https://doi.org/10.1016/j.tics.2004.08.007

131. Hommel, B. (2014). The role of action semantics in action control. Comment on “Action semantics: A unifying conceptual framework for the selective use of multimodal and modality-specific object knowledge” by Michiel van Elk, Hein van Schie and Harold Bekkering. *Physics of Life Reviews*, *11*(2), 259–260. https://doi.org/10.1016/j.plrev.2014.01.010

132. Hrkać, M., Wurm, M. F., & Schubotz, R. I. (2014). Action observers implicitly expect actors to act goal-coherently, even if they do not: An fMRI study. *Human Brain Mapping*, *35*(5), 2178–2190. https://doi.org/10.1002/hbm.22319

133. Hudson, M., Bach, P., & Nicholson, T. (2018). You said you would! The predictability of other’s behavior from their intentions determines predictive biases in action perception. *Journal of Experimental Psychology: Human Perception and Performance*, *44*(2), 320–335. https://doi.org/10.1037/xhp0000451

134. Hudson, M., Nicholson, T., Ellis, R., & Bach, P. (2016). I see what you say: Prior knowledge of other’s goals automatically biases the perception of their actions. *Cognition*, *146*,

245–250. https://doi.org/10.1016/j.cognition.2015.09.021

135. Hudson, M., Nicholson, T., Simpson, W. A., Ellis, R., & Bach, P. (2016). One step ahead: The perceived kinematics of others' actions are biased toward expected goals. *Journal of Experimental Psychology: General*, *145*(1), 1–7. https://doi.org/10.1037/xge0000126

136. Humphreys, G. W., Kumar, S., Yoon, E. Y., Wulff, M., Roberts, K. L., & Riddoch, M. J. (2013). Attending to the possibilities of action. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *368*(1628), 20130059–20130059. https://doi.org/10.1098/rstb.2013.0059

137. Hunnius, S., & Bekkering, H. (2014). What are you doing? How active and observational experience shape infants' action understanding. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *369*, 20130490. https://doi.org/10.1098/rstb.2013.0490

138. Iacoboni, M., Molnar-Szakacs, I., Gallese, V., Buccino, G., & Mazziotta, J. C. (2005). Grasping the intentions of others with one's own mirror neuron system. *PLoS Biology*, *3*(3), 0529–0535. https://doi.org/10.1371/journal.pbio.0030079

139. Jacob, P., & Jeannerod, M. (2005). The motor theory of social cognition: A critique. *Trends in Cognitive Sciences*, *9*(1), 21–25. https://doi.org/10.1016/j.tics.2004.11.003

140. Jacquet, P. O., & Avenanti, A. (2015). Perturbing the action observation network during perception and categorization of actions' goals and grips: State-dependency and virtual lesion TMS effects. *Cerebral Cortex*, *25*(3), 598–608. https://doi.org/10.1093/cercor/bht242

141. Jeannerod, M. (1994). The representing brain: Neural correlates of motor intention and imagery. *Behavioral and Brain Sciences*, *17*(2), 187–202. https://doi.org/10.1017/S0140525X00034026

142. Jeannerod, M. (1999). The 25th Bartlett Lecture: To act or not to Act: Perspectives on the Representation of Actions. *The Quarterly Journal of Experimental Psychology A*, *52*(1), 1–29. https://doi.org/10.1080/027249899391205

143. Jeannerod, M. (2009). *Le cerveau volontaire* (Odile Jaco).

144. Judd, C. M., Westfall, J., & Kenny, D. A. (2017). Experiments with more than one random factor: Designs, analytic models, and statistical power. *Annual Review of Psychology*, *68*(1), 601–625. https://doi.org/10.1146/annurev-psych-122414-033702

145. Kalénine, S., & Buxbaum, L. J. (2016). Thematic knowledge, artifact concepts, and the left posterior temporal lobe: Where action and object semantics converge. *Cortex*, *82*, 164–178. https://doi.org/10.1016/j.cortex.2016.06.008

146. Kalénine, S., Buxbaum, L. J., & Coslett, H. B. (2010). Critical brain regions for action recognition: Lesion symptom mapping in left hemisphere stroke. *Brain*, *133*(11), 3269–3280. https://doi.org/10.1093/brain/awq210

147. Kalénine, S., Mirman, D., Middleton, E. L., & Buxbaum, L. J. (2012). Temporal dynamics of activation of thematic and functional knowledge during conceptual processing of manipulable artifacts. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(5), 1274–1295. https://doi.org/10.1037/a0027626

148. Kalénine, S., Shapiro, A. D., & Buxbaum, L. J. (2013). Dissociations of action means and outcome processing in left-hemisphere stroke. *Neuropsychologia*, *51*(7), 1224–1233. https://doi.org/10.1016/j.neuropsychologia.2013.03.017

149. Kalénine, S., Wamain, Y., Decroix, J., & Coello, Y. (2016). Conflict between object structural and functional affordances in peripersonal space. *Cognition*, *155*, 1–7. https://doi.org/10.1016/j.cognition.2016.06.006

150. Kehrer, S., Kraft, A., Koch, S. P., Kathmann, N., Irlbacher, K., & Brandt, S. A. (2015). Timing of spatial priming within the fronto-parietal attention network: A TMS study. *Neuropsychologia*, *74*, 30–36. https://doi.org/10.1016/j.neuropsychologia.2014.11.017

151. Kennedy, D. P., & Adolphs, R. (2012). The social brain in psychiatric and neurological

disorders. *Trends in Cognitive Sciences*, *16*(11), 559–572. https://doi.org/10.1016/j.tics.2012.09.006

152. Kenward, M. G., & Roger, J. H. (1997). Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics*, *53*(3), 983–997. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9333350

153. Keysers, C., & Gazzola, V. (2006). Chapter 21 Towards a unifying neural theory of social cognition. *Progress in Brain Research*, *156*, 379–401. https://doi.org/10.1016/S0079-6123(06)56021-2

154. Kilner, J. M. (2011). More than one pathway to action understanding. *Trends in Cognitive Sciences*, *15*(8), 352–357. https://doi.org/10.1016/j.tics.2011.06.005

155. Kilner, J. M., Friston, K. J., & Frith, C. D. (2007). Predictive coding: An account of the mirror neuron system. *Cognitive Processing*, *8*(3), 159–166. https://doi.org/10.1007/s10339-007-0170-2

156. Kilner, J. M., & Frith, C. D. (2008). Action observation: Inferring intentions without mirror neurons. *Current Biology*, *18*(1), R32–R33. https://doi.org/10.1016/j.cub.2007.11.008

157. Kilner, J. M., & Lemon, R. N. (2013). What we know currently about mirror neurons. *Current Biology*, *23*(23), R1057–R1062. https://doi.org/10.1016/j.cub.2013.10.051

158. Kiverstein, J., & Miller, M. (2015). The embodied brain: towards a radical embodied cognitive neuroscience. *Frontiers in Human Neuroscience*, *9*(May), 1–11. https://doi.org/10.3389/fnhum.2015.00237

159. Koch, G., Versace, V., Bonnì, S., Lupo, F., Gerfo, E. Lo, Oliveri, M., & Caltagirone, C. (2010). Resonance of cortico-cortical connections of the motor system with the observation of goal directed grasping movements. *Neuropsychologia*, *48*(12), 3513–3520. https://doi.org/10.1016/j.neuropsychologia.2010.07.037

160. Kothari, R., Bokariya, P., Singh, S., & Singh, R. (2016). A Comprehensive Review on Methodologies Employed for Visual Evoked Potentials. *Scientifica*, *2016*, 1–9. https://doi.org/10.1155/2016/9852194

161. Koul, A., Soriano, M., Tversky, B., Becchio, C., & Cavallo, A. (2019). The kinematics that you do not expect: Integrating prior information and kinematics to understand intentions. *Cognition*, *182*(October), 213–219. https://doi.org/10.1016/j.cognition.2018.10.006

162. Kovács, G., Zimmer, M., Bankó, É., Harza, I., Antal, A., & Vidnyánszky, Z. (2006). Electrophysiological correlates of visual adaptation to faces and body parts in humans. *Cerebral Cortex*, *16*(5), 742–753. https://doi.org/10.1093/cercor/bhj020

163. Kozak, M. N., Marsh, A. a, & Wegner, D. M. (2006). What do i think you're doing? Action identification and mind attribution. *Journal of Personality and Social Psychology*, *90*(4), 543–555. https://doi.org/10.1037/0022-3514.90.4.543

164. Kumar, S., Riddoch, M. J., & Humphreys, G. (2013). Mu rhythm desynchronization reveals motoric influences of hand action on object recognition. *Frontiers in Human Neuroscience*, *7*(March), 66. https://doi.org/10.3389/fnhum.2013.00066

165. Kumar, S., Yoon, E. Y., & Humphreys, G. W. (2012). Perceptual and motor-based responses to hand actions on objects: Evidence from ERPs. *Experimental Brain Research*, *220*(2), 153–164. https://doi.org/10.1007/s00221-012-3126-4

166. Kutas, M., & Federmeier, K. D. (2011). Thirty Years and Counting: Finding Meaning in the N400 Component of the Event-Related Brain Potential (ERP). *Annual Review of Psychology*. https://doi.org/10.1146/annurev.psych.093008.131123

167. Lang, N., Siebner, H. R., Ernst, D., Nitsche, M. A., Paulus, W., Lemon, R. N., & Rothwell, J. C. (2004). Preconditioning with transcranial direct current stimulation sensitizes the motor cortex to rapid-rate transcranial magnetic stimulation and controls the direction of after-effects. *Biological Psychiatry*, *56*(9), 634–639. https://doi.org/10.1016/j.biopsych.2004.07.017

168. Lashley, K. S. (1951). The problem of serial order in behavior. In Bobbs-Merrill (Ed.), *The problem of serial order in behavior*. Retrieved from http://ling75.arts.ubc.ca/ling508/readings/Other/Lashley1951.pdf
169. Lautrey, J. (2003). La psychologie différentielle à l'épreuve de la variabilité intraindividuelle. In A. Vom Hofe, H. Charvin, J.-L. Bernaud, & D. Guédon (Eds.), *Psychologie differentielle: recherches et rélexions.* (PUR, pp. 9–28). Rennes.
170. Lee, C., Middleton, E., Mirman, D., Kalénine, S., & Buxbaum, L. J. (2013). Incidental and context-responsive activation of structure- and function-based action features during object identification. *Journal of Experimental Psychology: Human Perception and Performance, 39*(1), 257–270. https://doi.org/http://dx.doi.org/10.1037/a0027533
171. Lee, C., Mirman, D., & Buxbaum, L. J. (2014). Abnormal dynamics of activation of object use information in apraxia: Evidence from eyetracking. *Neuropsychologia*, *59*(1), 13–26. https://doi.org/10.1016/j.neuropsychologia.2014.04.004
172. Leek, E. C., Atherton, C. J., & Thierry, G. (2007). Computational mechanisms of object constancy for visual recognition revealed by event-related potentials. *Vision Research*, *47*(5), 706–713. https://doi.org/10.1016/j.visres.2006.10.021
173. Leonetti, A., Puglisi, G., Siugzdaite, R., Ferrari, C., Cerri, G., & Borroni, P. (2015). What you see is what you get: motor resonance in peripheral vision. *Experimental Brain Research*, *233*(10), 3013–3022. https://doi.org/10.1007/s00221-015-4371-0
174. Lepage, J. F., Tremblay, S., & Théoret, H. (2010). Early non-specific modulation of corticospinal excitability during action observation. *European Journal of Neuroscience*, *31*(5), 931–937. https://doi.org/10.1111/j.1460-9568.2010.07121.x
175. Lewkowicz, D., Quesque, F., Coello, Y., & Delevoye-Turrell, Y. N. (2015). Individual differences in reading social intentions from motor deviants. *Frontiers in Psychology*, *6*(August), 1–12. https://doi.org/10.3389/fpsyg.2015.01175
176. Leys, C., Ley, C., Klein, O., Bernard, P., & Licata, L. (2013). Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology*, *49*(4), 764–766. https://doi.org/10.1016/j.jesp.2013.03.013
177. Lhermitte, F. (1983). Utilization behaviour and its relation to lesions of the frontal lobes. *Brain*, *106*(2), 237–255. https://doi.org/10.1093/brain/106.2.237
178. Lillard, A. S. (2017). Why Do the Children (Pretend) Play? *Trends in Cognitive Sciences*, *21*(11), 826–834. https://doi.org/10.1016/j.tics.2017.08.001
179. Lingnau, A., & Downing, P. E. (2015). The lateral occipitotemporal cortex in action. *Trends in Cognitive Sciences*, *19*(5), 268–277. https://doi.org/10.1016/j.tics.2015.03.006
180. Logothetis, N. K. (2008). What we can do and what we cannot do with fMRI. *Nature*, *453*(7197), 869–878. https://doi.org/10.1038/nature06976
181. Loucks, J., & Baldwin, D. (2009). Sources of information for discriminating dynamic human actions. *Cognition*, *111*(1), 84–97. https://doi.org/10.1016/j.cognition.2008.12.010
182. Loucks, J., & Pechey, M. (2016). Human Action Perception is Consistent, Flexible, and Orientation Dependent. *Perception*, *45*(11), 1222–1239. https://doi.org/10.1177/0301006616652054
183. Loucks, J., & Sommerville, J. A. (2013). Attending to what matters: Flexibility in adults' and infants' action perception. *Journal of Experimental Child Psychology*, *116*(4), 856–872. https://doi.org/10.1016/j.jecp.2013.08.001
184. Luck, S. J. (2014). *An Introduction to the Event-Related Potential Technique*. Cambridge: The MIT Press.
185. Luck, S. J., & Gaspelin, N. (2017). How to get statistically significant effects in any ERP experiment (and why you shouldn't). *Psychophysiology*, *54*(1), 146–157. https://doi.org/10.1111/psyp.12639
186. Luke, S. G. (2017). Evaluating significance in linear mixed-effects models in R.

Behavior Research Methods, *49*(4), 1494–1502. https://doi.org/10.3758/s13428-016-0809-y

187. Maffongelli, L., Bartoli, E., Sammler, D., Kölsch, S., Campus, C., Olivier, E., … D'Ausilio, A. (2015). Distinct brain signatures of content and structure violation during action observation. *Neuropsychologia*, *75*(August 2015), 30–39. https://doi.org/10.1016/j.neuropsychologia.2015.05.020

188. Maffongelli, L., D'Ausilio, A., Fadiga, L., & Daum, M. M. (2019). The ontogenesis of action syntax. *Collabra: Psychology*, *5*(1), 1–11. https://doi.org/10.1525/collabra.215

189. Manera, V., Becchio, C., Cavallo, A., Sartori, L., & Castiello, U. (2011). Cooperation or competition? Discriminating between social intentions by observing prehensile movements. *Experimental Brain Research*, *211*(3–4), 547–556. https://doi.org/10.1007/s00221-011-2649-4

190. Manera, V., Becchio, C., Schouten, B., Bara, B. G., & Verfaillie, K. (2011). Communicative interactions improve visual detection of biological motion. *PLoS ONE*, *6*(1). https://doi.org/10.1371/journal.pone.0014594

191. Marneweck, M., & Vallence, A.-M. (2015). The neural bases of different levels of action understanding. *Journal of Neurophysiology*, (1088295), jn.00819.2015. https://doi.org/10.1152/jn.00819.2015

192. Marsh, A. A., Kozak, M. N., Wegner, D. M., Reid, M. E., Yu, H. H., & Blair, R. J. R. (2010). The neural substrates of action identification. *Social Cognitive and Affective Neuroscience*, *5*(4), 392–403. https://doi.org/10.1093/scan/nsq004

193. 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*(2013), 305–315. https://doi.org/10.1016/j.jml.2017.01.001

194. Memon, M. A., Hwa, C. J., Ramayah, T., & Ting, H. (2018). Mediation Analysis : Issues and Recommendations. *Journal of Applied Structural Equation Modeling*, *2*(January).

