**IMT School for Advanced Studies, Lucca**

Lucca, Italy

# The organization of action representation and the interaction between the action and perception systems

PhD Program in Cognitive and Cultural Systems – Track in Cognitive, Computational and Social Neuroscience

XXXIII Cycle

**By**

**Francesca Simonelli**

**2023**

**The dissertation of Francesca Simonelli is approved.**

PhD Program Coordinator: Prof. Emiliano Ricciardi, IMT School for Advanced Studies Lucca

Advisor: Prof. Pietro Pietrini, IMT School for Advanced Studies Lucca

Co-Advisor: Prof. Emiliano Ricciardi, IMT School for Advanced Studies Lucca

The dissertation of Francesca Simonelli has been reviewed by:

Prof. Luca Turella, University of Trento, Italy

Prof. Sara Parmigiani, Stanford University, CA, USA

**IMT School for Advanced Studies, Lucca**

**2023**

## Contents

# List of Figures

# Acknowledgments

Chapter 3 is based on a paper in progress co-authored with Giacomo Handjaras, Francesca Benuzzi, Giulio Bernardi, Andrea Leo, Davide Duzzi, Luca Cecchetti, Paolo F. Nichelli, Carlo A. Porro, Pietro Pietrini, Fausta Lui and Emiliano Ricciardi. The affiliation of Handjaras, Bernardi, Leo, Cecchetti, Pietrini, Ricciardi is IMT School for Advanced Studies Lucca. The affiliation of Benuzzi, Duzzi, Nichelli, Porro, Lui is University of Modena and Reggio Emilia. Chapter 4 is based on a paper in progress co-authored with Simone Toma, Marco Santello, Ioannis Delis. The affiliation of Toma and Santello is Arizona State University. The affiliation of Delis is University of Leeds.

I would like to thank Emiliano Ricciardi, Giacomo Handjaras, Andrea Leo, Simone Toma and Marco Santello for mentoring me during my PhD years and for their help in drafting this thesis.

I am grateful to my fellow PhD students, especially the CCSN XXXIII, who made this journey much more enjoyable.

And, as always, a special thanks to my family.

# Vita and publications

| | |
|---|---|
| **October 17, 1990** | Born in Fiesole (Florence), Italy |
| **2013** | Bachelor's Degree in Psychology<br>Final mark: 108/110<br>Università La Sapienza di Roma, Rome, Italy |
| **2016** | Master's Degree in Cognitive Neuroscience and Clinical Neuropsychology<br>Final mark: 110/110 cum laude<br>Università degli studi di Padova, Padua, Italy |
| **2017** | Admission to the PhD program in Cognitive and Cultural Systems, IMT School for Advanced Studies Lucca, Italy |
| **2020** | Visiting Research Student<br>Arizona State University, AZ, USA |

- Castellani N., **Simonelli F**., Bottari D., Manuello J., Liloia D., Duca S., …& Garbarini F. Inter-subject correlation unveils a shared visuo-tactile representation in primary and secondary somatosensory cortices. Poster session at: 21st International Multisensory Research Forum; 2023 June 27-30; Brussels, Belgium.

- **Simonelli F**., Handjaras G., Benuzzi F., Bernarni G., Leo A., Duzzi D., …& Ricciardi E. Voxel Sensitivity to Kinematic and Object-related Features During Action Observation. Poster session at: OHBM. 26th Annual Meeting of the Organization for Human Brain Mapping; 2020 Jun 23-Jul 3; virtual edition.

- Schwartzman, D. J., Schartner, M., Ador, B. B., **Simonelli, F**., Chang, A. Y. C., & Seth, A. K. "Increased spontaneous EEG signal diversity during stroboscopically-induced altered states of consciousness". *BioRxiv: 511766*, 2019.

- **Simonelli F**., Handjaras G., Benuzzi F., Bernarni G., Leo A., Duzzi D. …& Ricciardi E. Interaction of transitivity and meaning during action observation. Poster session at: OHBM. 25th Annual Meeting of the Organization for Human Brain Mapping; 2019 June 9- 3; Rome, Italy.

# Abstract

Action representation is thought to rely on hierarchical processing, whereby action-related information is represented across different levels of abstraction. The visual processing of human actions is subserved by the Action Observation Network (AON), spanning separate anatomical areas. Rather than relying on a modular organization based on the segregation of action features, the AON organization may be based on distributed representations, overlapping across a wide expanse of the cortex and supporting a high-dimensional representational space. Coherently with modern perspectives of distributed interactive cortical systems, it has been proposed that the same representational systems are shared between action, perception, and higher-level cognitive functions; in this view, action and perception are interdependent systems characterized by reciprocal influences.

In the first and second studies, we aimed to better characterize the organization of the AON at different levels of the action hierarchy by measuring brain hemodynamic activity and representational geometries during observation of transitive and intransitive gestures. In the third study, we propose a novel methodological and analytical framework aimed at investigating the interplay between action and perceptual processes during active perception; we tested the validity of the proposed approach by recording participants' motor and perceptual behavior during a perceptual decision-making task and assessing the influence of motor control features on decision formation.

The results of the first and the second study indicate that actions categories are not encoded in a segregated manner and that the same areas of the AON participate in the representations of multiple action features; moreover, the representational content of the AON is not constrained by

anatomical proximity, as anatomically distant regions shared similar information content. These findings support the notion that the AON relies on overlapping and distributed coding and may act as a unique representational space. In the third study, we established a link between motor and perceptual behavior, as changes in movement kinematics continuously tracked the decision formation process, and we revealed a significant contribution of grip force to perceptual decisions. These results support the validity of the proposed approach in exploring the relationship between patterns of motor behavior and decision processes and the neurophysiological correlates underlying decision-making mechanisms.

# Introduction

## 1. 1 Action Representation

A fundamental problem in motor neuroscience is understanding how the brain represents distinct motor features and combines them to form a high-level representation of a finalized act. Grafton and Hamilton (2007) proposed a model of action organization based on hierarchical processing, where stimuli are processed over multiple consecutive stages and information is carried on and made explicit from one level to another. Within this framework, action is thought to rely on a control hierarchy chaining and integrating movement elements into single units that can be adaptively recombined to deal with changing contexts and achieve distant goals. Action-related information is represented at a concrete level of specific motor features and at an abstract level (e.g., goal) to ensure performance and simultaneously allow for flexibility of behavior (Spunt et al., 2016; Turella et al., 2020). Evidence supporting the notion of an action hierarchy is provided by studies on apraxia, of which Liepmann (1988) described two forms, seemingly affecting different stages of the action hierarchy: ideomotor apraxia, that is, the inability to perform purposeful movements, despite preserved motor functions, and ideational apraxia, which consists in a deficit in concatenating motor elements.

### 1.1.1 Action Observation Network

A large corpus of evidence demonstrates a significant overlap between networks sub-serving action observation and execution (e.g., Buccino et al., 2001; Gazzola & Keysers, 2009; Mukamel et al., 2010; Molenberghs et al., 2012). A recent metanalysis (Hardwick et al., 2018) identified consistently shared activations in parietal areas, such as the inferior parietal lobule (*IPL*), and in premotor regions, including bilateral pre-supplementary motor area (*pre-SMA*), left supplementary area (*SMA*), bilateral ventral premotor cortex (*PMv*), right dorsal premotor cortex (*PMd*). This shared network is part of a broader action observation network (AON), specialized for the representation of human actions and spanning bilateral frontal, parietal, and posterior temporal areas (Culham & Valyear, 2006; Caspers et al., 2010; Kilner, 2011; Filimon et al., 2015; Reynaud et al., 2019; Vannuscorps et al., 2019).

#### 1.1.1.1 Parietal areas

The role of the posterior parietal cortex in action-related processing is well established (Culham & Valyear, 2006; Caspers et al., 2010; Hardwick et al., 2018), and evidence indicates it is functionally divided into sub-regions responding to observation of different classes of actions (Urgen & Orban, 2021; Jastorff et al., 2010; Abdollahi et al., 2013).

The intraparietal sulcus (*IPS*) and *IPL* have been implicated in the execution and observation of object-directed movements,

with anterior parts (*aIPS*) specialized for grasping and object manipulation (Culham & Valyear, 2006; Abdollahi et al., 2013), while activations in posterior parts reflect processing of action-relevant object features (Gallivan et al., 2013; Villarreal et al., 2008). The role of *aIPS* in grasping seems to relate to the level of precision required as it shows greater activation for precision grips regardless of the number of digits (Cavina-Pratesi et al., 2018), and it encodes action type (i.e., precision grip vs. whole-hand grasp) independently of effector and low-level kinematic features (Turella et al., 2020). These findings suggest that *aIPS* could integrate information about object properties and task demands to compute appropriate hand configurations (Cavina-Pratesi et al., 2018). Conversely, activations in superior parietal areas (*SPL*) are associated with directional arm movements, e.g., reaching and pointing, suggesting a role of *SPL* in processing the arm transport component of hand movements (Savaki et al., 2022; Filimon et al., 2007; Fabbri et al., 2014; Cavina-Pratesi et al., 2018).

Neuroimaging studies have also implicated *IPL* in encoding observed actions at different levels of abstraction, generalizing across kinematics and object category (Wurm & Lingnau, 2015), and in the representation of immediate action goals, understood as the observable end state of an action (Urgen & Orban, 2021; Jastorff et al., 2010; Wurm et al., 2016; Urgen et al., 2019). Koul et al. (2018) found that *IPL* and *SPL* encoded goal-related information conveyed by variations in movement kinematics, suggesting that agents' intentions can be inferred from

kinematic-related information coded in these areas even in the absence of contextual clues.

#### 1.1.1.2 Temporal areas

Temporal areas of the AON receive inputs from the visual system and relay this information to other action observation areas for further processing (Kilner, 2011; Urgen & Orban, 2021).

Recently, Wurm and Caramazza (2022) proposed that AON temporal areas, specifically the lateral occipitotemporal cortex (*LOTC*), may be part of a visual pathway dedicated to action recognition, analogous to the ventral stream for object recognition (Goodale & Milner, 1992). While frontoparietal areas are mainly sensitive to 'how' aspects of action representation, e.g., kinematics of manipulation actions, *LOTC* is involved in the representation of features relevant for action recognition and is organized along a posterior-to-anterior gradient from perceptual aspects, such as biological motion, to conceptual representations (Wurm & Caramazza, 2022). Consistently with this notion, the involvement of posterior temporo-occipital areas in the perception of biological motion is well established (Peuskens et al., 2005; Zeki, 2015), while the middle temporal gyrus (*MTG*) has been implicated in semantic processing (Villarreal et al., 2008; Xu et al., 2009; Kubiak & Króliczak, 2016). Moreover, the representational organization of the posterior superior temporal sulcus (*pSTS*) has been shown to correlate with models describing the visual form and trajectory of the stimuli and the type of action, suggesting this area integrates form and motion

information to encode action type (Urgen et al., 2019). Notably, action-related representations in *MTG* are also organized along a posterior-to-anterior gradient, with posterior regions encoding symbolic gestures and pantomimes and anterior areas showing higher activation to word stimuli (Papeo et al., 2019).

#### 1.1.1.3 Frontal areas

Frontal areas are considered a second-level anatomical stage in the AON (Urgen & Orban, 2021), receiving input from visual areas and other AON regions. Specifically, premotor areas receive sensorimotor and hand-object interaction information and integrate it into the appropriate concrete motor program to understand observed action (Kilner, 2011; Caspers et al., 2010).

Activations in *PMv* and *PMd* have been associated with both planning and observation of grasping movements, independently of the effector involved in the action (Villarreal et al., 2008; Caspers et al., 2010; Gallivan et al., 2013), although some evidence indicates a dissociation between these two areas in coding grasping and reaching, respectively (Filimon et al., 2007; Filimon, 2010; Fabbri et al., 2014; Cavina-Pratesi et al., 2018). Overall, *PMv* seems to encode mainly low-level features of observed movements (Wurm & Lingnau, 2015), suggesting it provides a kinematics-based representation of action (Savaki et al., 2022).

Conversely, *IFG* is involved in action representation at higher levels of abstraction, as it has been shown to respond to goal-

directed actions and to code agents' intentions (Koul et al., 2018; Savaki et al., 2022); moreover, lesions to this area have been associated with impairments in action goals recognition (Kalénine et al., 2013). *IFG* has been suggested to be part of a ventral pathway within the AON from middle temporal areas to anterior to posterior parts of *IFG*, representing actions along a gradient from conceptual to concrete level (Kilner, 2011; Bischoff et al., 2014; Möttönen et al., 2016): representations of the actions associated with an object would be encoded from an abstract level, such as goals, to the appropriate motor parameters for interacting with the object, along the anterior-to-posterior axis in the inferior frontal cortex (Badre & D'esposito, 2009).

### 1.1.2 Organization of the Action Observation Network

In accordance with the model proposed by Grafton and Hamilton (2007), action information processing in the AON is hypothesized to be hierarchical: visual information is transferred from one anatomical level of the AON to the other, representing increasingly abstract and complex aspects of the observed act (Urgen et al., 2019; Urgen & Orban, 2021). The hierarchical organization of the AON is supported by neuroimaging findings of action encoding at different levels of abstraction, from motion and kinematics to immediate goals and outcomes. These studies found differential activation in areas of the AON, representing actions at either a concrete level, e.g., encoding of appearance-based features, or intermediate and higher levels, generalizing

across kinematics, effector, or target object (Grafton & Hamilton, 2007; Tucciarelli et al., 2015; Wurm & Lingnau, 2015; Urgen et al., 2019; Tarhan et al., 2021).

Expanding on the hierarchical account of action representation, the predictive coding model postulates that the AON generates predictions about action goals and how the action will be performed and then uses the kinematics of the observed action to verify these predictions (Kilner, 2011; Koul et al., 2018; Decroix & Kalénine, 2019). Action observation is thus equated to a hypothesis testing process disambiguating between predictions (Donnarumma et al., 2017). According to this account, goal predictions are generated through the ventral pathway linking posterior temporal areas with *IFG*, while dorsal areas generate a concrete instance of the predicted actions and select the most likely hypothesis based on the observed kinematics. Predictive coding accounts are supported by evidence that goal-related information is more salient than appearance-based information during action processing (Decroix & Kalénine, 2019; Tarhan et al., 2021) and that *LOTC* is recruited earlier than frontal areas during action observation (Tucciarelli et al., 2015).

Other evidence suggests that the representational organization of the AON relies on distributed representations: observed actions are encoded at larger spatial scales by populations encompassing multiple voxels tuned to specific features (Giese & Rizzolatti, 2015; Handjaras et al., 2015; Tucciarelli et al., 2019). These distributed representations partially overlap, with voxels participating in multiple representations while carrying

discriminative information (Filimon et al., 2015; Handjaras et al., 2015). Localized activations to different categories of stimuli can be explained by distributed representations with specific topographies that involve neurons to different degrees (Pulvermüller, 2005).

The overlap between multiple representations means that different kinds of information coexist at different spatial scales (Grill-Spector & Weiner, 2014) and may reflect the necessity of supporting a high-dimensional representational space (Haxby et al., 2014).

## 1. 2 Action and Perception in Embodied Theories

Modern perspectives on brain organization have moved away from a modular view based on independent functional modules, suggesting instead that cortical functions are served by distributed interactive systems (Pulvermüller, 2005).

Coherently with this perspective, embodied theories of cognition postulate shared representational systems involved in action, perception, and higher-level cognitive abilities, with sensorimotor representations providing the scaffolding upon which cognitive functions are constructed (Gordon et al., 2021). Action perception in the embodied cognition framework is considered not a passive but an active process, embodied in the AON and engaging sensorimotor representations to recognize and understand the observed actions (Caramazza et al., 2014; Donnarumma et al., 2017). The embodied view also postulates that the action system affects perception processes. Since perception, understood as information sampling from the environment, necessarily requires movement, this results in an integrated perception-action loop characterized by reciprocal influences between the two systems (Creem-Regehr & Kunz, 2010).

For this reason, embodied theories offer an ecologically valid perspective for the study of perceptual decision-making, in that they account for the situated aspects of a decision and are better

suited to model the kind of scenarios in which neural systems underlying decision-making have evolved (Gordon et al., 2021).

### 1.2.1 The role of action in perceptual decision-making

Perceptual decision-making is a competitive process involving the accumulation and evaluation of sensory information guiding motor behavior (Hauser & Salinas, 2014). Several frameworks have been proposed to describe perceptual decision-making and its links to action. The classical framework generally views decision-making as a serial process, where a decision is made and then the action to report the choice is selected and executed (Newell & Simon, 1972). Thus, serial models neglect the role of action, assuming functional and temporal segregation between decision-related and motor processes. While serial models adapt well to the methodological constraints of laboratory settings, they fail to capture the range of behaviors observed in real-life scenarios in which decisions are deployed (Gordon et al., 2021).

Indeed, the temporal segregation between the decision and the response expression stages has been challenged by evidence of action preparation and initiation before decision completion (Cisek & Kalaska, 2005; McKinstry et al., 2008; Freeman et al., 2011; Selen et al., 2012), suggesting that decisional and motor processes may unfold in parallel. Consequently, parallel models of decision-making have been proposed, in which a continuous flow of information or evidence is transferred from the decision layer to the action component (Coles et al., 1985) and used to

update and adjust the motor response. Compared to serial models, the role of action is extended, starting early in the decision process and being continuously revised (Freeman et al., 2011; Selen et al., 2012). Both covert and overt components of action dynamics are now taken into account, reflecting uncertainty in the evolving decision-formation process (Lepora & Pezzulo, 2015). Continuous flow models can also explain changes of mind (Resulaj et al., 2009), i.e., revisions of decision once the response movement has already started: response initiation is triggered by the accumulated evidence passing a threshold, though new information or information not initially considered may be transferred to the motor system after movement initiation triggering the opposite response (Resulaj et al., 2009; Lepora & Pezzulo, 2015).

Parallel models consider only forward influences from the decision layer to action. However, from an ecological perspective, sensory information is not passively received from the external world but is shaped by active interaction with the environment. In decision-making contexts, this translates to the features of the action associated with the choice themselves influencing perceptual decisions (Gallivan et al., 2018).

In accordance with this notion, the Embodied Choice framework (Lepora & Pezzulo, 2015) considers action and decision-making as interdependent processes and emphasize the situated characteristics of choice. Action is no longer simply a way to express a choice but is part of the decision-making process, changing both the perceptual and the decision landscapes

(Gordon et al., 2021). Embodied decision models propose feedback influences from action dynamics to the decision process, which is biased by the constraints of the motor system (Lepora & Pezzulo, 2015). Specifically, the physical effort associated with the action to report the choice, i.e., the cost of action, can influence perceptual decisions towards the alternative with the lower cost (Marcos et al., 2013; Burk et al., 2014; Hagura et al., 2017). Motor costs are not determined a priori but may change depending on the action dynamics of already initiated responses (Lepora & Pezzulo, 2015). Once an action is initiated, the sensory evidence required to change the decision increases, as it must outweigh the commitment from having already selected a response: this generates commitment effects, which can also bias decisions and affect the occurrence of changes of mind (Thura & Cisek, 2014; Wiener et al., 2019; Cos et al., 2021).

Lepora and Pezzulo (2015) demonstrated that action preparation, changes of mind and commitment effects are all necessary to explain the full range of speed-accuracy tradeoffs observed in decision-making scenarios, supporting the embodied view of reciprocal interaction between action, perception, and decision-making.

# Interaction of Transitivity and Goal-directedness during Action Observation

## 2. 1 Introduction

Action Observation Network (AON) activity is modulated by a wide range of factors, mainly pertinent to features of the observed action, such as transitivity, kinematics of the movement and the goal of the actor (Kemmerer, 2021). Previous research primarily focused on investigating the neural correlates of action features in isolation, e.g., distinguishing transitive vs. intransitive (Villarreal et al., 2008; Handjaras et al., 2015; Kubiak & Króliczak, 2016) and meaningful vs. meaningless (Lui et al., 2008; Möttönen et al., 2016; Papeo et al., 2019) movements. Transitivity, which refers to whether actions involve interaction with and manipulation of objects, has been extensively investigated in research on action observation. Transitive and intransitive actions have been shown to recruit a shared network of areas within the AON (Villarreal et al., 2008; Króliczak & Frey, 2009; Andric et al., 2013; Kubiak & Króliczak, 2016), including the intraparietal sulcus (*IPS*), posterior superior and middle temporal cortex (*pSTS*, *pMTG*), and frontal areas such as ventral premotor cortex (*PMv*) and inferior frontal gyrus (*IFG*); however, object-directed actions elicit stronger and more widely spread responses than non-object-directed ones (Caspers et al., 2010; Kubiak & Króliczak, 2016).

Activity in the AON is also influenced by the goals of the observed actions (Kemmerer, 2021), defined in terms of the agent's abstract purposes. Goal-oriented actions include intransitive communicative gestures and transitive actions involving a meaningful interaction with the target object, such as grasping (Andric et al., 2013). Purposeful actions have been shown to recruit bilateral parietal and frontal areas (Jastorff et al., 2010; Andric et al., 2013; Wurm & Lingnau, 2015; Möttönen et al., 2016). Symbolic gestures, or emblems, constitute a particular subgroup of purposeful hand actions, in that they carry semantic content and subserve a communicative function; they are also distinct from speech-related gesticulation or pantomimes, as they are voluntary, learned and culturally shared, and function as independent units of speech (Xu et al., 2009). Compared to nonsense movements, symbolic gestures have been shown to evoke overlapping responses in frontal (e.g., *IFG*) and temporal (e.g., *MTG*) cortices (Lui et al., 2008; Möttönen et al., 2016; Papeo et al., 2019), whereas dissociations emerge in the parietal cortex: nonsense gestures recruit more dorsal regions in *IPS*/*SPL*, while emblems recruit the supramarginal gyrus (*SMG*) and inferior parietal lobule (*IPL*; Lui et al., 2008; Möttönen et al., 2016; Lesourd et al., 2018). Moreover, partial lateralization has been observed with emblems more strongly activating left *IFG* (Möttönen et al., 2016) and left *MTG* (Papeo et al., 2019). Indeed, emblems and speech engage partially overlapping areas in the temporal and frontal cortices (Xu et al., 2009; Andric et al., 2013; Papeo et al., 2019), indicating that observing symbolic gestures engages semantic processing and recruits some of the same

resources as language comprehension.

The evidence of partially overlapping representations across different action categories may suggest that they are not encoded independently. However, previous studies failed to thoroughly explore the neural representations of actions differing in multiple factors, limiting their findings to category-specific activations.

In the present functional magnetic resonance imaging (fMRI) study, we investigate the neural correlates of observing actions differing in both transitivity and purposefulness. Specifically, we measured neural responses to four types of actions: transitive, either purposeful or purposeless, and intransitive, either purposeful or purposeless; using a mixed effect approach, we assessed the main effects of transitivity and purposefulness, as well as their interaction. Contrasting all the possible combinations between factor levels allowed us to emphasize different features shared between or specific to each category and to explore how such features interact.

## 2. 2 Methods

### 2.2.1 Participants

Twenty-five healthy right-handed participants (11 males, mean age ± SD: 26±4 years) were enrolled in the fMRI protocol. All subjects received a medical examination, including a brain structural MRI scan, to exclude any disorder affecting brain structure or function. All participants gave their written informed consent after being explained the study procedures and potential risks. The study was conducted under a protocol approved by the University of Modena and Reggio Emilia Ethical Committee.

### 2.2.2 Stimuli and procedure

Participants were presented with a continuous video depicting an actor, whose arms and trunk only were visible, sitting at a table with nine everyday artificial objects placed in front of him; every 17.5 seconds, the actor performed a movement with the upper limb, which pertained to one of the following four classes: intransitive, either purposeful (i.e., symbolic) or purposeless (i.e., nonsense), and transitive, either purposeful (i.e., grasping an object) or purposeless, i.e., touching an object without hand pre-shaping for grasping. In the present study, we considered

touching as a purposeless gesture based on previous evidence that cortical activity associated with object-directed reaching is modulated by whether grasping is possible or not (Bozzacchi et al., 2012): this finding suggests that the act of touching an object without grasping it recruits different cognitive resources than when the goal of grasping is present. Some nonsense actions were taken from the Goldenberg apraxia test (Goldenberg, 1995), while others were created appositely for the study. Movies were presented using an in-house developed software via the IFIS-SA fMRI System (Invivo Corp, FL USA; visual field: 15°, 7'', 352 x 288 pixels, 60 Hz). Participants were instructed to observe and pay attention to the video while undergoing fMRI. Catch trials, lasting 20 seconds, were interspersed among experimental ones to ensure participants were attending the stimuli: during these trials, a static image of the same actor in the same environment was presented, and a red cross appeared on the screen as a cue for the subjects to press a button. Participants completed a total of 81 trials divided into three runs; in each run, six examples of each of the four actions were presented (24 experimental trials per run plus catch trials - 3 in the first run, 2 in the second and 4 in the third). No stimulus was repeated during each session.