195. Miall, R. C. (2003). Connecting mirror neurons and forward models. *NeuroReport*, *14*(17), 2135–2137. https://doi.org/10.1097/00001756-200312020-00001

196. Michael, J., Sandberg, K., Skewes, J., Wolf, T., Blicher, J., Overgaard, M., & Frith, C. D. (2014). Continuous theta-burst stimulation demonstrates a causal role of premotor homunculus in action understanding. *Psychological Science*, *25*(4), 963–972. https://doi.org/10.1177/0956797613520608

197. Miller, G. A., Gration, G., & Yee, C. M. (1988). Generalized Implementation of an Eye Movement Correction Procedure. *Psychophysiology*, *25*(2), 241–243. https://doi.org/10.1111/j.1469-8986.1988.tb00999.x

198. Miniussi, C., Harris, J. A., & Ruzzoli, M. (2013). Modelling non-invasive brain stimulation in cognitive neuroscience. *Neuroscience and Biobehavioral Reviews*, *37*(8), 1702–1712. https://doi.org/10.1016/j.neubiorev.2013.06.014

199. Mirman, D. (2014). *Growth Curve Analysis and Visualization Using R*. https://doi.org/978-1-4665-8433-4

200. Mirman, D., Dixon, J. A., & Magnuson, J. S. (2008). Statistical and computational models of the visual world paradigm: Growth curves and individual differences. *Journal of Memory and Language*, *59*(4), 475–494. https://doi.org/10.1016/j.jml.2007.11.006

201. Mirman, D., & Magnuson, J. S. (2009). Dynamics of activation of semantically similar concepts during spoken word recognition. *Memory and Cognition*, *37*(7), 1026–1039. https://doi.org/10.3758/MC.37.7.1026

202. Mirman, D., Strauss, T. J., Dixon, J. A., & Magnuson, J. S. (2010). Effect of representational distance between meanings on recognition of ambiguous spoken words. *Cognitive Science*, *34*(1), 161–173. https://doi.org/10.1111/j.1551-6709.2009.01069.x

203. Molenberghs, P., Cunnington, R., & Mattingley, J. B. (2012). Brain regions with mirror properties: A meta-analysis of 125 human fMRI studies. *Neuroscience and Biobehavioral Reviews*, *36*(1), 341–349. https://doi.org/10.1016/j.neubiorev.2011.07.004

204. Morgan, A. (2014). Representations gone mental. *Synthese*, *191*(2), 213–244. https://doi.org/10.1007/s11229-013-0328-7

205. Mudrik, L., Lamy, D., & Deouell, L. Y. (2010). ERP evidence for context congruity effects during simultaneous object-scene processing. *Neuropsychologia*, *48*(2), 507–517. https://doi.org/10.1016/j.neuropsychologia.2009.10.011

206. Mukamel, R., Ekstrom, A. D., Kaplan, J., Iacoboni, M., & Fried, I. (2010). Single-Neuron Responses in Humans during Execution and Observation of Actions. *Current Biology*, *20*(8), 750–756. https://doi.org/10.1016/j.cub.2010.02.045

207. Muthukumaraswamy, S. D., & Singh, K. D. (2008). Modulation of the human mirror neuron system during cognitive activity. *Psychophysiology*, *45*(6), 896–905. https://doi.org/10.1111/j.1469-8986.2008.00711.x

208. Naish, K. R., Houston-Price, C., Bremner, A. J., & Holmes, N. P. (2014). Effects of action observation on corticospinal excitability: Muscle specificity, direction, and timing of the mirror response. *Neuropsychologia*, *64*, 331–348. https://doi.org/10.1016/j.neuropsychologia.2014.09.034

209. Naish, K. R., Reader, A. T., Houston-Price, C., Bremner, A. J., & Holmes, N. P. (2013). To eat or not to eat? Kinematics and muscle activity of reach-to-grasp movements are influenced by the action goal, but observers do not detect these differences. *Experimental Brain Research*, *225*(2), 261–275. https://doi.org/10.1007/s00221-012-3367-2

210. Natraj, N., Pella, Y. M., Borghi, A. M., & Wheaton, L. A. (2015). The visual encoding of tool-object affordances. *Neuroscience*, *310*(October), 512–527. https://doi.org/10.1016/j.neuroscience.2015.09.060

211. Natraj, N., Poole, V., Mizelle, J. C., Flumini, A., Borghi, A. M., & Wheaton, L. A. (2013). Context and hand posture modulate the neural dynamics of tool–object perception. *Neuropsychologia*, *51*(3), 506–519. https://doi.org/10.1016/j.neuropsychologia.2012.12.003

212. Nicholson, T., Roser, M., & Bach, P. (2017). Understanding the goals of everyday instrumental actions is primarily linked to object, not motor-kinematic, information: Evidence from fMRI. *PLoS ONE*, *12*(1), 1–21. https://doi.org/10.1371/journal.pone.0169700

213. Novack, M. A., & Goldin-Meadow, S. (2016). Gesture as representational action: A paper about function. *Psychonomic Bulletin & Review*, 652–665. https://doi.org/10.3758/s13423-016-1145-z

214. Novack, M. A., Wakefield, E. M., & Goldin-Meadow, S. (2016). What makes a movement a gesture? *Cognition*, *146*, 339–348. https://doi.org/10.1016/j.cognition.2015.10.014

215. Ocampo, B., & Kritikos, A. (2011). Interpreting actions: The goal behind mirror neuron function. *Brain Research Reviews*, *67*(1–2), 260–267. https://doi.org/10.1016/j.brainresrev.2011.03.001

216. Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*, *9*(1), 97–113. https://doi.org/10.1016/0028-3932(71)90067-4

217. Ondobaka, S., & Bekkering, H. (2013). Conceptual and perceptuo-motor action control and action recognition. *Cortex*, *49*(10), 2966–2967. https://doi.org/10.1016/j.cortex.2013.06.005

218. Ondobaka, S., de Lange, F. P., Wittmann, M., Frith, C. D., & Bekkering, H. (2015). Interplay Between Conceptual Expectations and Movement Predictions Underlies Action Understanding. *Cerebral Cortex*, (September), 2566–2573. https://doi.org/10.1093/cercor/bhu056

219. Ortigue, S., Sinigaglia, C., Rizzolatti, G., & Grafton, S. T. (2010). Understanding actions of others: The electrodynamics of the left and right hemispheres. A high-density EEG neuroimaging study. *PLoS ONE*, *5*(8). https://doi.org/10.1371/journal.pone.0012160

220. Ortigue, S., Thompson, J. C., Parasuraman, R., & Grafton, S. T. (2009). Spatio-temporal

dynamics of human intention understanding in temporo-parietal cortex: A combined EEG/fMRI repetition suppression paradigm. *PLoS ONE*, *4*(9). https://doi.org/10.1371/journal.pone.0006962

221. Osiurak, F., & Badets, A. (2016). Tool use and affordance: Manipulation-based versus reasoning-based approaches. *Psychological Review*, *123*(5), 534–568. https://doi.org/10.1037/rev0000027

222. Oztop, E., Wolpert, D., & Kawato, M. (2005). Mental state inference using visual control parameters. *Cognitive Brain Research*, *22*(2), 129–151. https://doi.org/10.1016/j.cogbrainres.2004.08.004

223. Pacherie, E. (2008). The phenomenology of action: A conceptual framework. *Cognition*, *107*(1), 179–217. https://doi.org/10.1016/j.cognition.2007.09.003

224. Pascual-Leone, A., Bartres-Faz, D., & Keenan, J. (1999). Transcranial magnetic stimulation: studying the brain-behaviour. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *354*(1387), 1229–1238. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1692644/

225. Peelen, M. V, & Downing, P. E. (2007). The neural basis of visual body perception. *Nature Reviews. Neuroscience*, *8*(8), 636–648. https://doi.org/10.1038/nrn2195

226. Pek, J., & Flora, D. B. (2018). Reporting effect sizes in original psychological research: A discussion and tutorial. *Psychological Methods*, *23*(2), 208–225. https://doi.org/10.1037/met0000126

227. Perry, A., & Bentin, S. (2009). Mirror activity in the human brain while observing hand movements: A comparison between EEG desynchronization in the μ-range and previous fMRI results. *Brain Research*, *1282*, 126–132. https://doi.org/10.1016/j.brainres.2009.05.059

228. Perry, A., Troje, N. F., & Bentin, S. (2010). Exploring motor system contributions to the perception of social information: Evidence from EEG activity in the mu/alpha frequency range. *Social Neuroscience*, *5*(3), 272–284. https://doi.org/10.1080/17470910903395767

229. Pezzulo, G. (2011). Grounding Procedural and Declarative Knowledge in Sensorimotor Anticipation. *Mind and Language*, *26*(1), 78–114. https://doi.org/10.1111/j.1468-0017.2010.01411.x

230. Pfeifer, J. H., Iacoboni, M., Mazziotta, J. C., & Dapretto, M. (2008). Mirroring others' emotions relates to empathy and interpersonal competence in children. *NeuroImage*, *39*(4), 2076–2085. https://doi.org/10.1016/j.neuroimage.2007.10.032

231. Piccinini, G. (2018). Computation and Representation in Cognitive Neuroscience. *Minds and Machines*, *28*(1), 1–6. https://doi.org/10.1007/s11023-018-9461-x

232. Pobric, G., & Hamilton, A. F. de C. (2006). Action understanding requires the left inferior frontal cortex. *Current Biology*, *16*(5), 524–529. https://doi.org/10.1016/j.cub.2006.01.033

233. Pomiechowska, B., & Csibra, G. (2017). Motor activation during action perception depends on action interpretation. *Neuropsychologia*, *105*(October 2016), 84–91. https://doi.org/10.1016/j.neuropsychologia.2017.01.032

234. Proverbio, A. M., & Riva, F. (2009). RP and N400 ERP components reflect semantic violations in visual processing of human actions. *Neuroscience Letters*, *459*(3), 142–146. https://doi.org/10.1016/j.neulet.2009.05.012

235. Proverbio, A. M., Riva, F., & Zani, A. (2010). When neurons do not mirror the agent's intentions: Sex differences in neural coding of goal-directed actions. *Neuropsychologia*, *48*(5), 1454–1463. https://doi.org/10.1016/j.neuropsychologia.2010.01.015

236. Quesque, F., & Coello, Y. (2015). Perceiving what you intend to do from what you do: evidence for embodiment in social interactions. *Socioaffective Neuroscience & Psychology*, *5*, 28602. https://doi.org/10.3402/snp.v5.28602

237. Quesque, F., Lewkowicz, D., Delevoye-Turrell, Y. N., & Coello, Y. (2013). Effects of

social intention on movement kinematics in cooperative actions. *Frontiers in Neurorobotics*, *7*(October), 14. https://doi.org/10.3389/fnbot.2013.00014

238. Quinn, S., Donnelly, S., & Kidd, E. (2018). The relationship between symbolic play and language acquisition: A meta-analytic review. *Developmental Review*, *49*(May 2017), 121–135. https://doi.org/10.1016/j.dr.2018.05.005

239. R Core Team. (2019). *R: A language and environment for statistical computing.* Retrieved from https://www.r-project.org/

240. Renoult, L., Wang, X., Calcagno, V., Prévost, M., & Debruille, J. B. (2012). From N400 to N300: Variations in the timing of semantic processing with repetition. *NeuroImage*, *61*(1), 206–215. https://doi.org/10.1016/j.neuroimage.2012.02.069

241. Riach, M., Holmes, P. S., Franklin, Z. C., & Wright, D. J. (2018). Observation of an action with a congruent contextual background facilitates corticospinal excitability: A combined TMS and eye-tracking experiment. *Neuropsychologia*, *119*(August), 157–164. https://doi.org/10.1016/j.neuropsychologia.2018.08.002

242. Rizzolatti, G., & Craighero, L. (2004). the Mirror-Neuron System. *Annual Review of Neuroscience*, *27*(1), 169–192. https://doi.org/10.1146/annurev.neuro.27.070203.144230

243. Rizzolatti, G., Fadiga, L., Gallese, V., & Fogassi, L. (1996). Premotor cortex and the recognition of motor actions. *Cognitive Brain Research*, *3*(2), 131–141. https://doi.org/10.1016/0926-6410(95)00038-0

244. Rizzolatti, G., & Fogassi, L. (2014). The mirror mechanism: recent findings and perspectives. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *369*(1644), 20130420. https://doi.org/10.1098/rstb.2013.0420

245. Rizzolatti, G., & Sinigaglia, C. (2008). *Les Neurones miroirs* (Odile Jaco). Paris.

246. Romaiguère, P., Nazarian, B., Roth, M., Anton, J. L., & Felician, O. (2014). Lateral occipitotemporal cortex and action representation. *Neuropsychologia*, *56*(1), 167–177. https://doi.org/10.1016/j.neuropsychologia.2014.01.006

247. Rossini, P. M., Burke, D., Chen, R., Cohen, L. G., Daskalakis, Z., Di Iorio, R., … Ziemann, U. (2015). Non-invasive electrical and magnetic stimulation of the brain, spinal cord, roots and peripheral nerves: Basic principles and procedures for routine clinical and research application: An updated report from an I.F.C.N. Committee. *Clinical Neurophysiology*, *126*(6), 1071–1107. https://doi.org/10.1016/j.clinph.2015.02.001

248. Rossion, B., Gauthier, I., Tarr, M. J., Despland, P., Bruyer, R., Linotte, S., & Crommelinck, M. (2003). The N170 occipito-temporal component is delayed and enhanced to inverted faces but not to inverted objects. *NeuroReport*, *11*(1), 69–72. https://doi.org/10.1097/00001756-200001170-00014

249. Rossion, B., & Jacques, C. (2008). Does physical interstimulus variance account for early electrophysiological face sensitive responses in the human brain? Ten lessons on the N170. *NeuroImage*, *39*(4), 1959–1979. https://doi.org/10.1016/j.neuroimage.2007.10.011

250. Rossion, B., Joyce, C. A., Cottrell, G. W., & Tarr, M. J. (2003). Early lateralization and orientation tuning for face, word, and object processing in the visual cortex. *NeuroImage*, *20*(3), 1609–1624. https://doi.org/10.1016/j.neuroimage.2003.07.010

251. Ruggiero, M., & Catmur, C. (2018). Mirror neurons and intention understanding: Dissociating the contribution of object type and intention to mirror responses using electromyography. *Psychophysiology*, *55*(7), 1–8. https://doi.org/10.1111/psyp.13061

252. Sartori, L., Becchio, C., & Castiello, U. (2011). Cues to intention: The role of movement information. *Cognition*, *119*(2), 242–252. https://doi.org/10.1016/j.cognition.2011.01.014

253. Satterthwaite, F. E. (1941). Synthesis of variance. *Psychometrika*, *6*(5), 309–316. https://doi.org/10.1007/BF02288586

254. Schaafsma, S. M., Pfaff, D. W., Spunt, R. P., & Adolphs, R. (2015). Deconstructing and reconstructing theory of mind. *Trends in Cognitive Sciences*, *19*(2), 65–72.

https://doi.org/10.1016/j.tics.2014.11.007

255. Schenke, K. C., Wyer, N. A., & Bach, P. (2016). The Things You Do: Internal Models of Others' Expected Behaviour Guide Action Observation. *PLOS ONE*, *11*(7), e0158910. https://doi.org/10.1371/journal.pone.0158910

256. Schilbach, L. (2010). A second-person approach to other minds. *Nature Reviews Neuroscience*, *11*(6), 449–449. https://doi.org/10.1038/nrn2805-c1

257. Schmidt, R. A. (1975). A schema theory of discrete motor skill learning. *Psychological Review*, *82*(4), 225–260. https://doi.org/10.1037/h0076770

258. Schubotz, R. I., & von Cramon, D. Y. (2009). The Case of Pretense: Observing Actions and Inferring Goals. *Journal of Cognitive Neuroscience*, *21*(4), 642–653. https://doi.org/10.1162/jocn.2009.21049

259. Schubotz, R. I., Wurm, M. F., Wittmann, M. K., & von Cramon, D. Y. (2014). Objects tell us what action we can expect: dissociating brain areas for retrieval and exploitation of action knowledge during action observation in fMRI. *Frontiers in Psychology*, *5*(JUN), 1–15. https://doi.org/10.3389/fpsyg.2014.00636

260. Schuch, S., Bayliss, A. P., Klein, C., & Tipper, S. P. (2010). Attention modulates motor system activation during action observation: Evidence for inhibitory rebound. *Experimental Brain Research*, *205*(2), 235–249. https://doi.org/10.1007/s00221-010-2358-4

261. Scott, S. H. (2016). A Functional Taxonomy of Bottom-Up Sensory Feedback Processing for Motor Actions. *Trends in Neurosciences*, *39*(8), 512–526. https://doi.org/10.1016/j.tins.2016.06.001

262. Sebanz, N., & Knoblich, G. (2009). Prediction in joint action: What, when, and where. *Topics in Cognitive Science*, *1*(2), 353–367. https://doi.org/10.1111/j.1756-8765.2009.01024.x

263. Sharma, R., Joshi, S., Singh, K. D., & Kumar, A. (2015). Visual evoked potentials: Normative values and gender differences. *Journal of Clinical and Diagnostic Research*, *9*(7), 12–15. https://doi.org/10.7860/JCDR/2015/12764.6181

264. Shin, Y. K., Proctor, R. W., & Capaldi, E. J. (2010). A Review of Contemporary Ideomotor Theory. *Psychological Bulletin*, *136*(6), 943–974. https://doi.org/10.1037/a0020541

265. Siebner, H. R., Hartwigsen, G., Kassuba, T., & Rothwell, J. C. (2009). How does transcranial magnetic stimulation modify neuronal activity in the brain? Implications for studies of cognition. *Cortex*, *45*(9), 1035–1042. https://doi.org/10.1016/j.cortex.2009.02.007

266. Siebner, H. R., Lang, N., Rizzo, V., Nitsche, M. A., Paulus, W., Lemon, R. N., & Rothwell, J. C. (2004). Preconditioning of low-frequency repetitive transcranial magnetic stimulation with transcranial direct current stimulation: evidence for homeostatic plasticity in the human motor cortex. *Journal of Neuroscience*, *24*(13), 3379–3385. https://doi.org/10.1523/JNEUROSCI.5316-03.2004

267. Siegler, R., & Shipley, C. (1995). Variation, selection and cognitive change. In T. Simon & G. Halford (Eds.), *Developing cognitive competence: New approaches to process modeling.* (pp. 31–76). Erlbaum.

268. Silvanto, J., & Cattaneo, Z. (2017). Common framework for "virtual lesion" and state-dependent TMS: The facilitatory/suppressive range model of online TMS effects on behavior. *Brain and Cognition*, *119*(July), 32–38. https://doi.org/10.1016/j.bandc.2017.09.007

269. Silvanto, J., Muggleton, N., & Walsh, V. (2008). State-dependency in brain stimulation studies of perception and cognition. *Trends in Cognitive Sciences*, *12*(12), 447–454. https://doi.org/10.1016/j.tics.2008.09.004

270. Sim, E.-J., Helbig, H. B., Graf, M., & Kiefer, M. (2014). When Action Observation Facilitates Visual Perception: Activation in Visuo-Motor Areas Contributes to Object Recognition. *Cerebral Cortex*, (September), 1–41. https://doi.org/10.1093/cercor/bhu087

271. Sitnikova, T., Holcomb, P. J., Kiyonaga, K. A., & Kuperberg, G. R. (2008). Two neurocognitive mechanisms of semantic integration during the comprehension of visual real-

world events. *Journal of Cognitive Neuroscience*, *20*(11), 2037–2057. https://doi.org/10.1162/jocn.2008.20143

272. Sitnikova, T., Kuperberg, G., & Holcomb, P. J. (2003). Semantic integration in videos of real-world events: An electrophysiological investigation. *Psychophysiology*, *40*(1), 160–164. https://doi.org/10.1111/1469-8986.00016

273. Smith, E. D., Englander, Z. A., Lillard, A. S., & Morris, J. P. (2013). Cortical mechanisms of pretense observation. *Social Neuroscience*, *8*(4), 356–368. https://doi.org/10.1080/17470919.2013.807872

274. Sobel, M. E. (1982). Asymptotic Confidence Intervals for Indirect Effects in Structural Equation Models. *Sociological Methodology*, *13*(1982), 290. https://doi.org/10.2307/270723

275. Sobel, M. E. (1986). Some New Results on Indirect Effects and Their Standard Errors in Covariance Structure Models. *Sociological Methodology*, *16*(May), 159. https://doi.org/10.2307/270922

276. Southgate, V., Begus, K., Lloyd-Fox, S., di Gangi, V., & Hamilton, A. (2014). Goal representation in the infant brain. *NeuroImage*, *85*, 294–301. https://doi.org/10.1016/j.neuroimage.2013.08.043

277. Springer, A., Brandstädter, S., & Prinz, W. (2013). Dynamic simulation and static matching for action prediction: Evidence from body part priming. *Cognitive Science*, *37*(5), 936–952. https://doi.org/10.1111/cogs.12044

278. Springer, A., Parkinson, J., & Prinz, W. (2013). Action simulation: Time course and representational mechanisms. *Frontiers in Psychology*, *4*(JUL), 1–20.

279. Spunt, R. P., Falk, E. B., & Lieberman, M. D. (2010). Dissociable neural systems support retrieval of how and why action knowledge. *Psychological Science : A Journal of the American Psychological Society / APS*, *21*(11), 1593–1598. https://doi.org/10.1177/0956797610386618

280. Spunt, R. P., Kemmerer, D., & Adolphs, R. (2016). The neural basis of conceptualizing the same action at different levels of abstraction. *Social Cognitive and Affective Neuroscience*, *11*(7), 1141–1151. https://doi.org/10.1093/scan/nsv084

281. Spunt, R. P., & Lieberman, M. D. (2012). Dissociating Modality-Specific and Supramodal Neural Systems for Action Understanding. *Journal of Neuroscience*, *32*(10), 3575–3583. https://doi.org/10.1523/JNEUROSCI.5715-11.2012

282. Summers, J. J., & Anson, J. G. (2009). Current status of the motor program: Revisited. *Human Movement Science*, *28*(5), 566–577. https://doi.org/10.1016/j.humov.2009.01.002

283. Thill, S., Caligiore, D., Borghi, A. M., Ziemke, T., & Baldassarre, G. (2013). Theories and computational models of affordance and mirror systems: An integrative review. *Neuroscience and Biobehavioral Reviews*, *37*(3), 491–521. https://doi.org/10.1016/j.neubiorev.2013.01.012

284. Thioux, M., Gazzola, V., & Keysers, C. (2008). Action Understanding: How, What and Why. *Current Biology*, *18*(10), R431–R434. https://doi.org/10.1016/j.cub.2008.03.018

285. Thioux, M., & Keysers, C. (2015). Object visibility alters the relative contribution of ventral visual stream and mirror neuron system to goal anticipation during action observation. *NeuroImage*, *105*, 380–394. https://doi.org/10.1016/j.neuroimage.2014.10.035

286. Thompson, E. L., Bird, G., & Catmur, C. (2019). Conceptualizing and testing action understanding. *Neuroscience and Biobehavioral Reviews*, *105*(August), 106–114. https://doi.org/10.1016/j.neubiorev.2019.08.002

287. Tidoni, E., Borgomaneri, S., di Pellegrino, G., & Avenanti, A. (2013). Action simulation plays a critical role in deceptive action recognition. *Journal of Neuroscience*, *33*(2), 611–623. https://doi.org/10.1523/JNEUROSCI.2228-11.2013

288. Tidoni, E., & Candidi, M. (2016). Commentary: Understanding intentions from actions: Direct perception, inference, and the roles of mirror and mentalizing systems. *Frontiers in*

Behavioral Neurosciences, *10*, 1–3. https://doi.org/10.3389/fnbeh.2016.00013 A

289. Toates, F. (2006). A model of the hierarchy of behaviour, cognition, and consciousness. *Consciousness and Cognition*, *15*(1), 75–118. https://doi.org/10.1016/j.concog.2005.04.008

290. Uithol, S., Burnston, D. C., & Haselager, P. (2014). Why we may not find intentions in the brain. *Neuropsychologia*, *56*(1), 129–139. https://doi.org/10.1016/j.neuropsychologia.2014.01.010

291. Uithol, S., & Paulus, M. (2014). What do infants understand of others' action? A theoretical account of early social cognition. *Psychological Research*, *78*(5), 609–622. https://doi.org/10.1007/s00426-013-0519-3

292. Uithol, S., van Rooij, I., Bekkering, H., & Haselager, P. (2012). Hierarchies in Actions and Motor Control. *Journal of Cognitive Neuroscience*, *24*(5), 1077–1086. https://doi.org/10.1162/jocn

293. Umiltà, M. A., Kohler, E., Gallese, V., Fogassi, L., Fadiga, L., Keysers, C., & Rizzolatti, G. (2001). I know what you are doing: A neurophysiological study. *Neuron*, *31*(1), 155–165. https://doi.org/10.1016/S0896-6273(01)00337-3

294. Urgesi, C., Candidi, M., & Avenanti, A. (2014). Neuroanatomical substrates of action perception and understanding: an anatomic likelihood estimation meta-analysis of lesion-symptom mapping studies in brain injured patients. *Frontiers in Human Neuroscience, 8*(May), 344. https://doi.org/10.3389/fnhum.2014.00344

295. Urgesi, C., Candidi, M., Ionta, S., & Aglioti, S. M. (2007). Representation of body identity and body actions in extrastriate body area and ventral premotor cortex. *Nature Neuroscience*, *10*(1), 30–31. https://doi.org/10.1038/nn1815

296. Valchev, N., Gazzola, V., Avenanti, A., & Keysers, C. (2016). Primary somatosensory contribution to action observation brain activity-combining fMRI and cTBS. *Social Cognitive and Affective Neuroscience*, (January), nsw029. https://doi.org/10.1093/scan/nsw029

297. Vallacher, R. R., & Wegner, D. M. (1987). What do people think they're doing? Action identification and human behavior. *Psychological Review*, *94*(1), 3–15. https://doi.org/10.1037/0033-295X.94.1.3

298. Vallacher, R. R., & Wegner, D. M. (1989). Levels of personal agency: Individual variation in action identification. *Journal of Personality and Social Psychology*, *57*(4), 660–671. https://doi.org/10.1037/0022-3514.57.4.660

299. Vallacher, R. R., & Wegner, D. M. (2012). Action Identification Theory. In *Handbook of Theories of Social Psychology: Volume 1* (pp. 327–348). https://doi.org/10.4135/9781446249215.n17

300. van Elk, M., Bousardt, R., Bekkering, H., & van Schie, H. T. (2012). Using goal- and grip-related information for understanding the correctness of other's actions: An ERP study. *PLoS ONE*, *7*(5), 1–8. https://doi.org/10.1371/journal.pone.0036450

301. van Elk, M., van Schie, H., & Bekkering, H. (2014a). Action semantics: A unifying conceptual framework for the selective use of multimodal and modality-specific object knowledge. *Physics of Life Reviews*, *11*(2), 220–250. https://doi.org/10.1016/j.plrev.2013.11.005

302. van Elk, M., van Schie, H., & Bekkering, H. (2014b). The scope and limits of action semantics. *Physics of Life Reviews*, *11*(2), 273–279. https://doi.org/10.1016/j.plrev.2014.03.009

303. van Elk, M., Van Schie, H. T., & Bekkering, H. (2008). Conceptual knowledge for understanding other's actions is organized primarily around action goals. *Experimental Brain Research*, *189*(1), 99–107. https://doi.org/10.1007/s00221-008-1408-7

304. Van Overwalle, F., & Baetens, K. (2009). Understanding others' actions and goals by mirror and mentalizing systems: A meta-analysis. *NeuroImage*, *48*(3), 564–584. https://doi.org/10.1016/j.neuroimage.2009.06.009

305. Vilarroya, O. (2017). Neural Representation. A Survey-Based Analysis of the Notion. *Frontiers in Psychology*, *8*(AUG), 1–15. https://doi.org/10.3389/fpsyg.2017.01458

306. Walther, D., & Koch, C. (2006). Modeling attention to salient proto-objects. *Neural Networks*, *19*(9), 1395–1407. https://doi.org/10.1016/j.neunet.2006.10.001

307. Wamain, Y., Pluciennicka, E., & Kalénine, S. (2014). Temporal dynamics of action perception: Differences on ERP evoked by object-related and non-object-related actions. *Neuropsychologia*, *63*, 249–258. https://doi.org/10.1016/j.neuropsychologia.2014.08.034

308. Wamain, Y., Pluciennicka, E., & Kalénine, S. (2015). A saw is first identified as an object used on wood: ERP evidence for temporal differences between Thematic and Functional similarity relations. *Neuropsychologia*, *71*, 28–37. https://doi.org/10.1016/j.neuropsychologia.2015.02.034

309. Wamain, Y., Sahaï, A., Decroix, J., Coello, Y., & Kalénine, S. (2017). Conflict between gesture representations extinguishes μ rhythm desynchronization during manipulable object perception: an EEG study. *Biological Psychology*, *132*(June 2017), 202–211. https://doi.org/10.1016/j.biopsycho.2017.12.004

310. West, W. C., & Holcomb, P. J. (2002). Event-related potentials during discourse-level semantic integration of complex pictures. *Cognitive Brain Research*, *13*(3), 363–375. https://doi.org/10.1016/S0926-6410(01)00129-X

311. Westfall, J., Kenny, D. A., & Judd, C. M. (2014). Statistical power and optimal design in experiments in which samples of participants respond to samples of stimuli. *Journal of Experimental Psychology: General*, *143*(5), 2020–2045. https://doi.org/10.1037/xge0000014

312. Westra, E. (2019). Stereotypes, theory of mind, and the action–prediction hierarchy. *Synthese*, *196*(7), 2821–2846. https://doi.org/10.1007/s11229-017-1575-9

313. Wolpert, D., Doya, K., & Kawato, M. (2003). A unifying computational framework for motor control and social interaction. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *358*(1431), 593–602. https://doi.org/10.1098/rstb.2002.1238

314. Wong, A. L., Haith, A. M., & Krakauer, J. W. (2015). Motor planning. *Neuroscientist*, *21*(4), 385–398. https://doi.org/10.1177/1073858414541484

315. Woodruff, C. C., & Klein, S. (2013). Attentional distraction, μ-suppression and empathic perspective-taking. *Experimental Brain Research*, *229*(4), 507–515. https://doi.org/10.1007/s00221-013-3612-3

316. Wright, D. J., Wood, G., Franklin, Z. C., Marshall, B., Riach, M., & Holmes, P. S. (2018). Directing visual attention during action observation modulates corticospinal excitability. *PLoS ONE*, *13*(1), 1–15. https://doi.org/10.1371/journal.pone.0190165

317. Wurm, M. F., Ariani, G., Greenlee, M. W., & Lingnau, A. (2015). Decoding Concrete and Abstract Action Representations During Explicit and Implicit Conceptual Processing. *Cerebral Cortex*, 1–12. https://doi.org/10.1093/cercor/bhv169

318. Wurm, M. F., & Lingnau, A. (2015). Decoding actions at different levels of abstraction. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *35*(20), 7727–7735. https://doi.org/10.1523/JNEUROSCI.0188-15.

319. Wurm, M. F., & Schubotz, R. I. (2012). NeuroImage Squeezing lemons in the bathroom : Contextual information modulates action recognition. *NeuroImage*, *59*, 1551–1559. https://doi.org/10.1016/j.neuroimage.2011.08.038

320. Wurm, M. F., & Schubotz, R. I. (2016). What's she doing in the kitchen? Context helps when actions are hard to recognize. *Psychonomic Bulletin & Review*, 1–7. https://doi.org/10.3758/s13423-016-1108-4

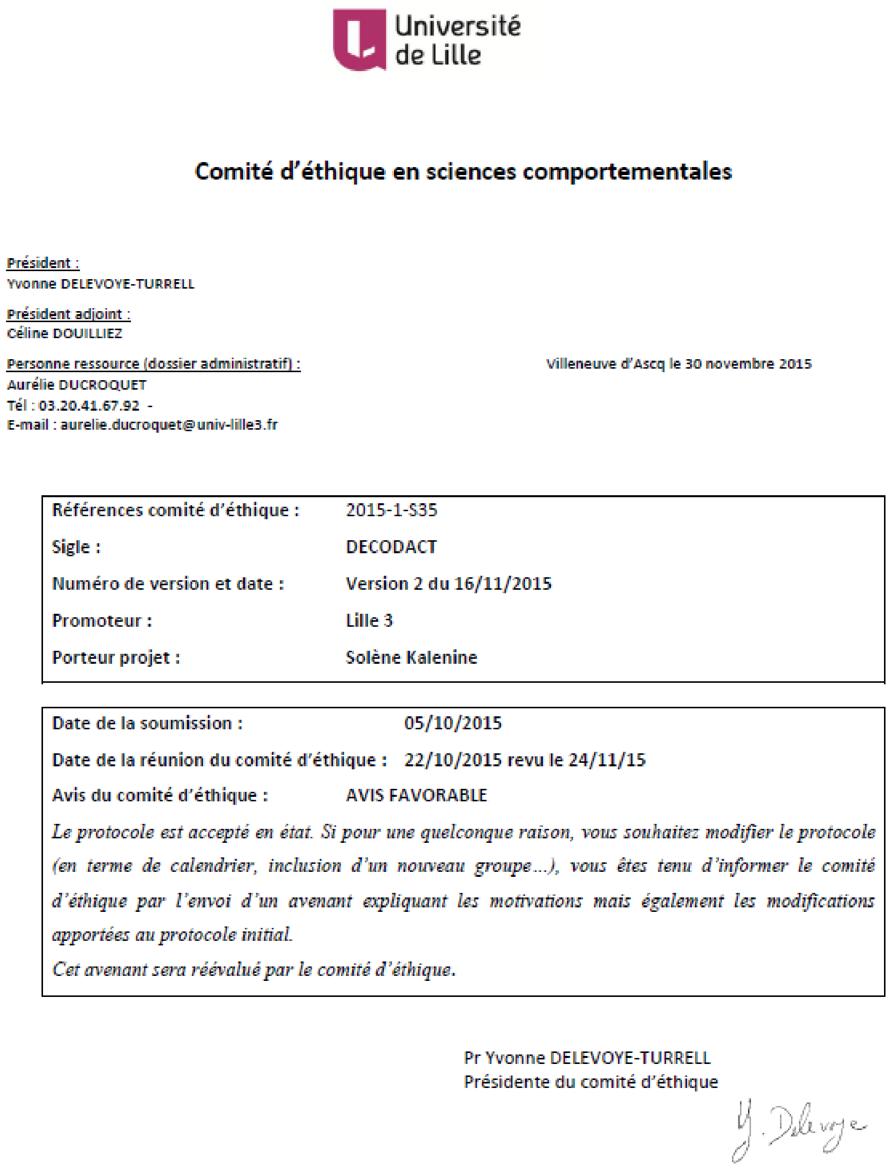
321. Yoon, E. Y., Humphreys, G. W., & Riddoch, M. J. (2010). The paired-object affordance effect. *Journal of Experimental Psychology: Human Perception and Performance*, *36*(4), 812–824. https://doi.org/10.1037/a0017175

322. Zacks, J. M., Tversky, B., & Iyer, G. (2001). Perceiving, remembering, and communicating structure in events. *Journal of Experimental Psychology: General*, *130*(1), 29–58. https://doi.org/10.1037//0096-3445.130.1.29
323. Zentgraf, K., Munzert, J., Bischoff, M., & Newman-Norlund, R. D. (2011). Simulation during observation of human actions - Theories, empirical studies, applications. *Vision Research*, *51*(8), 827–835. https://doi.org/10.1016/j.visres.2011.01.007
324. Zhang, L., Zhang, L., Mou, X., & Zhang, D. (2011). FSIM: A feature similarity index for image quality assessment. *IEEE Transactions on Image Processing*, *8*(20), 2378–2386.
325. Ziemann, U. (2010). TMS in cognitive neuroscience: Virtual lesion and beyond. *Cortex*, *46*(1), 124–127. https://doi.org/10.1016/j.cortex.2009.02.020

Appendices

APPENDIX 1

Ethical protocol (priming studies)



Université
de Lille

Comité d'éthique en sciences comportementales

Président :
Yvonne DELEVOYE-TURRELL

Président adjoint :
Céline DOUILLIEZ

Personne ressource (dossier administratif) :
Aurélie DUCROQUET
Tél : 03.20.41.67.92 -
E-mail : aurelie.ducroquet@univ-lille3.fr

Villeneuve d'Ascq le 30 novembre 2015

Références comité d'éthique :	2015-1-S35
Sigle :	DECODACT
Numéro de version et date :	Version 2 du 16/11/2015
Promoteur :	Lille 3
Porteur projet :	Solène Kalenine

Date de la soumission :	05/10/2015
Date de la réunion du comité d'éthique :	22/10/2015 revu le 24/11/15
Avis du comité d'éthique :	AVIS FAVORABLE

Le protocole est accepté en état. Si pour une quelconque raison, vous souhaitez modifier le protocole (en terme de calendrier, inclusion d'un nouveau groupe…), vous êtes tenu d'informer le comité d'éthique par l'envoi d'un avenant expliquant les motivations mais également les modifications apportées au protocole initial.