### 2.2.3 Image acquisition and analysis

Functional brain images were acquired using a 3T Philips Intera scanner (TR = 2.5 s, TE = 34 ms; in-plane matrix size = 64 × 64; slice thickness = 4.6 mm; FOV = 243.2 x 243.2 x 115 mm; voxel

size = 3.8 x 3.8 x 4.6 mm; 25 sequential descending axial slices; 576 volumes in total for the three runs). T1-weighted anatomical images (FOV = 170 x 256 x 256 mm, matrix size = 170 x 256 × 256; 1 mm isotropic voxel; 170 sagittal slices) were acquired at the beginning of the experimental session. The anatomical volume was corrected for field inhomogeneity, skull stripped using the Advanced Normalisation Tools package (ANTs; Avants et al., 2009), co-registered to the functional image and then aligned to the MNI template (Fonov et al., 2009). Functional preprocessing was performed using the AFNI software package (Cox, 1996): scanner-related noise was removed through despiking (*3dDespike*), and all volumes comprising a run were temporally aligned (*3dTshift*) to correct for slice timing delays. The resulting time series were corrected for head motion (*3dvolreg*) using the first run as the reference volume; spatial smoothing with a Gaussian kernel of 6mm (full width at half maximum, *3dBlurToFWHM*) was applied, and each run was normalized as percent signal change. Single-subject preprocessed images were entered a GLM (*3dDeconvolve*), modeling each unique stimulus as a separate regressor and including motion parameters and outlier frames (Power et al., 2012) as noise regressors.

At the group level, a 2x2 mixed effects analysis was performed in AFNI using *3dMEMA* (Chen et al., 2012), which takes into account within-subject variability and weights each subject's contribution to the final statistics accordingly. The main effects for the transitivity factor (grasping + touching vs. symbolic + nonsense) and the goal-directedness factor (grasping + symbolic

vs. touching + nonsense) were calculated, as well as their interaction (grasping + nonsense vs. touching + symbolic). Group maps were corrected for multiple comparisons using cluster-based correction (voxel-wise threshold of 0.001, corrected *p*-value of 0.05).

## 2. 3 Results

### 2.3.1 Main effects of Transitivity and Goal-directedness

The main effect of transitivity ($p$<0.05, corrected; Fig. 2.1A) highlighted significantly greater activation to observed grasping and touching in the superior parietal lobule (*SPL*), *IPS*, collateral sulcus (*CS*) and in the right superior frontal gyrus (*SFG*). Conversely, voxels encoding intransitive actions were found in *IFG*, the superior temporal sulcus (*STS*), the median wall of the *SFG*, the thalamus, the posterior cingulate cortex, and the right lateral occipital cortex (*LOC*). Although significant activations to these gestures were found bilaterally, they extended over a larger area in the left hemisphere, indicating that processing of intransitive actions recruits left-lateralized brain regions to a greater degree.

The activation map for the main effect of goal-directedness (Fig. 2.1B) showed recruitment of left *STS* extending to *SMG*, medial *SFG* and left superior frontal sulcus (*SFS*) during observation of purposeful actions. On the other hand, touching and nonsense gestures elicited increased activity in *SPL* and the post-central sulcus (*PCS*).

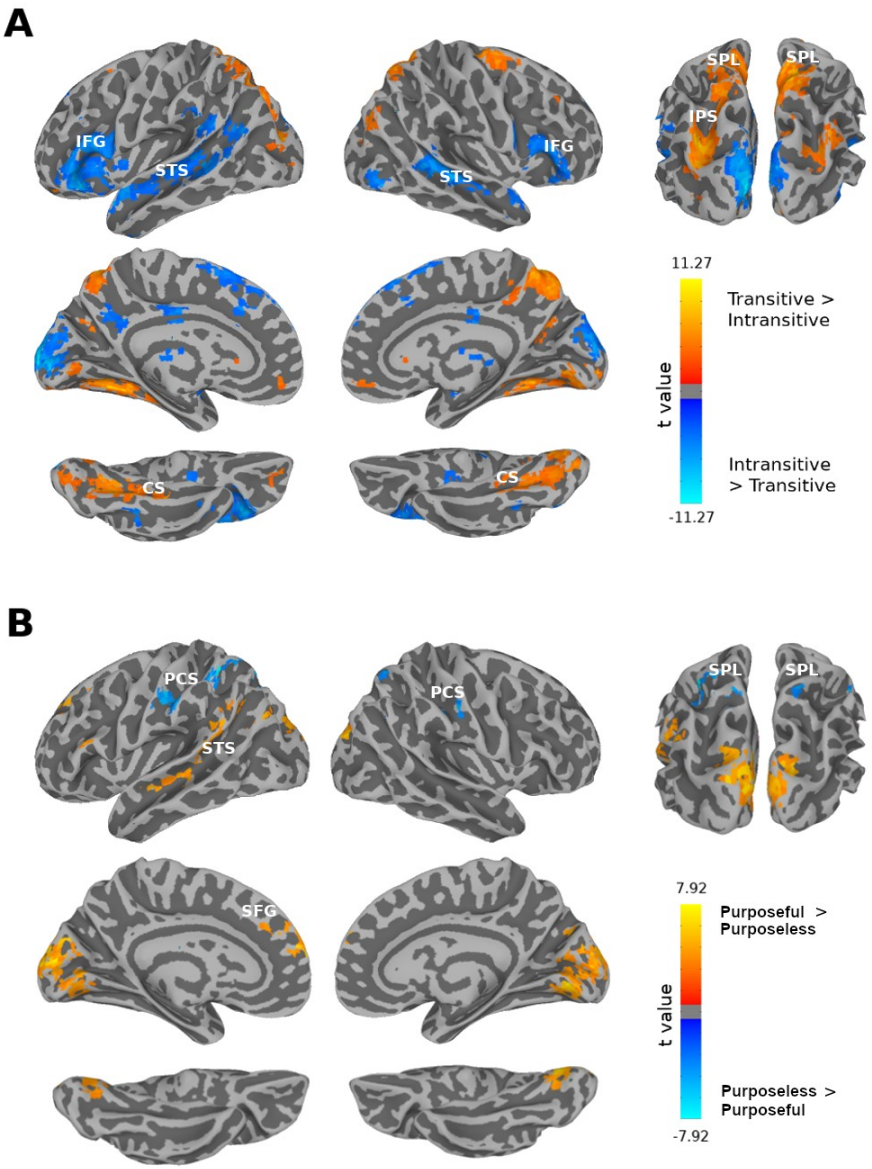



**Figure 2.1.** Brain areas encoding the dimensions of transitivity and goal-

directedness ($p < 0.05$, corrected). **A.** Main effect for transitivity factor: increased activity associated with grasping and touching (in yellow) and symbolic and nonsense actions (in blue). **B.** Main effect for goal-directedness factor: increased activity associated with grasping and symbolic actions (in yellow) and touching and nonsense actions (in blue). Statistical maps are projected onto the inflated surface template. *CS*: collateral sulcus; *IFG*: inferior frontal gyrus; *IPS*: intraparietal sulcus; *PCS*: post-central sulcus; *SFG*: superior frontal gyrus; *STS*: superior temporal sulcus; *SPL*: superior parietal lobule.

### 2.3.2 Interaction

Clusters of significant activations for the interaction effect between transitivity and goal-directedness ($p<0.05$, corrected) were found in an extended network of regions, spanning parietal, temporal and frontal cortices (Fig. 2.2): grasping and nonsense gestures compared to touching and symbolic actions elicited stronger responses in bilateral *SPL*, right *pMTG*, as well as right *SFG* and post-central sulcus. On the other hand, the left *SMG*, anterior cingulate cortex, precuneus and right central sulcus responded more strongly to touching and symbolic gestures.

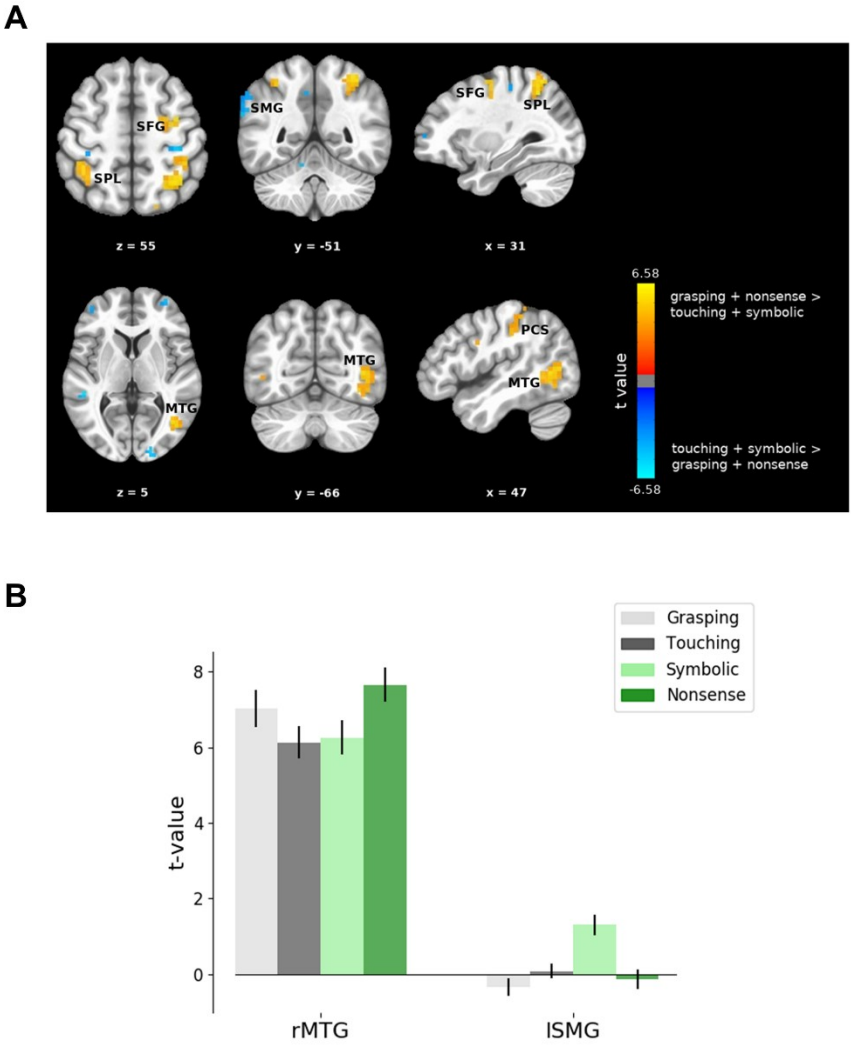



**Figure 2.2. A.** Brain areas showing significant interaction effect between factors ($p < 0.05$, corrected). The interaction effect between transitivity and goal-directedness was tested by contrasting grasping and nonsense actions against touching and symbolic actions (in yellow) and vice versa

(in blue). Statistical maps are rendered onto the ICBM152 template. **B.** Two brain areas showing an interaction effect. The right middle temporal gyrus responded more strongly to grasping and nonsense gestures compared to touching and symbolic actions, while the left supramarginal gyrus demonstrated the opposite response pattern. *MTG*: middle temporal gyrus; *PCS*: post-central sulcus; *SFG*: superior frontal gyrus; *SMG*: supramarginal gyrus; *SPL*: superior parietal lobule.

## 2. 4 Discussion

By contrasting actions differing in both transitivity and purposefulness, we emphasized different features shared between or specific to each category and investigated neural representations relating to particular features. The findings of a significant interaction effect between transitivity and goal-directedness and of partially non-overlapping activations across contrasts suggest that each type of action likely activates a distributed network of areas spanning the borders of commonly understood category-specific regions. Actions may be differentially encoded according to the combination of specific features, with categories differentially recruiting specific features-encoding regions.