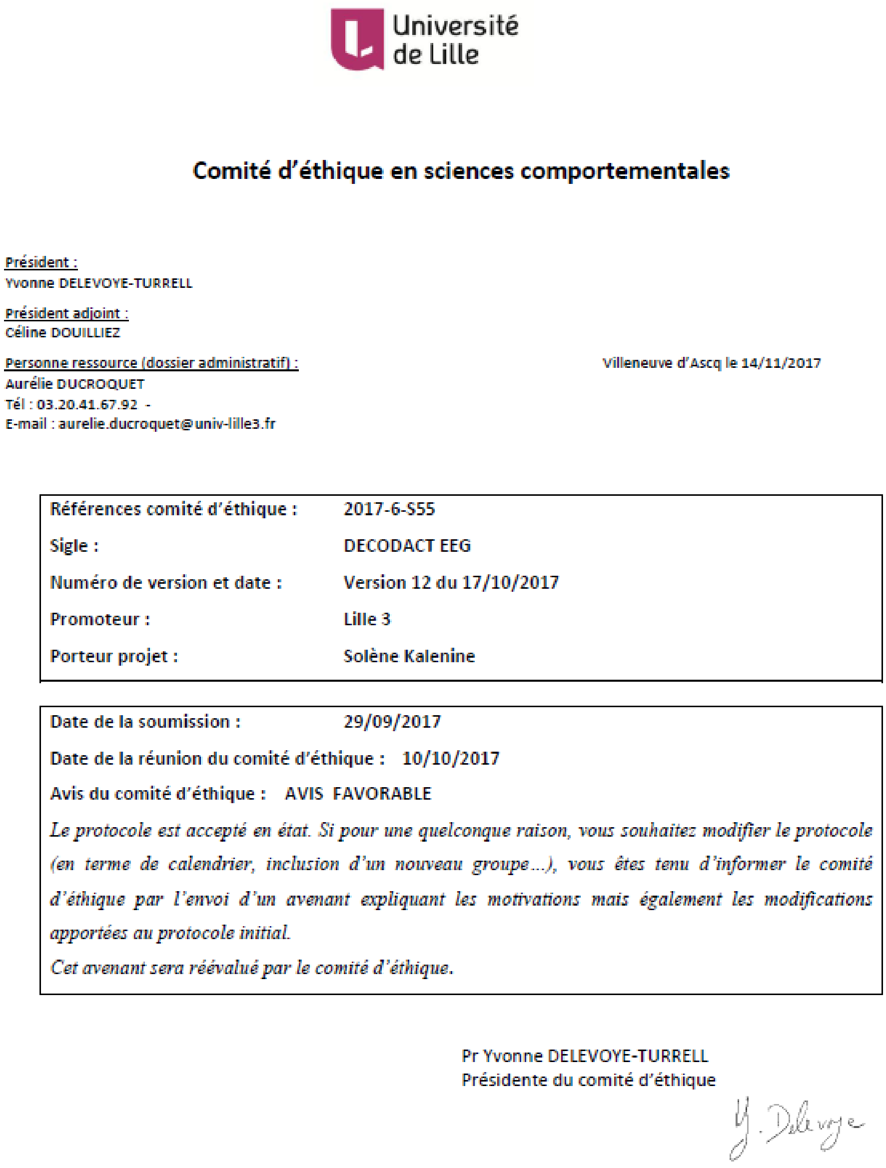
Cet avenant sera réévalué par le comité d'éthique.

Pr Yvonne DELEVOYE-TURRELL
Présidente du comité d'éthique

Y. Delevoye

APPENDIX 2

Ethical protocol (EEG study)



Université
de Lille

Comité d'éthique en sciences comportementales

Président :
Yvonne DELEVOYE-TURRELL

Président adjoint :
Céline DOUILLIEZ

Personne ressource (dossier administratif) :
Aurélie DUCROQUET
Tél : 03.20.41.67.92 -
E-mail : aurelie.ducroquet@univ-lille3.fr

Villeneuve d'Ascq le 14/11/2017

Références comité d'éthique :	2017-6-S55
Sigle :	DECODACT EEG
Numéro de version et date :	Version 12 du 17/10/2017
Promoteur :	Lille 3
Porteur projet :	Solène Kalenine

Date de la soumission : 29/09/2017

Date de la réunion du comité d'éthique : 10/10/2017

Avis du comité d'éthique : AVIS FAVORABLE

Le protocole est accepté en état. Si pour une quelconque raison, vous souhaitez modifier le protocole (en terme de calendrier, inclusion d'un nouveau groupe…), vous êtes tenu d'informer le comité d'éthique par l'envoi d'un avenant expliquant les motivations mais également les modifications apportées au protocole initial.

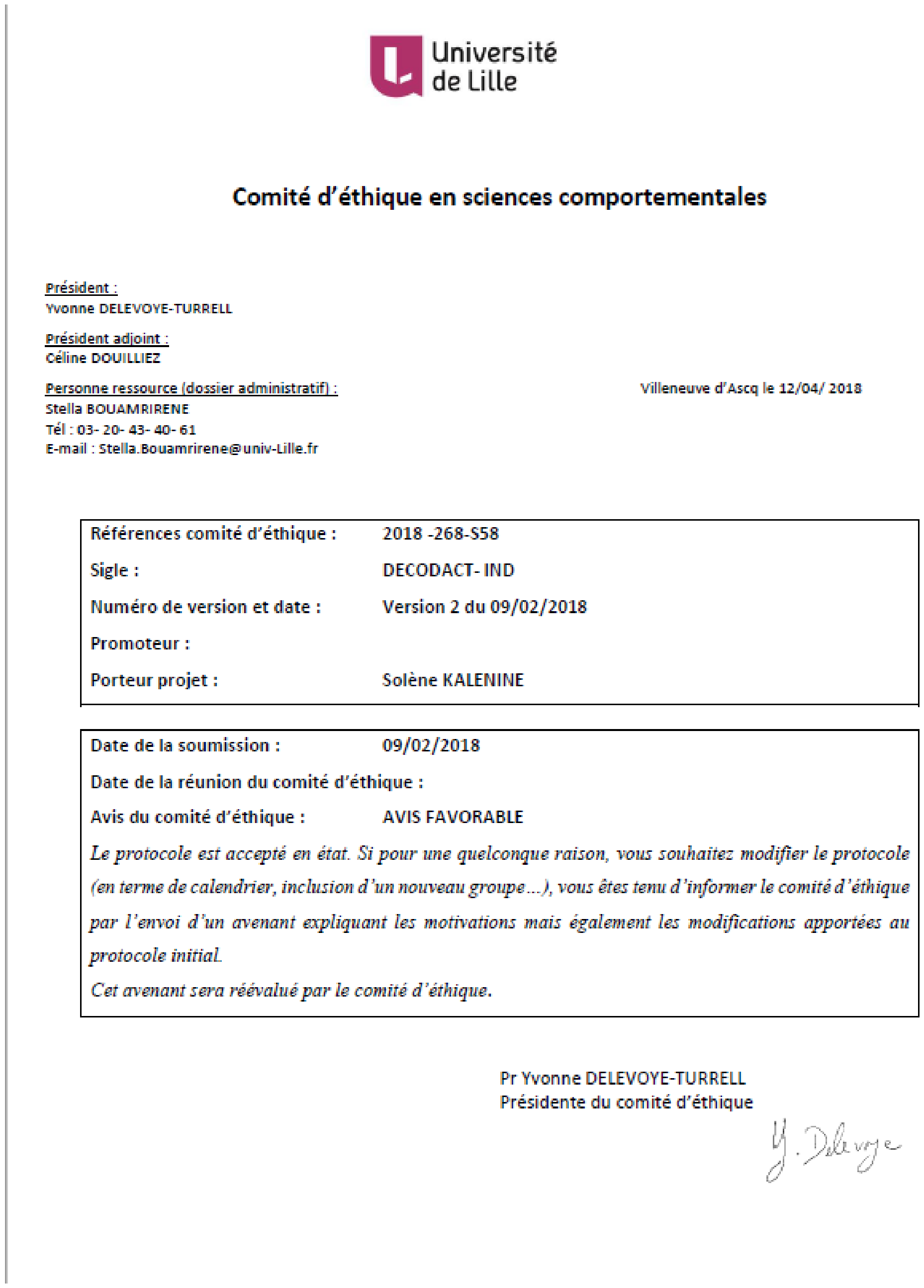
Cet avenant sera réévalué par le comité d'éthique.

Pr Yvonne DELEVOYE-TURRELL
Présidente du comité d'éthique

Y. Delevoye

APPENDIX 3

Ethical protocol (social power study)



Université de Lille

Comité d'éthique en sciences comportementales

Président :
Yvonne DELEVOYE-TURRELL

Président adjoint :
Céline DOUILLIEZ

Personne ressource (dossier administratif) :
Stella BOUAMRIRENE
Tél : 03- 20- 43- 40- 61
E-mail : Stella.Bouamrirene@univ-Lille.fr

Villeneuve d'Ascq le 12/04/ 2018

Références comité d'éthique : 2018 -268-S58
Sigle : DECODACT- IND
Numéro de version et date : Version 2 du 09/02/2018
Promoteur :
Porteur projet : Solène KALENINE

Date de la soumission : 09/02/2018
Date de la réunion du comité d'éthique :
Avis du comité d'éthique : AVIS FAVORABLE

Le protocole est accepté en état. Si pour une quelconque raison, vous souhaitez modifier le protocole (en terme de calendrier, inclusion d'un nouveau groupe…), vous êtes tenu d'informer le comité d'éthique par l'envoi d'un avenant expliquant les motivations mais également les modifications apportées au protocole initial.

Cet avenant sera réévalué par le comité d'éthique.

Pr Yvonne DELEVOYE-TURRELL
Présidente du comité d'éthique

Y. Delevoye

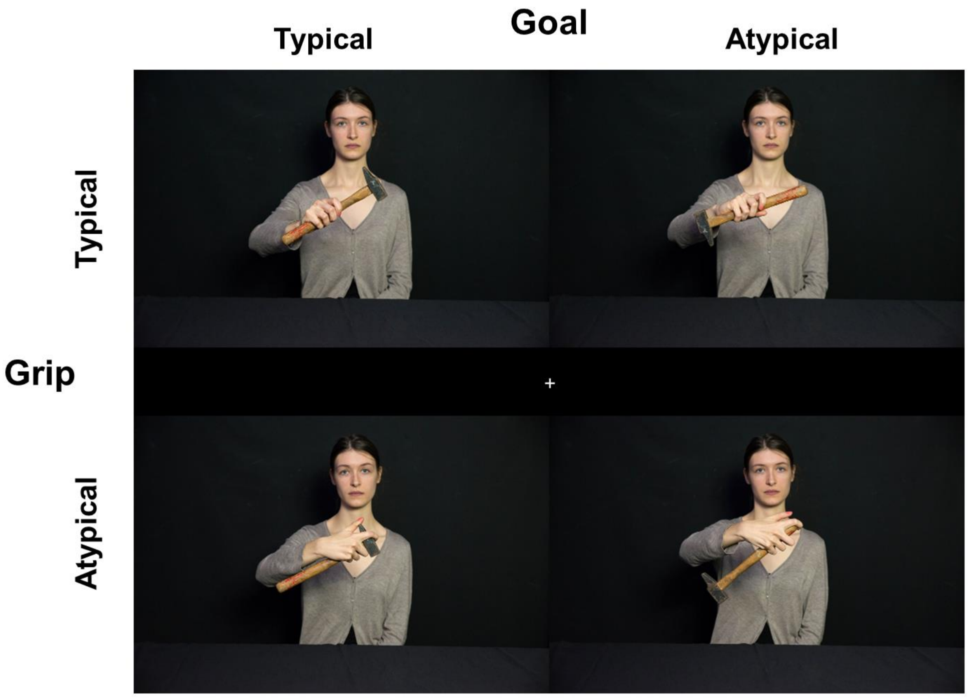
APPENDIX 4

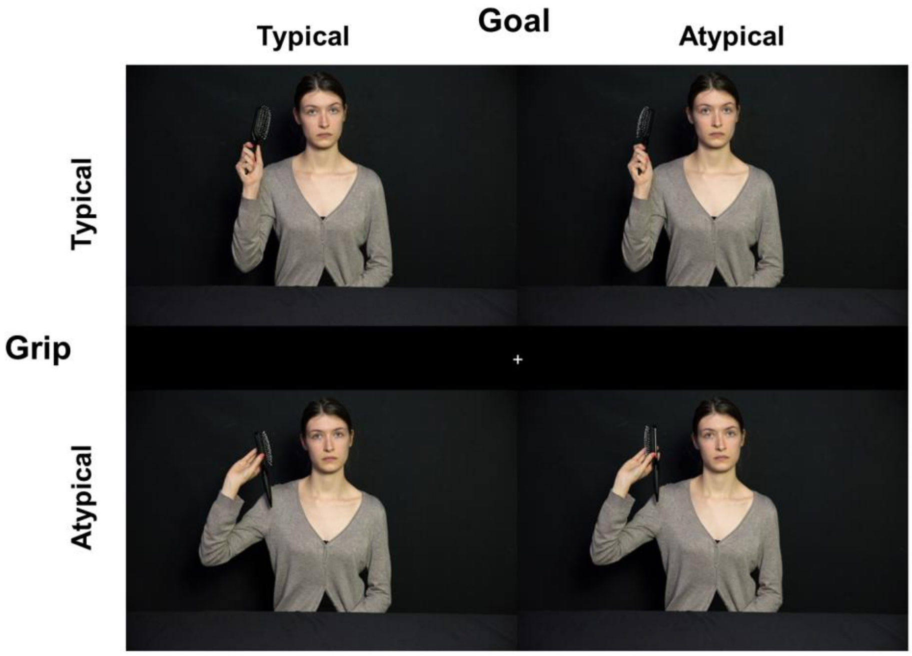
List of stimuli

Set 1	Set 2
Carafe	Bread knife
Coffee cup	Cleaning brush
Fork	Cleaning spray 1
Hairbrush	Cleaning spray 2
Hairdryer	Coffee cup
Hammer	Cooking painbrush
Knife	Cutter
Liquid soap	Deodorant spray
Magnifying glass	Doser
Pencil	Dusting spray
Phone	Hammer
Screwdriver	Hand squeegee
Tea spoon	Highlighter
Teapot	Paintbrush
Toothbrush	Paste stick
Torch	Pencil
Cream Tube	Spatula
Water bottle	Tea pot
Watering can	Whisk
Wine glass	Watering can

APPENDIX 5

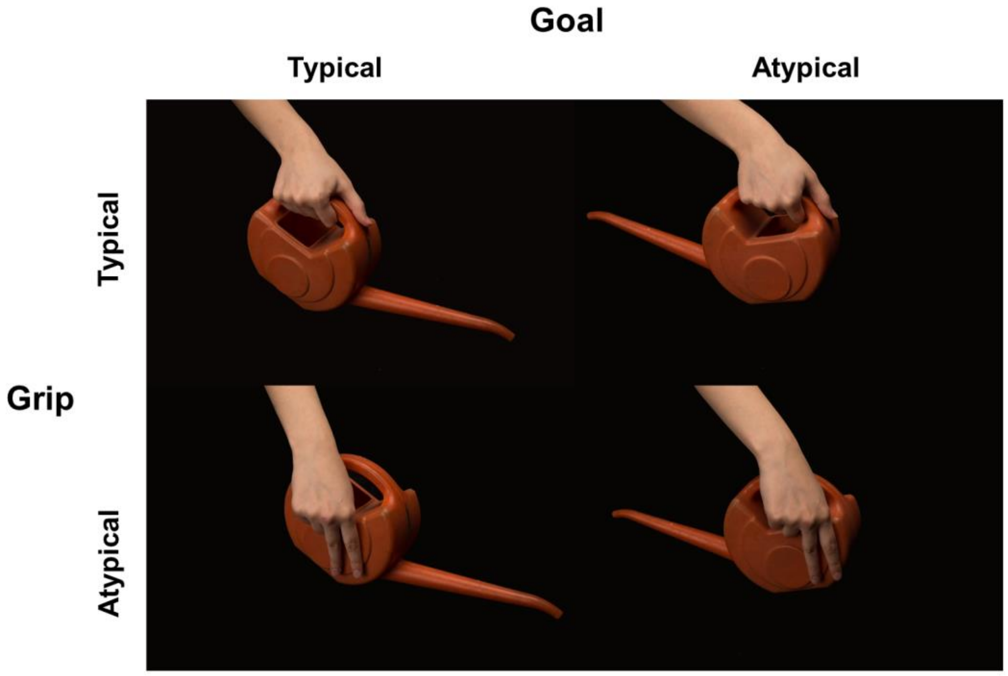
Example of stimuli from Set 1

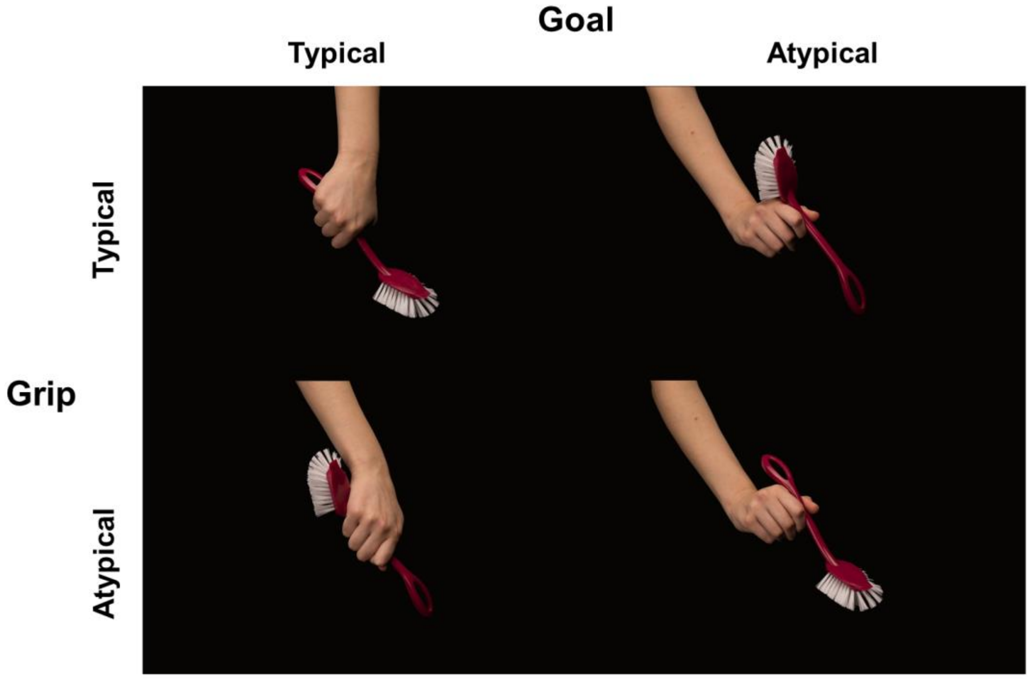




APPENDIX 6

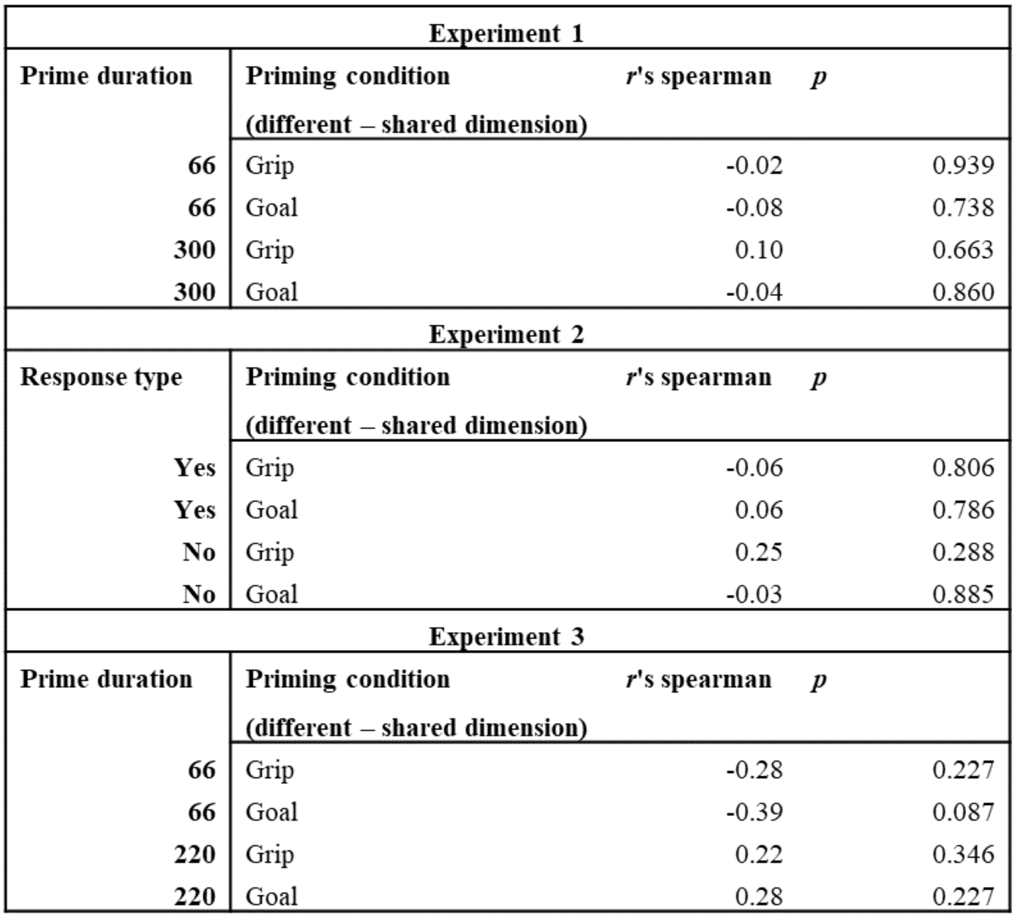
Example of stimuli from Set 2





APPENDIX 7

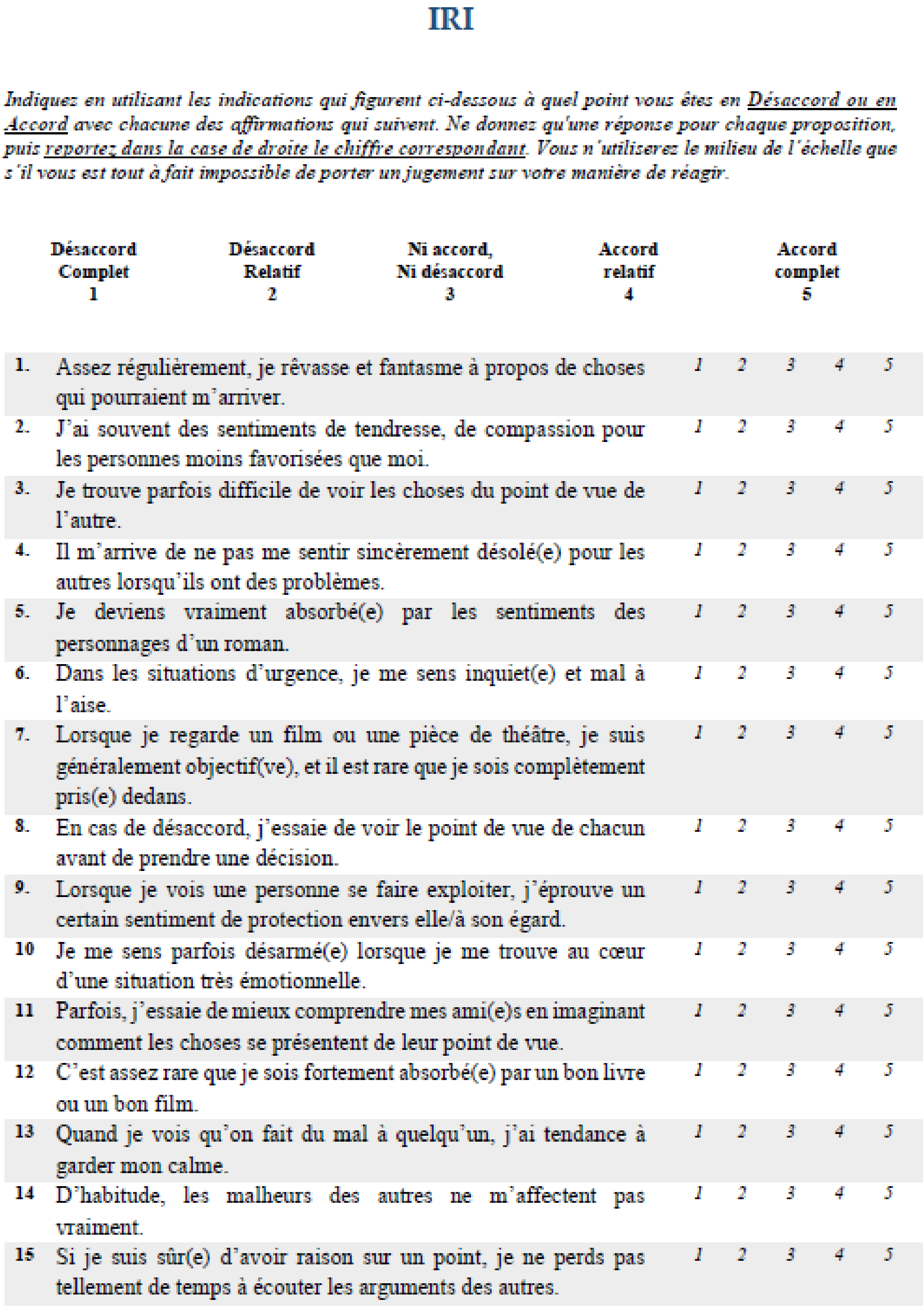
Correlation between the FSIM Index and the significant priming effects obtained in the priming studies



Experiment 1			
Prime duration	**Priming condition (different – shared dimension)**	***r*'s spearman**	***p***
66	Grip	-0.02	0.939
66	Goal	-0.08	0.738
300	Grip	0.10	0.663
300	Goal	-0.04	0.860
Experiment 2			
Response type	**Priming condition (different – shared dimension)**	***r*'s spearman**	***p***
Yes	Grip	-0.06	0.806
Yes	Goal	0.06	0.786
No	Grip	0.25	0.288
No	Goal	-0.03	0.885
Experiment 3			
Prime duration	**Priming condition (different – shared dimension)**	***r*'s spearman**	***p***
66	Grip	-0.28	0.227
66	Goal	-0.39	0.087
220	Grip	0.22	0.346
220	Goal	0.28	0.227

APPENDIX 8

Interpersonal Reactivity Index



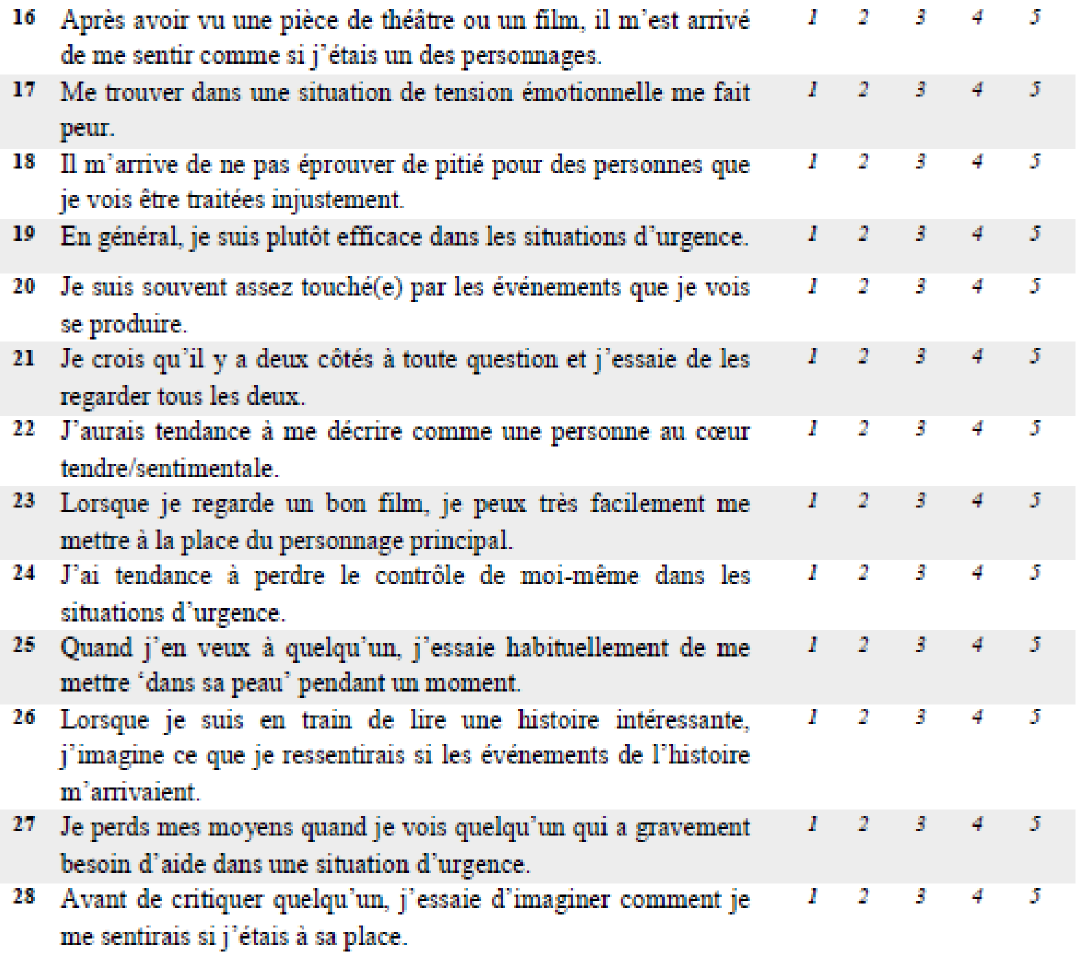
IRI

Indiquez en utilisant les indications qui figurent ci-dessous à quel point vous êtes en Désaccord ou en Accord avec chacune des affirmations qui suivent. Ne donnez qu'une réponse pour chaque proposition, puis reportez dans la case de droite le chiffre correspondant. Vous n'utiliserez le milieu de l'échelle que s'il vous est tout à fait impossible de porter un jugement sur votre manière de réagir.