*Interaction effect.* Significantly stronger responses for the grasping and nonsense vs. touching and symbolic contrast were identified in bilateral *SPL*, right *pMTG*, right *SFG* and post-central sulcus. Consistently with previous findings, parietal activation may reflect the processing of visuospatial features of hand actions (Caspers et al., 2010; Gallivan & Culham, 2015; Möttönen et al., 2016), e.g., hand postures, which would be the most salient feature characterizing both grasping and nonsense gestures. *pMTG* is known to be involved in the perception of biological motion (Peuskens et al., 2005; Zeki, 2015; Hodgson et al., 2022), but it has also been linked to transitivity, specifically to the processing of object features relevant to action recognition

(Wurm et al., 2017; Wurm & Caramazza, 2019). Responses to observed touching and symbolic gestures against grasping and nonsense were less pronounced and mainly located in left *SMG*; this area has been implicated in comprehension of intransitive gestures (Lui et al., 2008; Króliczak & Frey, 2009; Andric et al., 2013), but only in specific conditions suggesting it might serve a specialized function (Yang et al., 2015).

*Transitivity.* Results emerging from the transitivity contrast (i.e., transitive vs. intransitive actions) were in agreement with previous literature on action observation. Specifically, we found encoding of transitive actions in bilateral *SPL* and *IPS*, which have been implicated in the processing of directional arm movements and object features relevant for manipulation (Gallivan et al., 2013; Gallivan & Culham, 2015; Cavina-Partesi et al., 2018). This result provides further support to the well-established involvement of parietal regions during observation of object-oriented hand actions (Culham & Valyear, 2006; Caspers et al., 2010; Abdollahi et al., 2013; Gallivan & Culham, 2015; Reynaud e al., 2019; Aflalo et al., 2020; Urgen & Orban, 2021). Conversely, observed intransitive gestures recruited *IFG* and *STS*, consistently with evidence that these areas are involved in processing communicative actions (Villarreal et al., 2008; Yang et al., 2015; Möttönen et al., 2016). Moreover, this category of actions recruited left cortices to a greater extent, in accordance with findings of shared neural substrates for gestures and language comprehension (Xu et al., 2009; Andric et al., 2013; Papeo et al., 2019).

*Goal-directedness.* Results from the contrast between purposeful and purposeless actions are less conclusive: we found significant activations associated with purposeful actions in left *STS* and medial *SFG*, while activation of *SPL* and post-central sulcus was associated with observation of purposeless movements. These areas do not match brain regions areas identified by previous studies as representing action goals, e.g., *IPL* and *IPS* and posterior temporal cortex (Jastorff et al., 2010; Wurm & Lingnau, 2015; Wurm et al., 2016). A possible explanation for this difference is that rather than contrasting purposeful vs. purposeless actions, these studies tested discrimination of transitive action types generalizing across effector and object, thus identifying neural correlates of representing action identity at an abstract level rather than actor's intentionality. Another possibility is that, in our study, participants may have attributed a communicative intent to nonsense gestures, similar to emblems; this would explain why activations for the purposeful vs. purposeless contrast were not particularly extended, as well as the lack of activation in *IFG*, which has been implicated in representing purposeful hand action (Andric et al., 2013; Möttönen et al., 2016; Koul et al., 2018; Savaki et al., 2022).

In conclusion, we expanded on previous findings by demonstrating a significant interaction between the categories of transitivity and goal-directedness, indicating that these two dimensions are not represented in an independent manner.

Going beyond the view that different brain areas are dedicated to the processing of specific action classes, our results indicate that actions may be differentially encoded according to the combination of specific motor (i.e., kinematics, affordances) and cognitive (i.e., semantics) features, with categories differentially recruiting specific features-encoding regions depending on the distinct combination of such characteristics.

# Sensitivity and Specificity of the Action Observation Network to Kinematic, Target Object and Gesture Meaning

## 3. 1 Introduction

Several behavioral, neuropsychological, and brain-functional observations support a model of action processing based on a hierarchical organization (Filimon, 2010; Kilner, 2011; Gallivan & Culham, 2015; Tarhan et al., 2021). Within this framework, action-related information is thought to be represented across different levels of abstraction, with brain areas being differentially activated as a function of information complexity, from low-level appearance and kinematics (e.g., effector, the trajectory of the arm, grip configuration, the orientation of the hand to the object), to object-related features (i.e., identity and function of the object), to higher-level goal and action outcomes (Grafton & Hamilton, 2007). In support of this perspective, neuroimaging studies report the existence of different levels of feature representation during both observation and execution (Gallivan et al., 2013; Wurm & Lingnau, 2015; Urgen et al., 2019; Turella et al., 2020).

The visual processing of action-related information is organized over the Action Observation Network (AON), spanning separate anatomical areas (Culham & Valyear, 2006; Caspers et al., 2010; Kilner, 2011; Filimon et al., 2015; Reynaud et al., 2019;

Vannuscorps et al., 2019). The AON relies on distributed representations, where low-level visual features (e.g., biological motion; Giese & Poggio, 2003; Zeki et al., 2015) are processed in specific patches of the cortex, while more abstract properties (e.g., semantics) are encoded at a larger spatial scale (Giese & Rizzolatti, 2015; Handjaras et al., 2015; Tucciarelli et al., 2019). Consequently, distinct action representations overlap across a wide expanse of the cortex and support a high-dimensional representational space (Filimon et al., 2015; Handjaras et al., 2015; Wurm & Caramazza, 2019).

In accordance with this functional perspective of action representation, the present study aims to better characterize the complex hierarchy of the AON by studying action organization at different levels of granularity (i.e., differentiation), from the representation of elementary descriptors, such as kinematics, object categories and their mutual interactions, to the representation of higher-level information, such as the meaning of gestures. Consistently with previous accounts of distributed representations, we propose a novel functional parcellation of the cortex based on representational similarity, where the clustering of voxels is guided by their information content rather than their anatomical proximity. Our approach aims to better characterize the intrinsic dimensionality and organization of the AON by testing whether and to what extent similar categorical information is distributed in multiple functional parcels when the anatomical proximity constraint is deliberately neglected. For this purpose, representational similarity analysis (RSA;

Kriegeskorte et al., 2008a) and clustering techniques (Maaten & Hinton, 2008) were employed in two fMRI experimental protocols to describe the representational geometries of the AON by examining multiple sets of theoretical categorical models and using the sensitivity index *d'* (Macmillan & Creelman, 2004) to measure discriminability of each action model. In the first fMRI experiment, we identified functional clusters of voxels encoding observed object-directed actions and investigated their specific tuning to different dimensions characterizing transitive actions (i.e., kinematic trajectory, animacy, and target object category), which have been previously shown to modulate brain responses during action observation (Kemmerer, 2021). In the second fMRI experiment, we tested the reliability of voxel responses to a different set of transitive gestures. Finally, by looking at the representation of intransitive gestures and quantifying the cortical overlap among action properties, we evaluated the sensitivity and specificity of the AON.

We expected the same brain regions to participate at multiple levels of action representation and the great majority of the information content to be spread across the cortex. This result would favor the hypothesis of a unique, high-dimensional space for action representation rather than a representational system with a modular organization based on segregated action features.

## 3. 2 Methods

Two independent fMRI experiments involving two different action observation paradigms were used in the present study. The second experimental protocol has been already described in the previous chapter (chapter 2). Both experimental protocols employed visual presentation of action stimuli in the form of video clips: in the first protocol, distinct clips depicting only transitive actions directed at different categories of objects were presented; in the second protocol, transitive and intransitive actions were embedded in a continuous visual stimulation. The dimensions characterizing the stimuli feature space in the first experiment were: the kinematic trajectory of transitive actions, animacy, and category of the target of the action. The features taken into consideration in the second experiment were: the kinematics and the target object identity of transitive actions and the meaning of non-object-directed, intransitive gestures.

Representational similarity analysis (RSA) was performed in both experiments. Brain voxels were then clustered based on their similarity structures, and the resulting representational spaces were compared to multiple categorical models to highlight, in any given cortical region, the dominant dimensions driving stimulus representation.

### 3.2.1 Participants

Fourteen healthy right-handed subjects (4 males, mean age ± SD: 37 ± 6 years) participated in the first study. Twenty-five healthy right-handed participants (11 males, mean age ± SD: 26±4 years) were enrolled in the second experiment.

All participants received a medical examination, including a brain structural MRI scan, to exclude any disorder affecting brain structure or function. All participants gave their written informed consent after being explained the study procedures and potential risks. The studies were conducted under protocols approved by the University of Modena and Reggio Emilia Ethical Committees.

### 3.2.2 Stimuli and procedure

*Experiment 1*

Subjects were presented with movie clips depicting a hand-performed transitive action, represented in a third-person perspective, with only the agent's right arm and hand visible. The performed action varied in three categorical features: one kinematic-based dimension ('grasping to lift', 'pushing away', 'putting down'), and two object-based dimensions, i.e., animacy of the distal object (ten animate and ten inanimate exemplars for each kinematic-based dimension) and its category (i.e., animals, body parts, natural objects, such as vegetables and fruits, artificial objects). For a complete list, please refer to

Supplementary Table 1. The sixty clips were 3 seconds long, presented with an inter-trial interval of 7 seconds, and each clip was filmed twice. Thus, participants observed 120 clips, randomly alternating between different action types.

*Experiment 2*

The experimental procedure for the second experiment has already been described in detailed in section 2.2.2. Briefly, participants were presented with a continuous video depicting an actor performing an upper-limb movement every 17.5 seconds, which pertained to one of the following four classes: transitive, either grasping an object or touching an object, and intransitive, either symbolic or nonsense. For a complete list, please refer to Supplementary Table 2. Catch trials were interspersed among experimental ones to ensure participants' attention. Participants completed a total of three runs; in each run, six examples of each of the four actions were presented and no stimulus was repeated during each session.

### 3.2.3 Image acquisition and preprocessing

In the first experiment, functional brain images were acquired during stimulus presentation using a 3T Philips Intera scanner (gradient echo echo-planar images, 35 axial slices, TR 2.5s, TE 35 ms, 3 mm isovoxel) with an fMRI six runs slow event-related design. T1-weighted spoiled gradient recall images (TR 9.9 ms, TE 4.6 ms, 170 sagittal slices, 1 mm isovoxel) were obtained for each participant to provide detailed brain anatomy. In the

second experiment, functional brain images were acquired using the same scanner (TR: 2.5 s, TE: 34 ms, 25 axial slices, voxel size: 3.8x3.8x4.6 mm). T1-weighted anatomical images (170 sagittal slices, 1 mm isovoxel) were acquired at the beginning of the experimental session.

In both protocols, visual stimulation was provided with the same equipment, using an in-house developed software via the IFIS-SA fMRI System (Invivo Corp, FL USA; visual field: 15°, 7'', 352 x 288 pixels, 60 Hz).

MRI and fMRI data were preprocessed following the same pipeline for both experiments. The anatomical volume was corrected for field inhomogeneity, skull-stripped using the Advanced Normalization Tools package (ANTs; Avants et al., 2009), co-registered to the functional image and then aligned to the MNI template (Fonov et al., 2009). Functional data were processed using AFNI (Cox, 1996). Despiking (*3dDespike*) and slice timing correction (*3dTshift*) were applied to each functional volume; the resulting time series were motion corrected (*3dvolreg*) and spatially smoothed at full-width-at-half-maximum Gaussian kernel of 6 mm (*3dBlurToFWHM*). Single-subject images were entered into a GLM (*3dDeconvolve*), modeling each unique stimulus as a separate regressor and including motion parameters and outlier frames (Power et al., 2012) as noise regressors.

### 3.2.4 Representational similarity analysis

For both datasets, t-score maps from the GLM were spatially normalized to the MNI template and resampled into 2 mm isovoxels. A volume-based searchlight approach (Kriegeskorte et al., 2006) was employed, restricting the analysis to grey matter cortical voxels: for each voxel, data was extracted from a searchlight of a 10 mm radius, normalized and de-noised with principal component analysis (PCA), retaining only those components that explained 75% of the variance; median activity patterns were computed across subjects. Stimuli activity patterns were then compared to each other using a dissimilarity metric (1 - Pearson's r correlation coefficient) to obtain a Representational Dissimilarity Matrix (RDM; Kriegeskorte et al., 2008a).

Within each searchlight, the ability to discriminate between categorical features was evaluated with sensitivity index $d'$ (Macmillan & Creelman, 2004), which quantifies the separation between the average dissimilarity computed between stimuli of different classes and the average within-class dissimilarity. The index was calculated as:

$$d' = \frac{\bar{d}_b - \bar{d}_w}{\sqrt{\frac{1}{2}(\sigma_b^2 + \sigma_w^2)}}$$

where $\bar{d}_b$ is the average between-category dissimilarity, $\bar{d}_w$ is the average within-category dissimilarity, $\sigma_b^2$ is the variance of between-category dissimilarity, and $\sigma_w^2$ is the within-category dissimilarity variance. $d'$ is equal to 0 if the between-category

and within-category distances are comparable, and increases as the between-category dissimilarity broadens (Fig. 3.1A).

The first dataset was tested comparing single voxel RDMs to four theoretical categorical models; the models were constructed based on the kinematic dimension (kinematic model, i.e., 'grasping to lift', 'pushing away', 'putting down'), the animacy dimension (animacy model, i.e., animate vs. inanimate), the category of the action target (category model, i.e., animals, body parts, natural and artificial objects), and the specific combination of the kinematic and animacy dimensions (i.e., the granularity model). The granularity model encoded actions as gestures with specific kinematics and acted upon distinct targets. The resulting representational space was characterized by a higher level of specificity in the action hierarchy (Fig. 3.1B). The second experiment tested five theoretical categorical models, evaluating the following dimensions (Fig. 3.1C): transitivity (i.e., object-directed actions considered as within category), kinematics (i.e., 'grasping' and 'touching' considered as distinct categories), object identity (i.e., every unique object considered as a category) and meaning of intransitive gestures (i.e., symbolic and nonsense considered as separate categories).

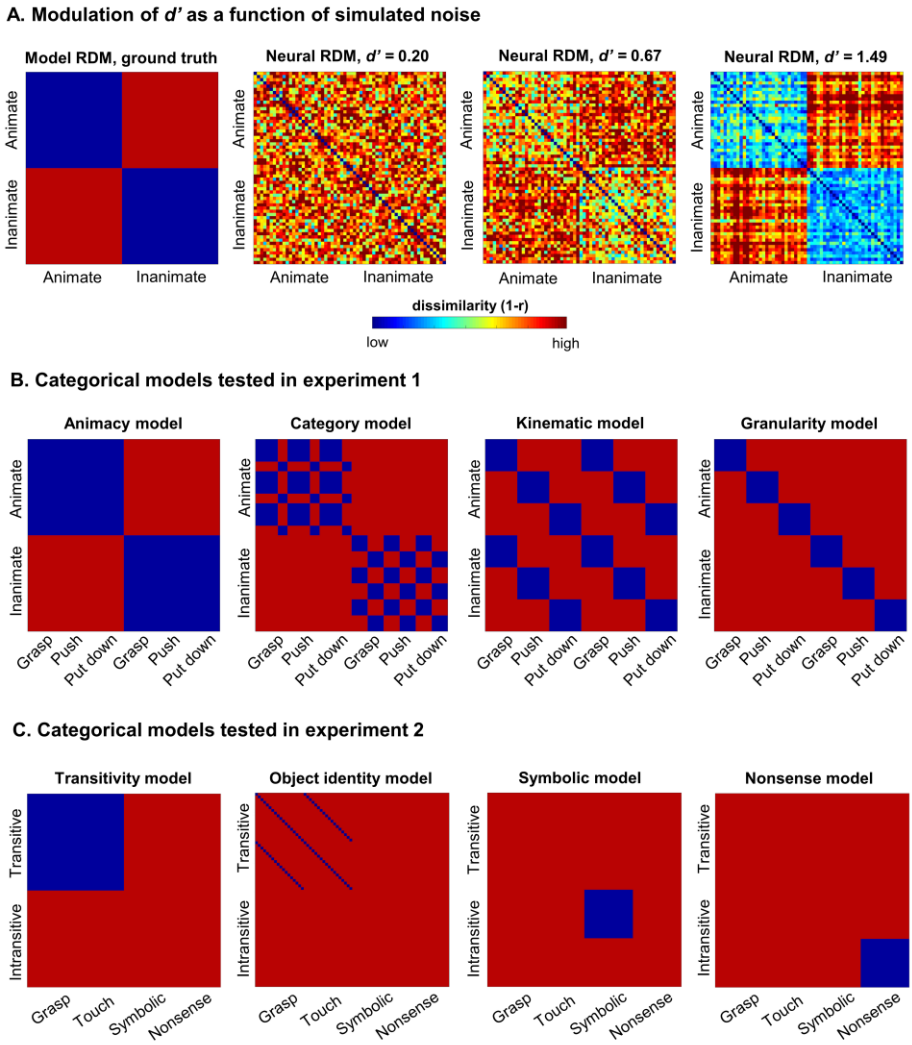



**Figure 3.1**. Categorical models used to obtain $d'$ values. Sensitivity index $d'$ is computed by contrasting the average between-category dissimilarity (in red) with the average within-category dissimilarity (in blue). If a voxel is tuned to a categorical dimension, dissimilarity averaged across blue squares should be lower than dissimilarity

averaged across red squares, resulting in a higher *d'* value. **A.** Examples of *d'* computed by comparing simulated data with decreasing noise levels (from left to right) for the animacy model; *d'* increases as a function of dissimilarity. **B.** Models tested in the first experiment: animacy model assessed voxel tuning to animacy of the target object; kinematic model tested voxel tuning to action type, independently of the action target; category model tested voxel tuning to target semantic domain; granularity model is built on the specific combination of movement type and animacy dimension of the target, describing a higher level of specificity in the action hierarchy. **C.** Models tested in the second experiment: transitivity model assessed voxel tuning to transitive actions using a different set of stimuli to demonstrate the reliability of voxels response; the object identity model assessed the generalizability of voxel tuning to object properties and the level of specificity in their representational content; symbolic and nonsense models tested discriminability of intransitive gestures to evaluate the specificity of voxel tuning.

Statistical significance of the obtained *d'* values was tested using a permutation test: for each voxel, a null distribution of *d'* values was obtained by shuffling the stimuli labels 1,000 times and computing dissimilarity and *d'* on the shuffled activity patterns; the exact *p*-value for each voxel was computed approximating the tail of the null distribution to the Pareto distribution (Winkler et al., 2016) and False Discovery Rate (FDR; Benjamini & Yekutieli, 2001) was used to correct for multiple comparisons.

### 3.2.5 Clustering and voxel tuning

In order to isolate brain regions responding to action observation, voxels from the first dataset with statistically significant *d'* values for the granularity model ($q$<0.01) were selected. t-distributed Stochastic Neighbor Embedding (t-SNE; Maaten & Hinton, 2008) was applied on the voxels RDMs obtained during the searchlight analysis and projected onto a 2D embedding space so that spatially close points reflect voxels with similar representational geometry. The *k-means* algorithm (Lloyd, 1982) was then used to identify the main clusters based on the spatial distance between voxels projections (Fig. 3.2).

Tuning of selected voxels to different stimulus features was evaluated by color coding *p*-values associated with *d'* for different models and mapping them onto the space defined by the t-SNE to highlight the relative contribution of each voxel to a specific dimension. Firstly, voxel sensitivity to kinematic, animacy, and object category dimensions was represented by associating each one to a different RGB color scale channel (Fig. 3.3). Secondly, data from the second experiment was extracted at the same coordinates of selected voxels, and *p*-values associated with the discriminability of object-directed actions were mapped onto the t-SNE space; the contribution of object identity was also assessed (Fig. 3.4). To explore the representational geometry at a larger scale, we selected the clusters with the highest *d'* values for the kinematic and the object-based models, and performed multidimensional scaling (Kruskal, 1964) on the RDMs (Fig. 3.5):

multidimensional scaling transforms information about the dissimilarity between stimuli into Euclidean distances, so that spatial arrangement of individual data points reflects activity pattern similarity, and allows to visualize the dominant features that drive stimulus representations, without any a priori assumption about the representational organization (Kriegeskorte et al., 2008b).

Lastly, we tested the specificity of selected clusters to transitive actions by evaluating the discriminability of intransitive gestures from the second experiment, that is, by mapping sensitivity to symbolic versus nonsense gestures (Fig. 3.6).

## 3. 3 Results

### 3.3.1 Sensitivity of the AON to action features

Brain areas showing sensitivity to both kinematic- and object-related features and therefore encoding actions at a high differentiation level were identified by selecting voxels associated with a statistically significant *d'* (granularity model, $q$<0.01, FDR corrected; Fig. 3.2A).

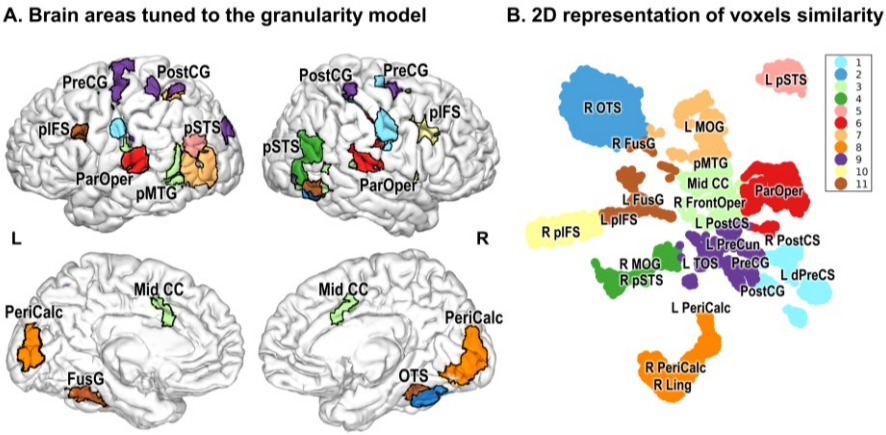



**Figure 3.2. A**. Areas showing significant sensitivity to both kinematic and object-based features (i.e., granularity model): voxels were selected by computing *d'* on the RDM ($q$<0.01, FDR corrected). **B.** Selected voxels were projected as points onto a 2D embedding space defined on RDM similarities, using t-SNE: the relative distance of voxels/points reflects similarities in the representational space; based on their spatial distance,

voxels were grouped into 11 clusters using the *k-means* algorithm. L: left; R: right; *FrontOper*: frontal operculum; *OTS*: occipito-temporal sulcus; *dPreCS*: dorsal precentral sulcus; *PostCS*: dorsal postcentral sulcus; *FusG*: fusiform gyrus; *Ling*: lingual gyrus; *Mid-CC*: middle cingulate cortex; *MOG:* middle occipital gyrus; *ParOper*: parietal operculum; *PeriCalc*: pericalcarine cortex; *pIFS*: posterior inferior frontal sulcus; *pMTG*: posterior middle temporal gyrus; *PostCG*: postcentral gyrus; *PostCS*: postcentral sulcus; *PreCS*: precentral sulcus; *PreCun*: precuneus; *pSTS*: posterior superior temporal sulcus; *TOS*: transverse occipital sulcus.

Significant voxels were found in ventral occipito-temporal cortices (occipito-temporal sulcus – *OTS*, fusiform gyrus – *FusG*, and lingual gyrus – *Ling*), sensory-motor cortices (precentral sulcus – *PreCS,* postcentral sulcus – *PostCS*, postcentral gyrus – *PostCG,* and parietal operculum – *ParOper*), posterior temporal areas (superior temporal sulcus – *pSTS,* and middle temporal sulcus – *pMTG),* occipital areas (middle occipital gyrus – *MOG*, pericalcarine cortex – *PeriCalc*, transverse occipital sulcus – *TOS)*, precuneus (*PreCun*), middle cingulate cortex (*Mid-CC)* and posterior inferior frontal sulcus (*pIFS*). These areas are consistent with previous literature, matching the well-known network of regions involved in action observation (e.g., Hardwick et al., 2018).

Next, a t-SNE dimensionality reduction was applied to perform a functional parcellation that grouped brain regions based on their similarity in the representational space without considering anatomical constraints. Using a clustering algorithm, 11 clusters

were identified (Fig. 3.2B), with individual clusters encompassing anatomically distant brain regions.