Désaccord Complet 1	Désaccord Relatif 2	Ni accord, Ni désaccord 3	Accord relatif 4	Accord complet 5

1.	Assez régulièrement, je rêvasse et fantasme à propos de choses qui pourraient m'arriver.	1	2	3	4	5
2.	J'ai souvent des sentiments de tendresse, de compassion pour les personnes moins favorisées que moi.	1	2	3	4	5
3.	Je trouve parfois difficile de voir les choses du point de vue de l'autre.	1	2	3	4	5
4.	Il m'arrive de ne pas me sentir sincèrement désolé(e) pour les autres lorsqu'ils ont des problèmes.	1	2	3	4	5
5.	Je deviens vraiment absorbé(e) par les sentiments des personnages d'un roman.	1	2	3	4	5
6.	Dans les situations d'urgence, je me sens inquiet(e) et mal à l'aise.	1	2	3	4	5
7.	Lorsque je regarde un film ou une pièce de théâtre, je suis généralement objectif(ve), et il est rare que je sois complètement pris(e) dedans.	1	2	3	4	5
8.	En cas de désaccord, j'essaie de voir le point de vue de chacun avant de prendre une décision.	1	2	3	4	5
9.	Lorsque je vois une personne se faire exploiter, j'éprouve un certain sentiment de protection envers elle/à son égard.	1	2	3	4	5
10	Je me sens parfois désarmé(e) lorsque je me trouve au cœur d'une situation très émotionnelle.	1	2	3	4	5
11	Parfois, j'essaie de mieux comprendre mes ami(e)s en imaginant comment les choses se présentent de leur point de vue.	1	2	3	4	5
12	C'est assez rare que je sois fortement absorbé(e) par un bon livre ou un bon film.	1	2	3	4	5
13	Quand je vois qu'on fait du mal à quelqu'un, j'ai tendance à garder mon calme.	1	2	3	4	5
14	D'habitude, les malheurs des autres ne m'affectent pas vraiment.	1	2	3	4	5
15	Si je suis sûr(e) d'avoir raison sur un point, je ne perds pas tellement de temps à écouter les arguments des autres.	1	2	3	4	5

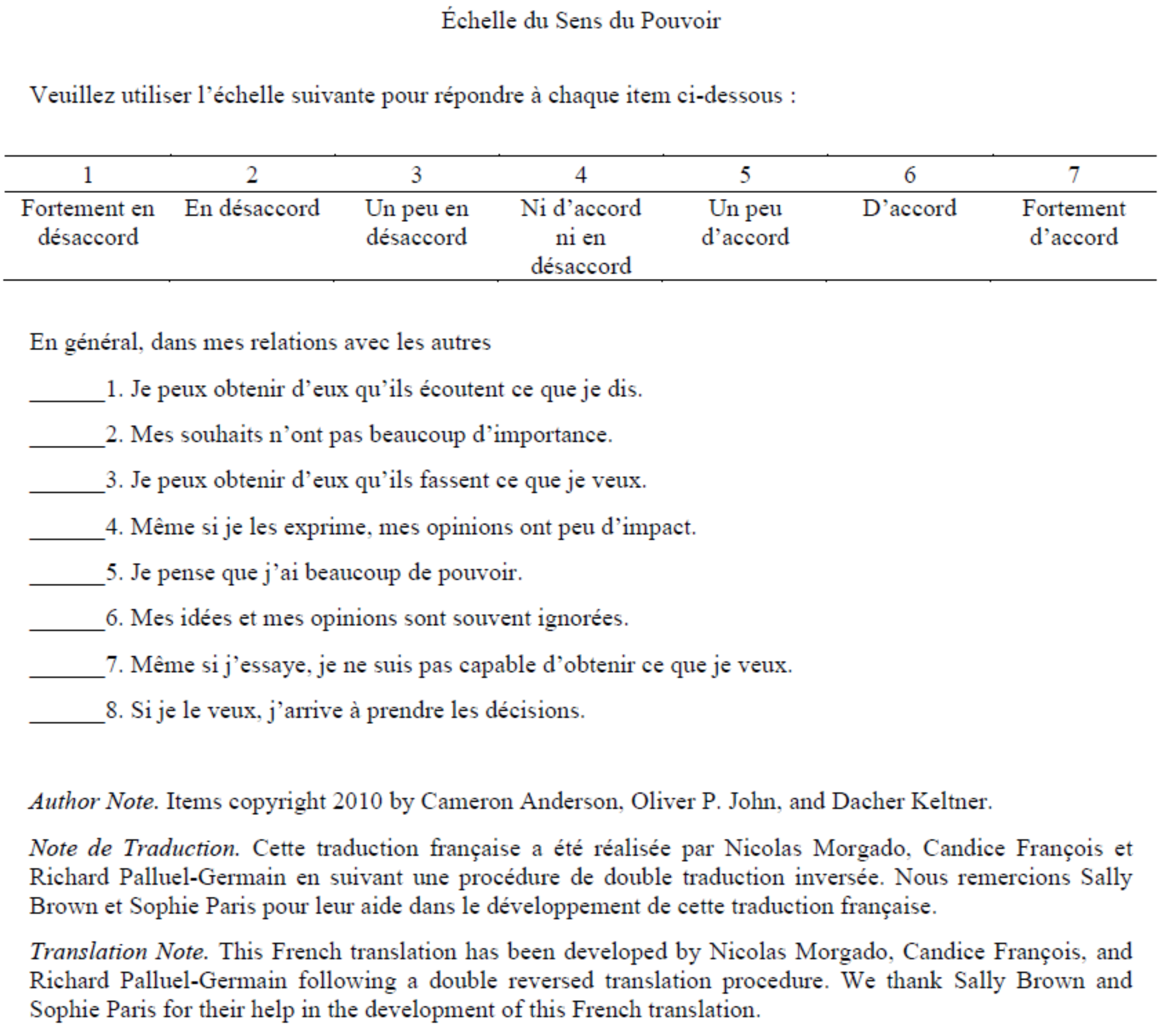
Appendix 8 (continued)



16	Après avoir vu une pièce de théâtre ou un film, il m'est arrivé de me sentir comme si j'étais un des personnages.	*1*	*2*	*3*	*4*	*5*
17	Me trouver dans une situation de tension émotionnelle me fait peur.	*1*	*2*	*3*	*4*	*5*
18	Il m'arrive de ne pas éprouver de pitié pour des personnes que je vois être traitées injustement.	*1*	*2*	*3*	*4*	*5*
19	En général, je suis plutôt efficace dans les situations d'urgence.	*1*	*2*	*3*	*4*	*5*
20	Je suis souvent assez touché(e) par les événements que je vois se produire.	*1*	*2*	*3*	*4*	*5*
21	Je crois qu'il y a deux côtés à toute question et j'essaie de les regarder tous les deux.	*1*	*2*	*3*	*4*	*5*
22	J'aurais tendance à me décrire comme une personne au cœur tendre/sentimentale.	*1*	*2*	*3*	*4*	*5*
23	Lorsque je regarde un bon film, je peux très facilement me mettre à la place du personnage principal.	*1*	*2*	*3*	*4*	*5*
24	J'ai tendance à perdre le contrôle de moi-même dans les situations d'urgence.	*1*	*2*	*3*	*4*	*5*
25	Quand j'en veux à quelqu'un, j'essaie habituellement de me mettre 'dans sa peau' pendant un moment.	*1*	*2*	*3*	*4*	*5*
26	Lorsque je suis en train de lire une histoire intéressante, j'imagine ce que je ressentirais si les événements de l'histoire m'arrivaient.	*1*	*2*	*3*	*4*	*5*
27	Je perds mes moyens quand je vois quelqu'un qui a gravement besoin d'aide dans une situation d'urgence.	*1*	*2*	*3*	*4*	*5*
28	Avant de critiquer quelqu'un, j'essaie d'imaginer comment je me sentirais si j'étais à sa place.	*1*	*2*	*3*	*4*	*5*

APPENDIX 9

Sense of social power scale



Échelle du Sens du Pouvoir

Veuillez utiliser l'échelle suivante pour répondre à chaque item ci-dessous :

1	2	3	4	5	6	7
Fortement en désaccord	En désaccord	Un peu en désaccord	Ni d'accord ni en désaccord	Un peu d'accord	D'accord	Fortement d'accord

En général, dans mes relations avec les autres

______1. Je peux obtenir d'eux qu'ils écoutent ce que je dis.

______2. Mes souhaits n'ont pas beaucoup d'importance.

______3. Je peux obtenir d'eux qu'ils fassent ce que je veux.

______4. Même si je les exprime, mes opinions ont peu d'impact.

______5. Je pense que j'ai beaucoup de pouvoir.

______6. Mes idées et mes opinions sont souvent ignorées.

______7. Même si j'essaye, je ne suis pas capable d'obtenir ce que je veux.

______8. Si je le veux, j'arrive à prendre les décisions.

Author Note. Items copyright 2010 by Cameron Anderson, Oliver P. John, and Dacher Keltner.

Note de Traduction. Cette traduction française a été réalisée par Nicolas Morgado, Candice François et Richard Palluel-Germain en suivant une procédure de double traduction inversée. Nous remercions Sally Brown et Sophie Paris pour leur aide dans le développement de cette traduction française.

Translation Note. This French translation has been developed by Nicolas Morgado, Candice François, and Richard Palluel-Germain following a double reversed translation procedure. We thank Sally Brown and Sophie Paris for their help in the development of this French translation.

APPENDIX 10

Behavioral Identification Questionnaire (BIF)

Information Générales

- DATE :
- AGE :
- SEXE :
- Avez-vous connaissance de ce questionnaire : OUI / NON

Toutes actions/comportements peuvent être décrites de différentes manières. L'action/le comportement d'écrire par exemple peut être décrite comme étant « le fait de tracer des lettres du papier » ou comme « le fait d'exprimer ses idées ». Nous sommes intéressés par vos préférences personnelles quant à la manière dont vous décrivez un certain nombre de comportement/d'action.

Sur les pages suivantes, plusieurs comportements/actions vont vous être proposer. Après chaque comportement/action, il y aura deux choix quant à la manière de décrire un comportement. Voici un exemple :

0. Assister à un cours

	Votre réponse
a. Etre assis sur une chaise	
b. Etre assidu à l'école	

Votre tâche est de choisir la proposition a ou b qui décrit le mieux le comportement selon vous. Si, « Être assis sur une chaise » est la meilleure description pour « Assister à un cours », alors placez un « x » en face de la proposition « a » et laissez vide la case en face de la proposition « b » :

	Votre réponse
a. Etre assis sur une chaise	X
b. Etre assidu à l'école	

Si « Etre assidu à l'école » est la meilleure description pour « Assister à un cours », alors placez un « x » en face de la proposition « b » et laissez vide la case en face de la proposition « a » :

	Votre réponse
a. Etre assis sur une chaise	
b. Etre assidu à l'école	X

Il n'y a pas de bonne ou de mauvaise réponse. Les personnes diffèrent simplement dans la manière dont ils décrivent un comportement, et nous sommes intéressés par vos préférences personnelles. Prenez soin de répondre à toutes les questions.

1. Faire une liste

	Votre réponse
a. S’organiser	
b. Ecrire des choses	

2. Lire

	Votre réponse
a. Suivre des lignes imprimées	
b. Acquérir des connaissances	

3. Rejoindre l’armée

	Votre réponse
a. Aider à la défense de la nation	
b. S’engager	

4. Laver des vêtements

	Votre réponse
a. Eliminer les odeurs des vêtements	
b. Mettre des vêtements dans la machine	

5. Cueillir une pomme sur une branche

	Votre réponse
a. Prendre quelque chose à manger	
b. Détacher une pomme d’une branche	

6. Abattre un arbre

	Votre réponse
a. Manier une hache	
b. Obtenir du bois pour le feu	

7. Prendre les mesures d’une pièce pour refaire la moquette

	Votre réponse
a. Se préparer à redécorer	
b. Utiliser un mètre	

8. Nettoyer la maison

	Votre réponse
a. Faire preuve de propreté	
b. Passer l’aspirateur	

9. Peindre la pièce

	Votre réponse
a. Appliquer des coups de pinceaux	
b. Décorer la pièce	

10. Payer le loyer

	Votre réponse
a. Conserver un endroit où vivre	
b. Faire un chèque/un virement	

11. Prendre soin des plantes d'intérieur

	Votre réponse
a. Arroser les plantes	
b. Garder la pièce agréable	

12. Fermer une porte à clé

	Votre réponse
a. Mettre une clé dans la serrure	
b. Sécuriser la maison	

13. Voter

	Votre réponse
a. Influencer les élections	
b. Remplir un bulletin de vote	

14. Grimper à un arbre

	Votre réponse
a. Avoir une meilleure visibilité	
b. S'accrocher aux branches	

15. Remplir un test de personnalité

	Votre réponse
a. Répondre à des questions	
b. Révéler ce que vous êtes	

16. Se brosser les dents

	Votre réponse
a. Prévenir la dégradation des dents	
b. Déplacer une brosse dans sa bouche	

17. Passer un test

	Votre réponse
a. Répondre à des questions	
b. Evaluer ses connaissances	

18. Saluer quelqu'un

	Votre réponse
a. Dire bonjour	
b. Faire preuve de politesse	

---

19. Résister à la tentation

	Votre réponse
a. Dire “non”	
b. Faire preuve de force de caractère	

20. Manger

	Votre réponse
a. Recevoir un apport nutritif	
b. Mâcher et avaler	

21. Cultiver un jardin

	Votre réponse
a. Semer des graines	
b. Avoir des légumes frais	

22. Voyager en voiture

	Votre réponse
a. Se déplacer	
b. Voir du pays	

23. Faire soigner une carie

	Votre réponse
a. Prendre soin de ses dents	
b. Aller chez le dentiste	

24. Parler à un enfant

	Votre réponse
a. Expliquer des choses à l’enfant	
b. Utiliser des mots simples	

25. Sonner à la porte

	Votre réponse
a. Appuyer avec son doigt sur un bouton	
b. Voir si quelqu’un est à la maison	

APPENDIX 11

Dominance questionaire

Ce questionnaire est une adaptation de l'échelle développée par Cheng et al. (2010).

Veuillez indiquer dans quelle mesure chacune des affirmations vous corresponds en général dans la vie de tous les jours (entourez).

En général,

J'aime avoir du contrôle sur les autres.

Pas du tout - Peu - Modérément - Fortement - Totalement

J'essaie souvent de faire les choses à ma manière, indépendamment de ce que les autres veulent.

Pas du tout - Peu - Modérément - Fortement - Totalement

Je suis prêt(e) à utiliser des techniques agressives pour obtenir ce que je veux.

Pas du tout - Peu - Modérément - Fortement - Totalement

Je n'ai pas une forte personnalité, ou une personnalité dominante.

Pas du tout - Peu - Modérément - Fortement - Totalement

Les autres savent qu'il est plus simple de me laisser faire ce que je veux.

Pas du tout - Peu - Modérément - Fortement - Totalement

Je n'apprécie pas avoir l'autorité sur les autres.

Pas du tout - Peu - Modérément - Fortement - Totalement

Certaines personnes ont peur de moi.

Pas du tout - Peu - Modérément - Fortement - Totalement

APPENDIX 12

Table of the cross-correlation between grip (Gi), goal (Go) priming effects, the index of goal priority, the score of IAT, BIF, Dominance, social power and perspective take. The significant correlations are in bold.

	Gi	Go	Goprio	IATScore	BIF	dominance	power	PT
Gi		0,13	**-0,53**	-0,05	-0,06	0,07	0,20	-0,23
Go			**0,78**	0,10	0,00	-0,13	0,24	-0,07
Goprio				0,11	0,04	-0,15	0,08	0,09
IATScore					0,00	-0,03	0,19	-0,13
BIF						-0,02	-0,03	0,15
dominance							**0,44**	**-0,32**
power								-0,13
PT								

French summary

Préface

“La plupart des gens, quand ils bougent, eh bien ils bougent en fonction de ce qu’il y a autour d’eux. Juste en ce moment, quand j’écris, il y a Constitution qui passe avec le ventre qui traîne par terre. Cette chatte n’a aucun projet construit dans la vie mais elle se dirige vers quelque chose, probablement un fauteuil. Et ça se voit dans sa façon de bouger : elle va vers. Maman vient de passer en direction de la porte d’entrée, elle sort faire des courses et en fait, elle est déjà dehors, son mouvement s’anticipe lui-même.”

L’Elégance du Hérisson, Muriel Barbery

La majeure partie de notre vie revêt un caractère social. La plupart de nos actions n’ont pas d’intérêt si personne n’est là pour les remarquer, les évaluer ou les juger. Un dicton bouddhiste questionne « *Quel bruit fait l’arbre qui tombe lorsque personne ne l’entend* » ? Dans le même temps, nous excellons à remarquer, évaluer et juger les actions de l’autre. Cette activité n’est pas seulement volontaire, elle est aussi spontanée. Il est presque impossible pour nous de ne pas comprendre directement ce que les autres font et pourquoi ils le font. D’une certaine manière, c’est à cette question que cette thèse s’intéresse. Nous nous sommes intéressés ici à quelques-unes des bases qui fondent notre capacité à faire sens du comportement de l’autre, car, avant toute évaluation subjective ou tout jugement de valeur, la reconnaissance de l’action précède.

Pourtant, les actions sont des mouvements complexes, organisés et dirigés vers des buts ; buts qui restent terrés dans l’esprit de l’acteur. Les dernières décennies ont fortement contribué à démontrer que ces buts transparaissent, en fait, dans les mouvements de l’acteur avant même qu’il n’ait fini de bouger ; mais pas seulement, nous anticipons aussi les buts et actions à venir grâce à l’information que nous avons de l’environnement de l’acteur. Dans cette thèse, nous nous sommes d’abord attelés à passer en revue les différents concepts utilisés par la littérature sur la reconnaissance des actions pour définir l’action ; concepts bien souvent empruntés à la littérature sur le contrôle moteur et la planification des actions. Dans les sections qui suivent, nous avons discutés les apports et limites des théories qui tentent d'expliquer notre capacité à reconnaitre les actions d'autrui, ainsi que les données empiriques utilisées pour les défendre. A l'issue de cette discussion, nous concluons qu'améliorer la manière dont nous conceptualisons la reconnaissance d'action nécessite la prise en compte des aspects temporels de ce phénomène ; or ces données manquent cruellement dans la littérature. Après avoir décrit et justifié notre stratégie expérimentale, nous apportons des arguments empiriques dans cette direction. Nous avons tenté de comprendre comment l'observateur, et son cerveau, fait usage des informations visuelles auxquelles il accède. Dans une dernière partie, nous envisageons que les caractéristiques propres à l'observateur influencent la manière dont il utilise ces informations, et donc la manière dont il reconnait les actions d'autrui. Notre objectif était d'apporter des éléments temporels sur la reconnaissance d'action à travers différents angles d'étude : via les mécanismes cognitifs et leur implémentation neuronale mais aussi en s'intéressant à l'impact de la variabilité inhérente au sujet humain.

Comprendre la manière dont nous reconnaissons les actions de l'autre est, je crois, une étape importante avant de pouvoir accéder à une compréhension plus étendue de la cognition sociale humaine. Si le langage reste au cœur de la cognition sociale, de nombreux arguments démontrent que des pans entiers de notre cognition sociale outrepassent la communication linguistique. Tous les niveaux de notre société sont concernés par la cognition sociale, ou du moins, chaque fois que l'humain est impliqué. Je crois que beaucoup des défis rencontrés aujourd'hui par nos sociétés impliquent, dans une moindre mesure, des aspects de notre cognition sociale. Améliorer notre compréhension de la cognition sociale aura d'importante conséquence sur la manière dont nous ferons face à ces défis.

Introduction

La reconnaissance des actions d'autrui est une capacité fondamentale de l'être humain, une compétence essentielle à nos comportements sociaux. Lorsque nous observons une action, nous ne traitons pas passivement l'action à mesure qu'elle se déroule, nous anticipons sa fin ainsi que l'objectif final de l'action, son but (Ansuini, Cavallo, Bertone, & Becchio, 2014; Buresh & Woodward, 2007; Novack & Goldin-Meadow, 2016; Quesque & Coello, 2015; Zacks, Tversky, & Iyer, 2001). Les buts sont souvent définis comme la représentation des conséquences sur l'environnement désiré par l'acteur (Jacob & Jeannerod, 2005; Kilner, 2011; Rizzolatti & Fogassi, 2014; Thill, Caligiore, Borghi, Ziemke, & Baldassarre, 2013), ils sont nécessaires à la production d'une action (Scott, 2016; Wolpert, Doya, & Kawato, 2003), et seraient représentés à différents niveaux de complexités (R. P. Cooper & Shallice, 2000, 2006; Lashley, 1951). Les mouvements ne sont, en conséquence, que la partie visible d'une action (Jeannerod, 1994, 2009). Si les mouvements sont la partie visible de l'action, comment expliquer alors que nous nous représentons les actions d'autrui majoritairement en termes de buts (Vallacher & Wegner, 1989, 2012; Zacks et al., 2001) ?

4.1. Approches sensorimotrices et prédictives

La première réponse, apportée par ce que nous désignerons comme les approches sensorimotrices, repose sur le traitement des informations visuelles du mouvement. L'observateur parviendrait à extraire des informations sur le but de l'acteur à travers le traitement de l'acte moteur. Cette capacité reposerait sur un phénomène de simulation motrice par lequel l'observateur réactiverait les programmes moteurs utilisés par l'acteur dans son système moteur (Gallese, Rochat, Cossu, & Sinigaglia, 2009; Rizzolatti & Fogassi, 2014). Ce mécanisme serait similaire à celui utilisé dans la production des actions, lorsque l'acteur utilise les conséquences perceptives désirées pour évaluer et sélectionner les commandes motrices les plus appropriées pour son objectif (Jeannerod, 1994; Oztop, Wolpert, & Kawato, 2005; Scott, 2016; Wolpert et al., 2003). La deuxième réponse est apportée par ce que nous désignerons comme les approches prédictives (Csibra, 2008; Kilner, 2011; Kilner, Friston, & Frith, 2007). Ces approches mettent en avant la relation difficile entre buts et commandes motrices puisqu'il n'existe pas de relation unique entre ces deux composantes de l'action. Il ne serait donc pas possible, en se basant sur le mouvement seul, d'accéder aux buts d'une action. Lorsqu'on agit, c'est le but qui oriente la sélection des commandes motrices ; en perception, le même mécanisme serait à l'œuvre. Les observateurs prédiraient le but d'une action en se basant sur d'autres informations, et utiliseraient cette prédiction pour faire sens des mouvements observés.

Les approches sensorimotrices ont été largement popularisées par la découverte des neurones miroirs. Ces neurones visuomoteurs ont la particularité d'avoir un pattern de décharge similaire lors de l'exécution et l'observation d'une action (Di Pellegrino, Fadiga, Fogassi, Gallese, & Rizzolatti, 1992; Kilner & Lemon, 2013). Découverts initialement chez le singe, certains arguments démontrent leur existence chez l'être humain (Mukamel, Ekstrom, Kaplan, Iacoboni, & Fried, 2010). L'observation, chez l'être humain, de pattern de résultats similaires entre observation et exécution de l'action a été initialement interprété comme un argument en faveur de l'utilisation des réseaux sensorimoteurs dans l'accès aux buts des actions observées (Cochin, Barthelemy, Roux, & Martineau, 1999; Fadiga, Fogassi, Pavesi, & Rizzolatti, 1995; Fox et al., 2016; Molenberghs, Cunnington, & Mattingley, 2012). La reconnaissance des actions d'autrui est ici conceptualisée comme un processus ascendant par lequel la compréhension du but de l'acteur émerge du traitement du geste observé. L'étude des propriétés miroirs chez l'être humain est néanmoins complexe. Sans remettre en considération le caractère ascendant de la reconnaissance des actions, certaines approches ont suggéré que les neurones miroirs pourraient acquérir leurs propriétés au cours du développement (Catmur, Press, &

Heyes, 2016; Heyes, 2010) quand d’autres ont suggéré que les propriétés miroirs du système cognitif pourraient ne pas nécessairement impliquer les neurones miroirs (Csibra, 2008; Miall, 2003). Une opposition plus forte issue des approches prédictives remet en cause, cette fois, le caractère ascendant. Les mêmes études sur les neurones miroirs sont cette fois évaluées à travers le prisme d’une prédiction (Bach, Nicholson, & Hudson, 2014; Donnarumma, Costantini, Ambrosini, Friston, & Pezzulo, 2017; Kilner, 2011; Kilner et al., 2007). Les neurones miroirs semblent donc impuissant pour opposer les différentes approches de reconnaissance des actions. Presque 30 ans après leurs découvertes, le rôle fonctionnel des neurones miroirs et de leurs propriétés restent un mystère (Enticott, 2015).

Lorsque l’on se penche sur les arguments expérimentaux en faveurs de l’une ou l’autre des approches, on s’aperçoit qu’elles possèdent, toutes deux, leurs propres justifications. Les approches sensorimotrices sont essentiellement défendues par des paradigmes où le mouvement est isolé de toutes autres informations visuelles, et où les participants observateurs doivent anticiper la finalité du mouvement observé (Ansuini et al., 2014; Quesque & Coello, 2015). Si le manque de rigueur et la faible reproductibilité des résultats sont pointés du doigts (Becchio, Koul, Ansuini, Bertone, & Cavallo, 2018), il semble néanmoins établi que les composantes dynamiques et statiques des mouvements observés sont utilisés dans la reconnaissance des actions (Ansuini et al., 2014; Buxbaum, Shapiro, & Coslett, 2014). Les approches prédictives sont quant à elles défendues par un ensemble plus éclectique de paradigmes expérimentaux. D’une manière générale, les paradigmes démontrant l’impact du contexte visuel (Amoruso, Finisguerra, & Urgesi, 2016; Amoruso & Urgesi, 2016; Wurm & Schubotz, 2012, 2016), des connaissances préalables sur les intentions de l’acteur (Hudson, Bach, & Nicholson, 2018; Hudson, Nicholson, Ellis, & Bach, 2016; Hudson, Nicholson, Simpson, Ellis, & Bach, 2016; Schenke, Wyer, & Bach, 2016) ou de variables attentionnelles (Flanagan, Rotman, Reichelt, & Johansson, 2013; Muthukumaraswamy & Singh, 2008; Perry, Troje, & Bentin, 2010; Schuch, Bayliss, Klein, & Tipper, 2010) sur le traitement du mouvement observé ont été utilisés pour montrer l'existence de processus descendants par lesquels le traitement du mouvement est biaisé. De notre point de vue, la question n'est plus de savoir quelle information est traitée ou impliquée dans la reconnaissance des actions d'autrui mais plutôt quand et comment cette information est-elle impliquée (Kilner & Frith, 2008; Thioux, Gazzola, & Keysers, 2008) ?

4.2. Problématique et stratégie méthodologique

Le point de débat critique dans la littérature sur la reconnaissance visuelle des actions concerne la place des informations sur le geste. Pour les approches sensorimotrices, l'information visuelle sur le geste est prioritaire et donne accès aux informations sur le but de l'acteur. Pour les approches prédictives au contraire, il n'est pas possible de comprendre l'information du geste sans information préalable sur le but de l'acteur ; l'information sur le geste intervient donc plus tard dans le traitement des actions observées. Tenter de dissocier ces deux approches requiert de manipuler indépendamment l'information sur le geste et l'information sur le but. Il est donc nécessaire de trouver des actions où les sources visuelles d'informations ne reposeraient pas uniquement sur le geste, composante centrale et nécessaire des actions, mais aussi sur d'autres éléments. Si les approches prédictives n'émettaient pas d'hypothèse forte sur le type d'information contextuelle permettant de générer des prédictions, certains auteurs ont proposé plus tardivement que les objets pourraient être de bons candidats pour permettre de dériver des informations sur le but (Bach et al., 2014). Les actions dirigées vers des objets nous semblaient donc une manière appropriée de répondre à notre question. En effet, réussir une action dirigée vers un objet nécessite non seulement un geste approprié à l'usage de l'objet, mais aussi de tenir l'objet dans le bon sens. Nous avons donc construit des photographies d'actions avec des objets dans lesquelles l'information sur le but était répartie de manière équilibrée sur le geste utilisé et sur la position relative de l'objet par rapport à la main.

Ainsi, la configuration de la main pouvait correspondre, ou non, à la configuration typique correspondant à l'utilisation de l'objet (geste typique versus atypique). L'objet pouvait être tenu à l'endroit ou à l'envers, et ainsi permettre, ou non, l'utilisation typique de l'objet pour sa fonction (but typique versus atypique). Les deux dimensions sont indépendantes dans la mesure où le geste pouvait être approprié quelques soit la position de l'objet et vice versa. Les différentes photographies ont ensuite été implémentées dans différents paradigmes expérimentaux permettant d'accéder plus précisément à la dynamique temporelle du traitement du geste et des autres informations relatives au but de l'action.

Partie expérimentale

Pour étudier la dynamique temporelle du traitement du geste et des informations liées au but de l'action, différentes stratégies ont été adoptées. Dans un premier temps, nous avons cherché à obtenir des indices comportementaux relatifs au moment où le participant accède à chacune des informations. Nous avons ensuite cherché à obtenir des indices neurophysiologiques de cet accès pour tenter d'appréhender les mécanismes neuronaux sous-jacents aux processus de décodage des actions. Enfin, nous nous sommes intéressés à la variabilité de ces processus, et plus particulièrement, au lien entre cette variabilité et les caractéristiques de l'observateur.

4.3. Arguments comportementaux

Dans une première série d'expérience, ces photographies d'actions ont été implémentées dans un paradigme d'amorçage[20]. Les photographies cibles étaient brièvement amorcées par des photographies partageant soit le même geste, soit le même but, soit les deux informations, soit aucune des deux. Dans tous les cas, l'objet était le même dans l'amorce et la cible. Pour accéder aux informations sur la dynamique temporelle de traitement, le temps de présentation de l'amorce était manipulé. Deux premières expériences, impliquant respectivement 28 et 25 participants, ont permis d'évaluer différents temps de présentation d'amorce : 66, 120, 220 et 300 ms. Les résultats montrent un effet d'amorçage plus important lorsque le but est partagé entre l'amorce et la cible que lorsque c'est le geste qui est partagé entre l'amorce et la cible dès 66ms de temps de présentation d'amorce et ce jusqu'à 220 ms de

---

[20] **Decroix, J.** & Kalénine, S. (2018). Timing of grip and goal activation during action perception: A priming study. *Experimental Brain Research, 1-16.* doi: *10.1007/s00221-018-5309-0*

temps de présentation d'amorce. Lorsque l'amorce est présentée 300 ms, l'effet d'amorçage par le geste devient plus important que celui par le but. Une 3ème expérience, impliquant 64 participants, a permis de répliquer les résultats obtenus pour 66 et 220 ms de temps de présentation d'amorce. Ces résultats suggèrent que l'information sur le but véhiculée par d'autres informations que le geste semble être priorisée par l'observateur dans les premières étapes du traitement visuel des actions, alors que l'information sur le geste sera utilisée plus tard dans le traitement de l'action.

Les résultats de cette première série d'étude ont été corroborés par une deuxième étude impliquant une tâche de recherche visuelle[21]. Quatre types d'image étaient présentés au participant correspondant aux quatre combinaisons possibles entre la typicalité du but et la typicalité du geste. Dix-huit participants ont dû identifier et sélectionner la photographie correspondant à une utilisation typique de l'objet (la cible), et par conséquent éliminer la photographie dans laquelle le geste était atypique (distracteur geste), celle où le but était atypique (distracteur but) et celle où les deux dimensions étaient atypiques (non relié). Dans cette étude, le comportement oculaire des participants étaient enregistrés. Nous nous sommes intéressés à la proportion de fixation en fonction du temps sur chaque photographie. A terme, 100% des fixations devraient tomber sur la photographie à identifier. L'évolution de la proportion de fixation sur chacune des images avant l'identification de la cible nous donne un indicateur sur l'évolution de l'attribution des ressources visuo-attentionnelles sur chacune des images (De Groot, Huettig, & Olivers, 2016; Huettig, Olivers, & Hartsuiker, 2011; Kalénine, Mirman, Middleton, & Buxbaum, 2012; Mirman & Magnuson, 2009). Si un distracteur partage un élément avec la cible, on s'attend à ce qu'une partie des ressources visuo-attentionnelles

---

[21] **Decroix, J.** & Kalénine, S. (2019). What first drives visual attention during the recognition of object-directed action? The role of kinematics and goal information. *Attention, Perception & Psychophysics, 1-10.* doi: 10.3758/s13414-019-01784-7

soient capturées par ce distracteur. Lorsque l'on confronte la cible à deux distracteurs partageant deux dimensions distinctes avec la cible, la distribution des fixations au cours du temps nous permet de savoir laquelle des deux dimensions capture les ressources visuo-attentionnelles en premier. Les résultats montrent une proportion de fixation plus importante sur le distracteur partageant le même but que le distracteur partageant le même geste que la cible dans les premières centaines de millisecondes de recherche visuelle, alors que c'est l'inverse dans les dernières centaines de millisecondes de recherche visuelle précédant l'identification de la cible. Corroborant les résultats de l'étude d'amorçage, cette seconde étude suggère à nouveau que les informations sur le but autre que le geste sont priorisées dans les premières étapes du traitement visuel des actions.

4.4. Arguments neurophysiologiques

Jeannerod (1994, 1999, 2009) défendait l'idée que le traitement des actions n'était pas seulement l'affaire de processus périphériques mais aussi l'affaire de processus centraux, autrement dit, de processus cérébraux. Bien qu'elles poursuivent des objectifs sensiblement distincts, la psychologie et les neurosciences partagent un intérêt commun pour la compréhension de la cognition, et des processus qui la sous-tendent. Comprendre un processus cognitif requiert donc aussi de comprendre les mécanismes neuronaux qui le génèrent. En outre, les méthodologies comportementales sont souvent limitées par le fait que, pour être observable, un phénomène cognitif doit avoir un impact sur le comportement. Or il est difficile de clairement identifier les relations entre différentes étapes de traitement de l'information, et affecter une étape de traitement peut générer des compensations sur les étapes ultérieures. Ainsi, une absence de modification comportementale peut néanmoins dissimuler des différences de traitement. Dans un second axe de recherche, nous nous sommes donc intéressés aux substrats neurophysiologiques impliqués dans la reconnaissance des actions en gardant un intérêt particulier pour des aspects dynamiques.