The relative contribution of the kinematic, animacy and category dimensions at the voxel level was mapped to evaluate the tuning of selected voxels to different stimulus features. Results showed a partial dissociation between the kinematic-based and the object-based dimensions (Fig. 3.3A): inferior temporal and premotor areas retained primarily object-related information, while precentral/postcentral and temporo-parieto-occipital areas exhibited higher sensitivity to kinematic-related features, reflecting the dichotomy between ventral and dorsal streams (Goodale & Milner, 1992).

Notably, 30% of the selected voxels showed significant tuning ($q$<0.05, FDR corrected) to all three dimensions (Fig. 3.3B), revealing overlapping representations of action features. Considerable overlap for the category and animacy dimensions was found in most clusters (Fig. 3.3C), particularly in inferior temporal areas showing sensitivity to both features.

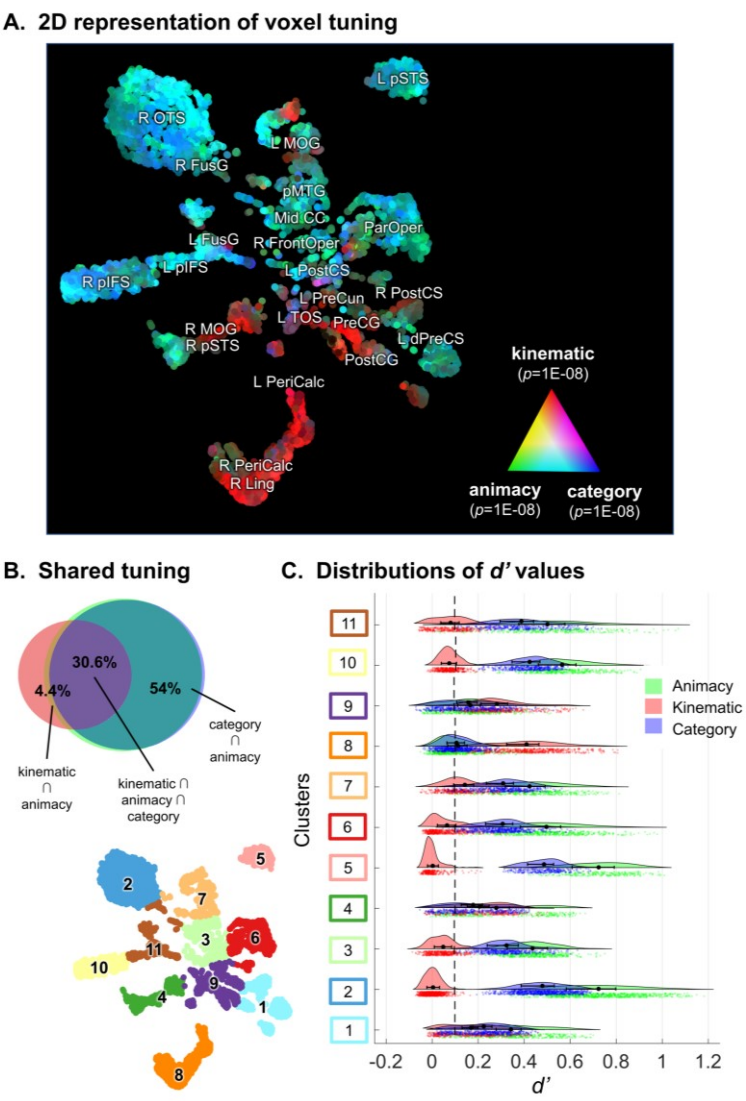



**Figure 3.3. A.** Dimensions describing the first set of stimuli were mapped using the *p*-value associated with *d'* onto the space defined by the t-SNE; kinematic dimension was coded in red, animacy in green and object category in blue; maximal saturation of each channel reflects

an uncorrected $p$ < 1E-08. **B.** The Venn diagram represents the proportion of voxels tuned to each dimension: voxels associated with a statistically significant $d'$ ($q$<0.05, FDR corrected) were 45.8%, 89.1% and 86.2% for kinematic, animacy and category, respectively; 4.4% of the voxels showed sensitivity to both kinematic and animacy dimensions, but not to object category, 54.0% were tuned to both animacy and object category but not to kinematic features, while 9.3% were tuned to kinematic features only; 30.6% of the voxels showed sensitivity to all three dimensions. **C.** $d'$ distributions and clusters averages (black dot) for the animacy, kinematic and category dimensions; bootstrap confidence intervals (95%) were computed for each $d'$ value and averaged across voxels (solid black line) within a cluster; the critical value at $\alpha$ = 0.05 (dotted line) was obtained from the null distribution.

### 3.3.2 Generalizability of the AON to different kinematics, objects and action categories

To assess whether identified clusters were consistently tuned to transitive gestures, the discriminability of transitive actions was further tested in the second experiment (Fig. 3.4A, left plot): roughly 73% of the identified voxels were still significant ($q$<0.05, FDR corrected; Fig. 3.4B), demonstrating the generalizability of the effect to a new stimulus set and tuning to a different feature space characterizing transitive actions.

Sensitivity to object identity was also tested using stimuli from the second experiment to confirm tuning to object-related features. Object identity discriminability was observed in most selected voxels, indicating high specificity in their representational content (Fig. 3.4A, right plot).

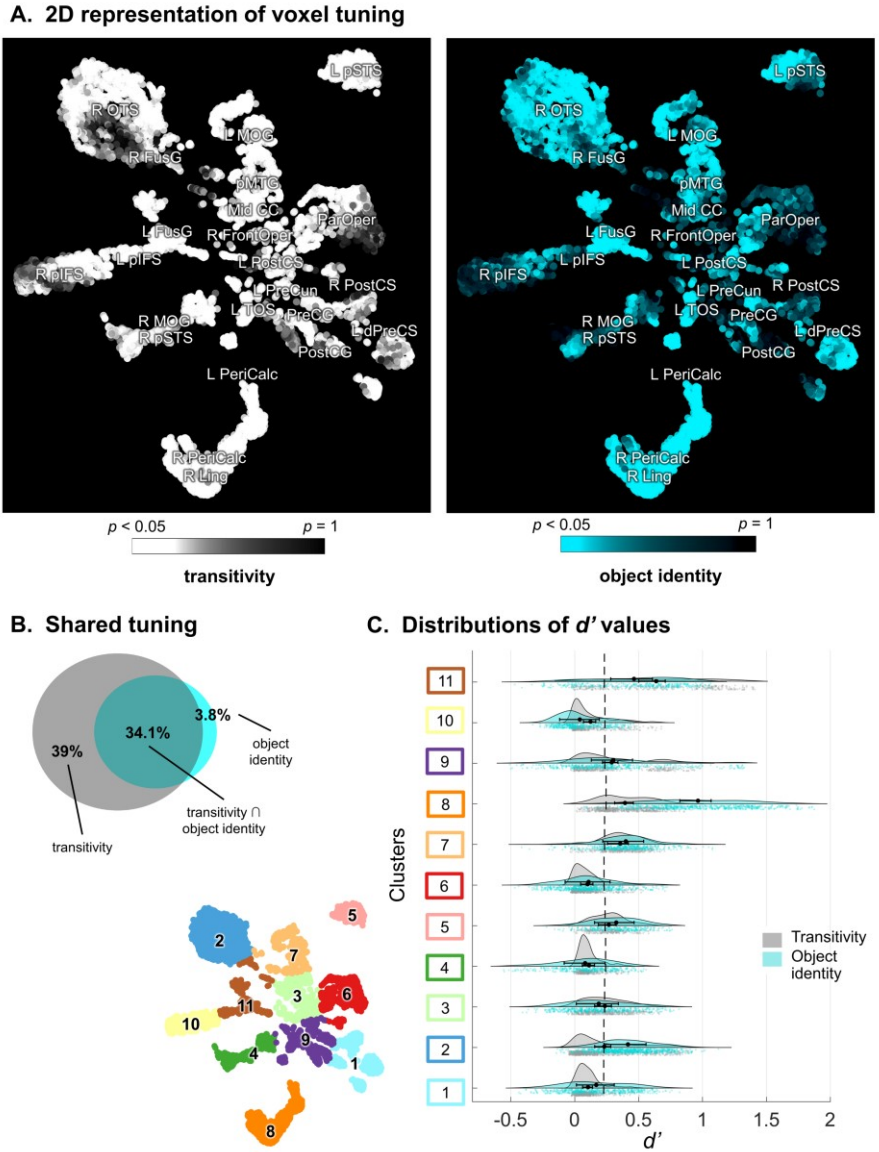



**Figure 3.4. A.** Voxels tuning to transitive actions was further tested using an independent dataset, and *p*-values from the second set were

mapped onto the same 2D space. Voxels associated with uncorrected *p*-values < 0.05 were mapped in white (left). Sensitivity to the object-identity feature was also mapped, and voxels associated with uncorrected *p*-values < 0.05 were colored in cyan (right). **B.** The Venn diagram represents the proportion of voxels tuned to each dimension: voxels associated with a statistically significant $d'$ ($q$<0.05, FDR corrected) were 73.2 % and 37.9% for transitivity and object identity, respectively, with 34.1% of voxels tuned to both dimensions. **C.** $d'$ distributions and clusters averages (black dot) with 95% bootstrap confidence intervals for transitivity and object identity dimensions; dotted lines represent the critical value at $\alpha$ = 0.05 obtained from the null distribution.

Substantial tuning to this feature was found in the left superior temporal cortex and bilaterally in the occipito-temporal cortex, fusiform gyri, and occipital regions, including the pericalcarine cortex.

Moreover, most of the voxels significantly encoded the categories of movement types (grasping vs. touching): voxels tuned primarily to grasping were found in frontal areas, such as the precentral gyrus, while touching-only tuning was observed in occipito-temporal areas, such as the middle occipital gyrus, the occipito-temporal sulcus and the posterior middle temporal gyrus, and parietal areas (see Supplementary Fig. S1).

To further characterize the organization of action representations along different dimensions, clusters with the highest sensitivity for object-related features, aside from early visual areas, were selected (Fig. 3.3C). These clusters included: cluster 2, a set of cortical areas pertaining to the ventral visual stream (e.g.,

occipito-temporal sulcus, fusiform gyrus), and cluster 5, encompassing the left superior temporal cortex. We also identified two clusters for the kinematic model: cluster 4, which included the right middle occipital cortex and superior temporal cortex, and cluster 9, comprising dorso-occipital, superior parietal and sensorimotor cortex and pertaining to the dorsal visual stream. Multidimensional scaling on the first experimental protocol showed that the same dimension-based dissociation found at the voxel level was also present at the cluster level: stimulus representations were organized primarily around the animacy of the object in clusters 2 and 5 and around movement types in clusters 4 and 9 (Fig. 3.5A). The same effect, although less pronounced, was found when performing multidimensional scaling on the second experimental protocol (Fig. 3.5B): representations in clusters 4 and 9 were organized primarily around kinematic-based features, as revealed by the separation in the representational space between grasping and touching, whereas the same organization was not present in clusters 2 and 5.

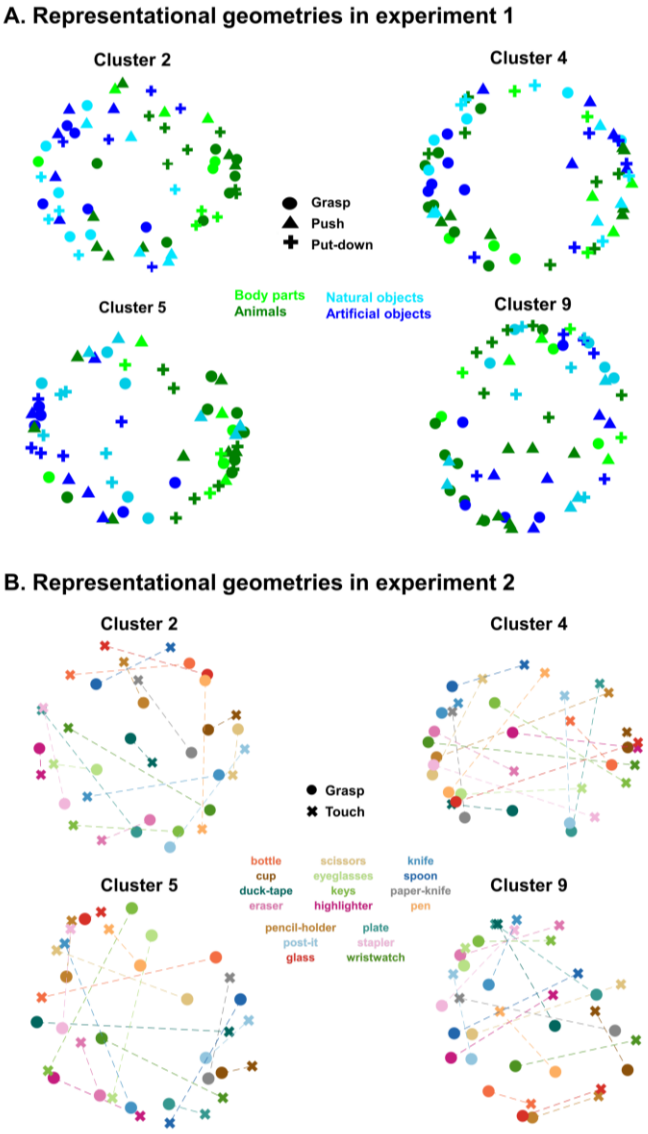



**Figure 3.5.** Multidimensional scaling for clusters 2 and 5 from the ventral stream, and 4 and 9 for the dorsal stream. Multidimensional scaling was performed on the cluster RDM. Euclidean distances

between data points reflect activity pattern similarity between stimuli. **A.** First experimental protocol: action type is coded by symbols, while animacy and object category are color-coded; in clusters 2 and 5, animate and inanimate objects are further apart, reflecting higher *d'* for the animacy model, while the effect for object category is weaker, but still present; in clusters 4 and 9, stimuli are grouped closer together based on action type, rather than object-related dimensions. **B.** Second experimental protocol: action type is coded by symbols, object identity is color-coded, and dashed lines connect data points representing the same object.

Lastly, we tested the sensitivity of the AON to different categories of actions, a set of intransitive gestures which do not involve objects and subserve symbolic and/or communicative functions. Specifically, we considered two classes of intransitive actions, meaningful symbolic (e.g., waving) and meaningless, nonsense gestures. Information related to intransitive gestures was represented in the AON: results showed that areas coding transitive actions in the first experiment generalized to novel features, with voxels exhibiting significant tuning to both symbolic and nonsense stimuli of the intransitive dimension (Fig. 3.6).

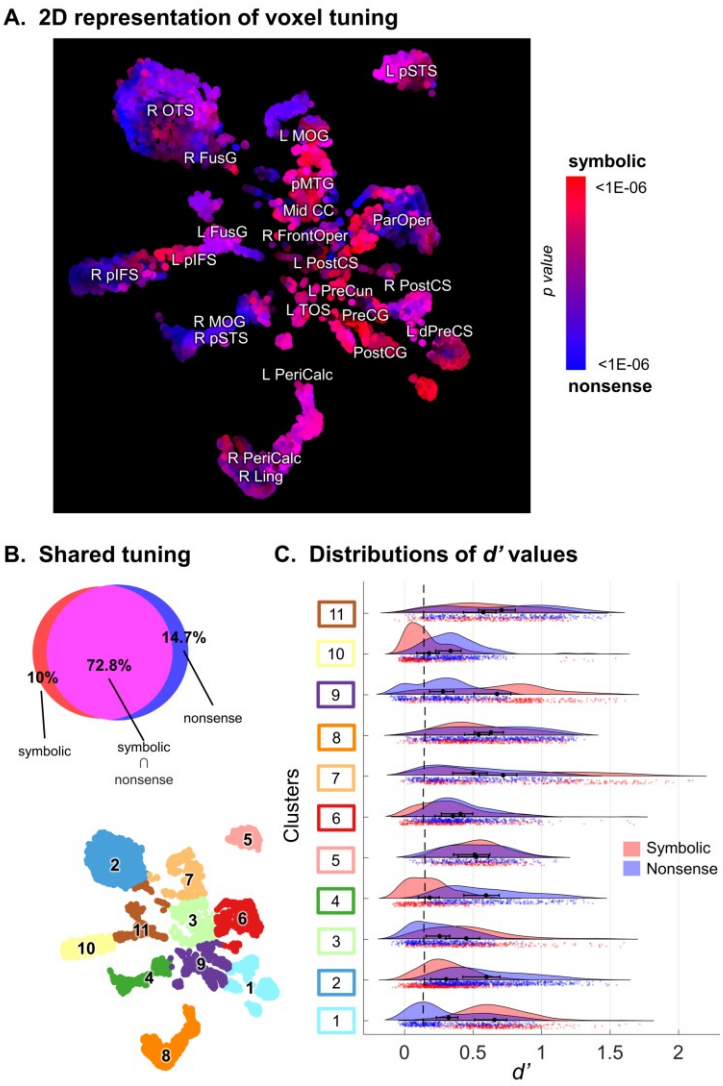



**Figure 3.6. A.** The same voxels identified through their tuning to transitive action features were tested for sensitivity to other action dimensions (meaning of intransitive actions), and the relative

contribution of voxels' sensitivity to symbolic to nonsense gestures was mapped with a color scale ranging from red to magenta to blue; maximal saturation of each channel reflects an uncorrected $p < 1E\text{-}06$. **B.** The Venn diagram represents the proportion of voxels tuned to each dimension: voxels associated with a statistically significant $d'$ ($q<0.05$, FDR corrected) were 82.8% and 87.5% for symbolic and nonsense gestures, respectively, with 72.8% of voxels tuned to both stimuli. **C.** $d'$ distributions for clusters 2, 4, 5 and 9 with cluster average (black dot) and 95% bootstrap confidence intervals; the dotted line represents the critical value at $\alpha = 0.05$ obtained from the null distribution.

## 3. 4 Discussion

Using representational similarity analysis and measuring sensitivity index *d'*, we localized cortical areas responding to a high level of representation (granularity model), where actions were encoded as gestures with specific kinematics and distinct targets. These AON regions were projected onto a two-dimensional embedding space based on voxel similarity to extract a set of functional clusters representing the various action features. Replicability and generalizability of action representation were tested in a second experiment using transitive gestures involving different target objects and kinematics. Finally, voxel specificity to the transitivity dimension was evaluated by means of a set of intransitive gestures modulating their information content. We provided evidence only for a partial dissociation between kinematic, object characteristics and meaning in the occipito-parietal, ventro-temporal and lateral occipito-temporal cortex, respectively. Notably, the great majority of the AON exhibited low specificity to all the explored action dimensions, and its representational content did not depend on proximity constraints, as anatomically distant cortical patches shared similar tunings.

### 3.4.1 A common representational space for action recognition

Wanting to explore action representation across the action hierarchy, we first selected brain areas responding to action observation at a high level of differentiation, i.e., tuned to the specific combination of different action features (e.g., grasping an inanimate target). Then, we assessed voxel tunings to action dimensions and across different sets of transitive and intransitive gestures. We found considerable overlap between the representation of action features in the AON. Specifically, of the voxels identified through the first experiment, 30% were tuned to all three dimensions (i.e., kinematic, animacy and target category), 34% responded to both transitivity and object identity in the second experiment, and 73% of voxels encoded intransitive gestures, regardless of their meaning. This evidence suggests that the same cortical patches participated in representations of multiple stimuli characterized by diverse features.

Moreover, we purposely neglected anatomical constraints in defining our clusters, which were identified based solely on their representational content. The resulting functional parcellation included anatomically distant areas grouped in the same cluster, such as the inferior temporal cortex and the inferior frontal sulcus in cluster 11, which shared related tunings. This functional parcellation, together with the finding of substantial overlap in features tuning, supports the notion that the

organization of the AON relies on overlapping and distributed representations, whereby, on the one hand, the same brain regions participate at multiple levels of the representation, and, on the other, the same information is spread across the cortex without being necessarily constrained by anatomical proximity.

This organization of the AON is also supported by clinical and neuromodulation studies and evidence on the elusive association between brain regions and deficits using lesion mapping (Mah et al., 2014). Specifically, impairments in recognition of both transitive and intransitive actions have been reported following lesions to the *IFG* (Pazzaglia et al., 2008) or by applying TMS on the frontal cortex (Ward et al., 2022). Left hemisphere stroke patients with damage to *pMTG* showed lower performance on action recognition tasks based on either semantic or kinematic features (Kalénine et al., 2010). Moreover, performance in goal recognition was affected by stimulation of both the frontal and the somatosensory cortex (Jacquet & Avenanti, 2015), while impairments in stroke patients did not seem to be associated with selective brain lesions (Kalénine et al., 2013), suggesting that the coding of action goals may depend on a more distributed network.

Our results are in line with evidence from other cognitive domains, like vision (Haxby et al., 2001), language (Huth et al., 2016), and semantics of object categories (Connolly et al., 2012; Huth et al., 2012), and suggest the existence of a unique, high-dimensional space even for action observation (Haxby et al., 2011).

### 3.4.2 The cortical representation of transitive actions

Despite overlapping responses, mapping of single voxel tuning still revealed incomplete segregation between movement types and object-related properties, reflecting the classical distinction between ventral and dorsal streams (Goodale & Milner, 1992). In particular, we found that inferior temporal areas, i.e., right *OTS* and bilateral *FusG*, retained primarily object-related information, consistent with previous literature indicating the involvement of ventral temporal areas in object processing and encoding of object categories (Haxby et al., 2001; Haxby et al., 2011; Grill-Spector & Weiner, 2014). On the other hand, dorsal occipital, precentral and postcentral regions (*PreCS, PostCS, PostCG*) exhibited higher sensitivity for kinematic-related features: these areas have been shown to process kinematic-based components of movement and trajectory (Grafton & Hamilton, 2007; Cavina-Pratesi et al., 2018; Turella et al., 2020) and are part of the dorsal pathway, which supports visually guided actions (Kravitz et al., 2011). The dissociation between ventral and dorsal pathways appeared to generalize to the second paradigm used, as representational geometries at the cluster level were similar between the first and second experiment: representations in clusters pertaining to the ventral stream organized along the animacy and object-identity dimensions, respectively, while representations in the dorsal stream organized primarily around kinematic-based features.

Growing evidence suggests object-related processing in the dorsal pathway, with caudal-medial portions responding to more visual and perceptual properties of the object and the rostral-lateral part representing action-based properties (Freud et al., 2016). In line with this evidence, we found object-related tuning in voxels pertaining to posterior parts of the dorsal stream (e.g., *TOS* and precuneus). Overall, the evidence suggests that a degree of functional segregation between the two streams exists, while their interconnectedness might subserve the integration of information processed by the two pathways to support more complex behaviors (Budisavljevic et al., 2018).

Interestingly, we also found a dissociation between left and right posterior temporo-occipital areas: voxels in left *MOG*, *pSTS* and *pMTG* were mostly tuned to animacy and category, while right *MOG* and *pSTS* showed higher discriminability for kinematic features. Posterior temporal regions are considered part of a third pathway, separate from ventral and dorsal streams (Pitcher & Ungerleider, 2021) and involved in the perception of biological motion (Peuskens et al., 2005; Zeki, 2015). Wurm and Carammazza (2022) proposed a conceptualization of this lateral pathway, particularly the lateral occipitotemporal cortex (LOTC), as an analogous of the ventral stream, subserving recognition of actions rather than objects. This third lateral pathway is characterized by a posterior-to-anterior gradient processing action-related representations from more concrete and perceptual features (Han et al., 2013) to more conceptual semantic aspects, mainly left-lateralized (Tucciarelli et al., 2019).

Other clusters responding to action observation were identified in bilateral *pIFS,* which showed high sensitivity to object-related properties, comparably to inferior temporal representations. The prefrontal cortex has been suggested to play an important role in spatial- and feature-based attention, in coordination with early visual and inferior temporal areas, to maintain the attentional template on the relevant location or object (Baldauf & Desimone, 2014). Thus, the involvement of *pIFS* may be explicitly ascribed to information coding in an object-based representational format (Bedini & Baldauf, 2021).