Ces questions ont d'abord été abordées à travers les potentiels évoqués obtenus en électroencéphalographie (EEG)[22]. L'EEG est réputée pour sa finesse temporelle et permet donc d'obtenir des données intéressantes sur les différentes étapes de traitement sensibles aux informations sur le geste et aux informations sur le but lors de la reconnaissance de photographies d'actions. Trente et un participants ont pris part à l'étude et avaient pour tâche d'évaluer la typicalité des actions présentées alors que leur activité cérébrale était enregistrée. Le traitement d'action dirigée vers des objets a généré les composantes P100, N170, P300 avec une topographie postérieure et les composantes N300 et N400 avec une topographie antérieure, en accord avec ce qui est classiquement observé lors du traitement d'informations visuelles (Bach, Gunter, Knoblich, Prinz, & Friederici, 2009; Bledowski et al., 2004; Chang et al., 2018; Peelen & Downing, 2007; Wamain, Pluciennicka, & Kalénine, 2014). La P100 et la N170 ont montré une sensibilité au traitement du geste et au traitement du but, ce qui reflète la capacité des observateurs à distinguer, dès les premières étapes de traitement, les informations visuelles relatives aux gestes et aux buts. La N300 est sensible aux différences de but mais pas aux différences de geste, ce qui suggère que les premières étapes de traitement sémantique pourraient être dirigées par le traitement des informations relatives aux buts. La N400 est sensible aux différences de but et de geste, et également à l'intégration des deux informations. En effet, les actions atypiques sur les deux dimensions entrainent une négativité plus importante que n'importe quelle autre combinaison. Globalement, les résultats obtenus en EEG miment et précisent les résultats obtenus en amorçage. Le cerveau distingue les deux dimensions dès les

---

[22] **Decroix, J.**, Roger, C. & Kalénine, S. (2020). Neural dynamics of grip and goal integration during the processing of others' actions with objects: An ERP study. *Scientific Reports, 1-11.* doi: 10.1038/s41598-020-61963-7

étapes perceptives, et les données suggèrent que les étapes sémantiques du traitement des actions pourraient être dirigées par le traitement du but. En effet, la sensibilité de la N300 aux différences de but pourrait indiquer la création d'une prédiction sur le but de l'acteur qui seraient par la suite comparée aux informations relatives au geste, ce qui serait reflété par les modulations de la N400.

Les données obtenues en EEG donnent accès à des informations intéressantes sur le traitement des actions observées mais ne permettent pas d'établir de relations causales entre chaque dimension et les substrats neurophysiologiques identifiés. Une des rares techniques permettant d'établir des relations causales entre des processus cognitifs et des substrats neuronaux impliquent la stimulation magnétique transcrânienne (TMS). Un champ magnétique est appliqué prêt du crâne, et induit une perturbation des neurones situés en dessous de la zone stimulée (Siebner, Hartwigsen, Kassuba, & Rothwell, 2009; Ziemann, 2010). Pour permettre une mesure plus fine, nous avons associé la TMS avec un paradigme d'amorçage[23]. Dans ce type de paradigme, la TMS est utilisée pour perturber l'effet de l'amorce. Par conséquent, la TMS est appliquée au moment de l'affichage de l'image cible. On s'attend à ce que l'effet d'amorçage soit diminuée après l'application de la TMS (Cattaneo, Devlin, Salvini, Vecchi, & Silvanto, 2010; Silvanto & Cattaneo, 2017). Cela permet de cibler plus précisément une information particulière (i.e., celle concernée par l'amorçage) et permet en outre de s'intéresser à une étape de traitement de l'action particulière. Dans notre cas, l'amorce était présentée pendant 220 ms, ce qui laisse assez de temps aux deux dimensions d'influencer la cible et d'être intégrées. L'intégration des deux dimensions se traduisaient dans notre paradigme d'amorçage par un coût de traitement plus important lorsque l'amorce partage une seule des deux

---

[23] *Decroix, J.*, Borgomaneri, S., Kalénine, S. & Avenanti, A. (submitted). Frontal and parietal integration of visual kinematics and functional goals during object-directed action recognition: Evidence from TMS-priming?

informations avec la cible que lorsqu'elle partage les deux dimensions ou aucune des dimensions (notion de coût de répétition partiel; Hommel, 2004). La TMS a été appliquée selon trois conditions : sur le gyrus frontal inférieur gauche, sur le gyrus pariétal inférieur gauche et selon une condition sham correspondant à une ligne de base où la TMS n'est pas directement appliquée sur le crâne. Les deux zones cérébrales ont été préalablement associées au traitement des actions observées, mais il reste difficile d'identifier précisément les dimensions qu'elles traitent. La présence de résultat contradictoire dans la littérature sur la reconnaissance des actions pousse de plus en plus d'auteurs à souligner l'importance de l'aspect dynamique, et l'objectif ici était donc d'identifier quelle est la dimension traitée par le gyrus frontal inférieur et le gyrus pariétal inférieur. Dix-huit participants ont pris part à l'étude. Les résultats montrent que le traitement du but indépendamment du type de geste et le traitement du geste indépendamment du type de but ne sont impactés ni par la stimulation du gyrus frontal inférieur, ni par la stimulation du gyrus pariétal inférieur. En revanche, le marqueur d'intégration des deux dimensions est bien diminué après l'application de la stimulation sur chacune des zones cérébrales. Nos résultats suggèrent donc que le gyrus frontal inférieur et le gyrus pariétal inférieur sont tous deux impliqués dans le traitement visuel des actions et qu'après 220 ms de présentation visuelle, ils sont impliqués dans l'intégration des deux dimensions.

Globalement, les résultats des deux études neurophysiologiques s'insèrent bien dans les modèles neurocognitifs des approches prédictives. L'étude EEG en appuyant la dynamique temporelle du traitement des informations sur le geste et sur le but ; l'étude TMS en impliquant causalement le réseau fronto-pariétal dans l'intégration des deux informations. En effet, selon les approches prédictives, le réseau fronto-pariétal serait notamment impliqué dans la comparaison entre les gestes prédits obtenus grâce à la prédiction du but de l'acteur et les gestes observés (Kilner, 2011).

4.5. Variabilité inter-individuelle et caractéristiques de l'observateurs

Après nous être intéressés aux mécanismes cognitifs et cérébraux qui sous-tendent la compréhension des actions d'autrui, nous nous sommes demandés si la priorité donnée aux informations sur le but par rapport aux informations sur le geste d'une action était différente d'un participant à l'autre[24]. Plus précisément, nous nous sommes demandés si ces différences entre individus pouvaient être expliquées par les caractéristiques propres aux observateurs. Ainsi, le degré de familiarité avec l'action observée (Nicholson, Roser, & Bach, 2017), l'appréciation que l'observateur a de l'acteur (Kozak, Marsh, & Wegner, 2006; Marsh et al., 2010) ou encore les connaissances que l'observateur a sur l'acteur (Schenke et al., 2016) sont autant de variables qui influencent la manière dont nous reconnaissons les actions d'autrui. De nombreuses études ont également trouvé des liens entre la compréhension des actions et diverses caractéristiques des observateurs comme celles reliées à leurs habiletés sociales (Borgomaneri, Gazzola, & Avenanti, 2015; Y. Cheng, Yang, Lin, Lee, & Decety, 2008; Dapretto et al., 2006; DiGirolamo, Simon, Hubley, Kopulsky, & Gutsell, 2019; Gazzola, Aziz-Zadeh, & Keysers, 2006; Keysers & Gazzola, 2006; Lewkowicz, Quesque, Coello, & Delevoye-Turrell, 2015; Pfeifer, Iacoboni, Mazziotta, & Dapretto, 2008). Dans une dernière étude corrélationnelle, nous avons donc chercher à faire le lien entre différents facteurs associés aux habiletés sociales et la priorité donnée aux informations sur le but de l'action. Les 64 participants de la 3ème expérience d'amorçage ont également eu à remplir divers questionnaires relatifs à leur sentiment de pouvoir social (Anderson, John, & Keltner, 2012), de dominance (J. T. Cheng, Tracy, & Henrich, 2010), à leur tendance à prendre la perspective psychologique d'autrui (Davis, 1983; Gilet, Mella, Studer, Griihn, & Labouvie-Vief, 2013) et au niveau

---

[24] *Decroix, J.*, Morgado, N. & Kalénine S. (submitted). Preference for visual goal over grip explained by individual characteristics during the recognition of object-directed actions.

d’abstraction auquel ils décrivent habituellement une action (Vallacher & Wegner, 1989). Nous avons essayé de prédire l’importance donnée aux informations de buts par rapport aux informations de gestes par les scores de ces différents questionnaires. Nos résultats montrent que la priorité donnée aux informations de but est prédite par le score de sentiment de pouvoir social et par le score de dominance. Plus le score de pouvoir social est élevé, plus le participant donne du poids à l’information de but. Au contraire, plus le score de dominance est élevé, moins le participant donne du poids à l’information de but. Ces résultats suggèrent que le sentiment de pouvoir social et l’utilisation de stratégie dominante pourraient influencer le poids donné aux informations de buts de manière différente. De manière plus importante, ces résultats suggèrent que les caractéristiques propres de l’observateur/trice, ici dans le domaine des habiletés sociales, pourraient influencer la manière dont il/elle perçoit les actions d’autrui.

Discussion

La présente thèse avait pour objectif d'étudier la dynamique temporelle de traitement des informations relatives aux gestes et des informations relatives aux buts. Après avoir dissociés les deux types d'informations, nous avons cherché à savoir le moment à partir duquel les observateurs accèdent à chaque catégorie d'information lors de la reconnaissance d'action. Notre question de recherche était guidée par le besoin de dissocier deux grandes approches. Les approches sensorimotrices suggèrent que c'est le geste qui permet d'accéder au but de l'action, le traitement du geste précède donc celui du but. A l'inverse, les approches prédictives suggèrent qu'il n'est pas possible de faire sens de geste observé sans attente préalable. Dans ces approches, les informations sur le but de l'acteur sont d'abord dérivées d'informations non motrices et vont ensuite guider le traitement du geste. Nos résultats suggèrent que les informations relatives aux buts de l'actions possèdent un poids plus important dans les premières étapes de traitement de l'action, alors que les informations sur le geste prendraient le relai dans les dernières étapes de traitement de l'action. Ces données favorisent donc les approches prédictives. Même lorsque le contexte est réduit au maximum, le traitement du but prime.

4.6. L'objet, un prédicteur improbable

Pour pouvoir dissocier les informations relatives au but et celles relatives au geste de l'action, nous nous sommes basés sur l'objet. Quelle que soit la configuration, le but ou le geste d'une action n'était jamais incorrect dans le sens où l'action était toujours faisable bio-mécaniquement parlant. L'a-typicalité d'un geste ou d'un but était systématiquement dépendante de l'objet considéré. L'objet est donc central dans notre stratégie expérimentale, et les participants devaient nécessairement traiter l'identité de l'objet en premier. Il est donc possible que l’objet puisse expliquer le fait que les participants utilisent les informations de but avant les informations de geste. Nous ne nierons pas le rôle potentiel de l’objet dans nos résultats. Néanmoins, l’identité de l’objet ne pourrait pas en être directement l’origine. En effet, l’identité de l’objet est partagée entre toutes les conditions, et l’orientation de l’objet n’a pas d’effet sur l’accès à l’identité de l’objet (E. E. Cooper, Biederman, & Hummel, 1992; Dilks, Julian, Kubilius, Spelke, & Kanwisher, 2011; Peelen & Downing, 2007; Rossion et al., 2003; Rossion & Jacques, 2008). Si l’identité de l’objet a un rôle à jouer dans l’émergence du pattern de résultat que nous obtenons, c’est à travers l’accès aux connaissances sur la manipulation (qui orienterait vers le traitement du geste geste) et l’accès aux connaissances sur la fonction (qui orienterait vers le traitement du but). Néanmoins, l’identité de l’objet donne accès aux deux types de connaissances de manière équivalente et indépendante (Buxbaum & Kalénine, 2010; Campanella & Shallice, 2011; van Elk, van Schie, & Bekkering, 2014). A ce jour, on ne sait pas dans quelle mesure l’accès aux deux types de connaissance est simultané ou non (Buxbaum & Kalénine, 2010; Collette, Bonnotte, Jacquemont, Kalénine, & Bartolo, 2016; van Elk et al., 2014). Au mieux, nos résultats indiquent que l’accès aux connaissances sur la fonction est priorisé lors de la reconnaissance des actions dirigées vers un objet. Si l’objet a pu orienter nos résultats, ce pattern n’était en tout cas pas prévisible.

4.7. L’implication tardive du geste lors de la reconnaissance d’action

Les informations sur le geste sont importantes et nécessaires à la reconnaissance d’action. Les études émanant des approches sensorimotrices ont largement démontré que ces informations alimentent la reconnaissance d’action, et contribuent à la reconnaissance des buts (Ansuini et al., 2014; Cavallo, Koul, Ansuini, Capozzi, & Becchio, 2016; Lewkowicz et al., 2015; Quesque & Coello, 2015). Néanmoins, c’est lorsque l’objet n’est pas visible (Thioux & Keysers, 2015) ou que l’action est difficile à reconnaitre (Nicholson et al., 2017) que l’observateur bénéficie le plus des informations sur le geste. Dans le même ordre d’idée, les regards proactifs, souvent considérés comme la preuve que l’observateur anticipe le but de l’action, ne sont possibles que lorsque l’observateur a suffisamment d’information pour prédire le but de l’action. En l’absence d’information contextuelle (Donnarumma et al., 2017), ou lorsque l’observateur n’a pas l’expertise motrice suffisante (Ambrosini et al., 2013; Geangu, Senna, Croci, & Turati, 2015), ces regards ne sont plus observés. Nos résultats s’inscrivent dans cette lignée, et montre que lorsque les deux informations sont disponibles simultanément, l’implication des informations reliées au geste reste assez tardive. Notre manipulation expérimentale se limite aux éléments posturaux des informations relatives au geste, et il faudrait pouvoir généraliser ce pattern a des stimuli dynamiques. Néanmoins, les éléments posturaux sont des aspects importants du geste (Buxbaum et al., 2014), et les données actuelles ne permettent pas d’identifier quels aspects du geste sont les plus important dans la reconnaissance d’action (Becchio et al., 2018). Quoiqu’il en soit, au moins pour les aspects posturaux, lorsque deux sources d’information sont présentes, le geste n’est pas priorisé.

4.8. Vers une approche pluraliste

Lorsque d’autres informations sont présentes, le geste ne semble pas diriger la reconnaissance de l’action. Certaines études montrent néanmoins qu’en l’absence de d’information non motrice, nous sommes capable d’identifier l’action (e.g., Avenanti, Paracampo, Annella, Tidoni, & Aglioti, 2017; Lewkowicz et al., 2015; Manera, Becchio, Schouten, Bara, & Verfaillie, 2011). Nos résultats montrent d’ailleurs que l’information relative au geste influence la reconnaissance de l’action dès les premières étapes de traitement, bien que cette influence soit moins importante que celle des informations relatives au but de l’action. Pour expliquer cette apparente contradiction, un nombre croissant d’auteurs adoptent une vision pluraliste de la reconnaissance de l’action. La reconnaissance de l’action ne serait pas un processus unitaire, mais une boite à outil avec plusieurs stratégies ou processus disponible pour comprendre les actions réalisées par autrui (Bach et al., 2014; Quesque & Coello, 2015; Schilbach, 2010; Springer, Parkinson, & Prinz, 2013; Uithol & Paulus, 2014). Un nombre croissant d'étude montre que la reconnaissance d'action est sensible à des facteurs attentionnels (D'Innocenzo, Gonzalez, Nowicky, Williams, & Bishop, 2017; Donaldson, Gurvich, Fielding, & Enticott, 2015; Leonetti et al., 2015; Muthukumaraswamy & Singh, 2008; Perry et al., 2010; Pomiechowska & Csibra, 2017; Riach, Holmes, Franklin, & Wright, 2018; Schuch et al., 2010; Woodruff & Klein, 2013; Wright et al., 2018). L'importance de la tâche a également été plusieurs fois soulignées (Flanagan et al., 2013; van Elk, Van Schie, & Bekkering, 2008). Selon la tâche, les effets observés vont varier. Evaluer le poids d'un objet soulevé par un acteur induit un pattern de recherche visuel similaire à celui que le participant aurait eu pour interagir lui-même avec l'objet (Flanagan & Johansson, 2003; Flanagan et al., 2013), ce qui n'est pas le cas lorsque le participant doit simplement prédire l'objet que l'acteur va saisir (Flanagan et al., 2013). Dans notre dernière étude, nous avons également observé que les caractéristiques propres de l'individu pouvaient être associées à différentes manières de traiter une même action. L'observateur lui-même aurait donc une influence sur la manière dont les actions sont traitées, et différents observateurs pourraient avoir différentes stratégies pour reconnaitre la même action. En considérant ces différents aspects, il est possible que les actions dirigées vers les objets constituent une catégorie d'action à part entière, et donc que nos résultats puissent être limités à cette catégorie. Néanmoins, si les résultats concernant les actions dirigées vers les objets ne sont pas généralisables à tout type d'action, l'inverse est aussi vrai : les résultats obtenus avec des actions non dirigées vers les objets ne devraient pas être librement généralisés à tout type d'action.

Conclusion

Dans l’ensemble nos résultats favorisent les approches prédictives de la reconnaissance des actions. En effet, les travaux menés dans cette thèse soulignent l’importance des informations relatives aux buts des actions dès les premières étapes du traitement des actions. Nos résultats apportent des arguments nouveaux, soulignent l’importance de la dynamique temporelle dans l’étude de la reconnaissance des actions, et fournissent des balises pour les études à venir. Nous suggérons néanmoins que le bénéfice de ces données serait plus important si l’on considère une approche pluraliste de la reconnaissance des actions.