### 3.4.3 The specificity of the AON: intransitive gestures

In order to test the specificity of the AON areas, identified solely based on their response to features characterizing transitive actions, we assessed sensitivity to a different category of actions, i.e., intransitive gestures, which do not involve objects and subserve symbolic and/or communicative functions. Research on non-object-related actions described greater contributions of left-lateralized brain regions in the representation of these gestures (Króliczak & Frey, 2009; Kubiak & Króliczak, 2016), with particular involvement of posterior temporal areas (Caspers et al., 2010; Handjaras et al., 2015; Kubiak & Króliczak, 2016; Papeo et al., 2019). Our results showed that intransitive gestures, either symbolic or nonsense, were encoded across the AON. In the lateral pathway, we reported a dissociation between the left and right hemispheres for intransitive gestures: left *pSTS* and *pMTG*

primarily encoded symbolic gestures, consistently with reports of semantic processing of actions (Wurm & Caramazza, 2019; Tucciarelli et al., 2019), while right posterior temporal areas showed higher sensitivity to nonsense stimuli, probably reflecting salience of kinematic information. This additional evidence from our second experimental protocol further supports the hypothesis that the representation of actions in the left hemisphere is connected to language (Pulvermüller, 2005; Hodgson et al., 2022), whereas, in the right hemisphere, neural activity during action observation was more strongly associated with visual properties, akin to biological motion (Han et al., 2013; Sokolov et al., 2018).

### 3.4.4 Action representation in early visual areas

In both experiments, we highlighted the role of early visual cortices (i.e., cluster 8) in representing action features (Nishimoto et al., 2011). Specifically, voxels pertaining to the visual cortex were tuned solely to kinematics in the first experiment, while in the second experiment, they mapped kinematics, objects and even classes of intransitive gestures. This complete representation may be related to the naturalistic set of visual stimuli, where objects were always located in the lower part of the visual field, and hand trajectories were consistently present in the upper part. This bias was particularly evident in the second experiment, where stimuli were administered to participants as a continuous video, and objects remained at the

same location in space both when grasped and when touched, linking the specific identity of that object to its position in the visual field.

By testing the sensitivity of the AON, we demonstrated a general low specificity to action properties, with anatomically distant voxels sharing the great majority of their information content. Overall, results support the view that the AON relies on overlapping and distributed coding and may be considered a unique representational space instead of a system with a rigid modular organization based on the segregation of action features.

# Action-Perception Continuum: A novel approach for investigating the interplay between action, perception and decision-making

## 4. 1 Introduction

The ability to collect information via interaction with the environment and to make perceptual decisions based on the sampled information depends on the interplay between action, perception, and decision-formation processes. Theoretical and methodological approaches proposed to explore and explain this interaction have focused on three main components (cf. Lepora & Pezzulo, 2015; Hagura et al., 2017; Ozbagci et al., 2021):

a) *sensory encoding*: accumulation of sensory information about object features to be assessed and their encoding into decision-related evidence,
b) *decision expression*: choice selection and implementation based on sampled information,
c) *motor component*, underlying the physical interaction with the environment to collect information or execute a response, i.e., interaction movements.

Experimental approaches investigating perceptual decision-making have assigned different roles to each of these components according to their theoretical frameworks.

Experimental designs based on serial models of decision-making have usually neglected the role of the motor component. The serial framework is characterized by decide-then-act protocols (Gordon et al., 2021; Fig. 4.1A), where a decision is made, and the associated response is executed after deliberation (e.g., button press). On the other hand, parallel models (e.g., continuous flow; Coles et al., 1985) regard decision-making as an evolving process, continuously supplying the motor system with the accumulated evidence in order to prepare response execution. Consequently, continuous flow experimental designs entail the investigation of covert and overt components of action dynamics to track the evolution of the decision process (Fig. 4.1B), e.g., motor system preparation reflecting evidence accumulation (Selen et al., 2012) and changes of mind reflecting the revaluation of sampled information (Resulaj et al., 2009). More recent approaches, proposed in the context of the embodied choice framework (Lepora & Pezzulo, 2015), have taken into account also feedback processes from the motor component to the decision layer. Embodied experimental approaches allow evaluation and quantification of the role of action on perceptual and decision processes (Fig. 4.1C) by measuring, for example, the cost of the action involved in expressing a decision and commitment effects biasing the decision process (Marcos et al., 2013; Burk et al., 2014; Hagura et al., 2017; Thura & Cisek, 2014; Wiener et al., 2019).

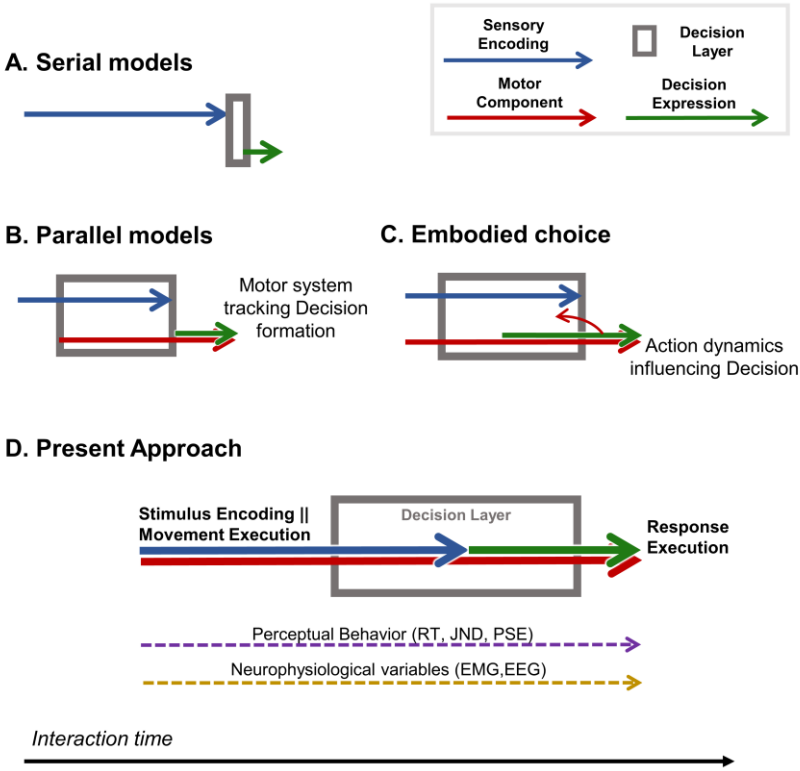



**Figure 4.1.** Previous and present approaches to the study of the interplay between action, perception, and decision-making. **A.** Serial models are characterized by a decide-then-act approach (e.g., key-press task), whereby the motor component and its influence on decision-making are neglected. **B.** Parallel models have the advantage of regarding decision-making as an evolving process, continuously feeding accumulated evidence to the motor system. By tracking changes in motor system states, it is possible to investigate the evolution of the decision process. **C.** Embodied choice approaches conceptualize perceptual decision-making as a perception-action loop, where action dynamics have a feedback influence on perceptual and decision layers. **D.** The present approach aims at investigating what

and to what extent motor behaviors contribute to perceptual decision-making. The novelty of the approach is to establish an action continuum along which all the stages of the decision process lie: interaction movements underlying stimulus exploration and those involved in decision expression are performed using the same end-effector, thus ensuring that response execution is a direct continuation of the stimulus encoding phase. Continuous recording of multiple variables throughout the interaction time allows the evaluation of perceptual and decision-making behavior at any time during the decision process and the investigation of neurophysiological correlates of perceptual decision-making.

Although suitable for advancing the understanding of perceptual decision-making, the previous approaches were not designed to explore the mechanisms linking spatiotemporal features of motor control to sensory encoding and decision formation. In particular, the main limitation of these protocols lies in the dissociation between interaction movements and sensory encoding or decision expression. Decoupling stimulus information sampling from motor behaviors involved in decision expression prevents the investigation of the interplay between sensory encoding and interaction movements and the extent to which motor patterns may correlate with or influence perceptual decision formation (Ozbagci et al., 2021). Conversely, decoupling interaction movements underlying stimulus information sampling from response execution limits the investigation of the evolution of pre- and post-decision behaviors. Importantly, previous approaches lack a continuous assessment of the motor component throughout the interaction time, which prevents testing correlative and causative hypotheses on the relation

between action, perception, and decision-making (Gordon et al., 2021).

In this chapter, we present a novel methodological and analytical framework to integrate and advance previous designs by addressing their limitations (Fig 4.1D). The rationale of this framework is to provide a comprehensive approach capable of a) describing the modulation of perceptual decision formation at any time of the interaction; b) determining the minimum amount of time and sampled information necessary to make a perceptual decision in interactive contexts; c) establishing correlations between patterns of motor behavior and the unfolding of perceptual decisions; d) testing hypotheses on the underlying neurophysiological correlates of perceptual decision-making mechanisms. The proposed approach was therefore designed to assess the spatiotemporal relationship between interaction movements, information sampling, and decision-related actions in different contexts of active perceptual decision-making.

# 4. 2 Methods

## 4.2.1 Methodological framework

The proposed experimental and analytical approach aims to advance the understanding of the role of motor behavior in active perceptual decision-making. Prior approaches were not designed to address *what* and *to what extent* motor behaviors contribute to perceptual decisions in interactive contexts. To fill this gap, we developed a novel methodological framework based on the combination of the following principles:

- definition of an a*ction continuum,* from early phases of stimulus interaction to response execution, along which all the decision phases lie, i.e., stimulus encoding, decision formation and expression.
- *interdependence* between interaction movements and stimulus encoding, i.e., stimulus features to be evaluated can be accessed exclusively through active movements.
- seamless and synchronous tracking of motor, psychophysical and neurophysiological aspects of the interaction underlying decision formation.

These principles allow a thorough and quantitative exploration of the mechanisms linking action, perception and decision from a motor, psychophysical and neurophysiological standpoint.

#### 4.2.1.1 Action continuum from movement initiation to response completion

A crucial characteristic of the present approach is establishing an action continuum whose beginning is defined by the motor components driving stimulus encoding and whose end consists of the action patterns expressing the decision, i.e., response execution and completion. Previous approaches did not consider this continuum in its entirety, for instance, by decoupling the action involved in the interaction from the stimulus features to be judged or decoupling interaction movements from the actions involved in response execution.

To address this limitation, we designed our approach, ensuring that pre- and post-decision motor patterns occurred along the same continuum: we asked participants to express their perceptual decisions using the same end-effector involved in the early phase of the interaction. Consequently, response execution is a direct continuation of the stimulus encoding phase (Fig. 4.1D). Notably, establishing an action continuum allows the decomposition of the entire decision-formation process into different interaction epochs. These epochs can be identified by detecting changes in motor behavior associated with the beginning of the interaction, the response initiation, and the completion of response execution (Fig. 4.3 and 4.2.3.1 in *Data Analysis*).

#### 4.2.1.2 Interdependence between sensory encoding and interaction movements

In the current approach, the task is designed so that the stimulus features to be evaluated can only be accessed through active interaction. The interdependent relation between interaction movements and stimulus encoding ensures a spatial and temporal coupling between the sensory inflow underlying perceptual decision formation and the unfolding of motor actions (i.e., blue, red and green arrows in Fig. 4.1D). An advantage of this coupling, which can be examined seamlessly throughout the entire interaction time, is the possibility to identify and assess salient changes in the motor and perceptual domains before, during and after a response has been initiated. Thus, the proposed framework can potentially uncover relationships between the perceptual and motor domains that would otherwise remain unobserved.

#### 4.2.1.3 Synchronous tracking of motor, psychophysical and neurophysiological variables

A third characteristic of our approach is the seamless and continuous recording of multiple variables describing participants' motor, perceptual and decision-making behavior emerging during the interaction, as well as their neurophysiological underpinnings. Motion tracking was included in our paradigm to continuously track the modulation of motor variables across all interaction phases. This aspect enables considering and correlating changes in action kinematics

and dynamics with psychophysical measures throughout the interaction time (Coles et al., 1985; Gratton et al., 1988; Lepora & Pezzulo, 2015; Wiener et al., 2019).

Unlike most previous approaches, our method allows the extraction of psychophysical variables, such as reaction times and accuracy, at any point in the interaction, rather than only when the decision has been expressed (Fig. 4.3 and 4.4). Reaction/response time (RT) is the time elapsed from stimulus presentation to response initiation/completion, whereas response accuracy is the rate of correct responses. The relation between these two variables has been shown to be highly effective in describing decision-making strategies (e.g., speed-accuracy tradeoff; Luce, 1986) and for implementing computational models of decision-making (e.g., the drift-diffusion model; Ratcliff, 1978).

Furthermore, our approach allows for the continuous recording of neurophysiological variables, consistent with previous work correlating EMG (Gratton et al., 1988; Selen et al., 2012) and EEG (O'Connell et al., 2012; Delis et al., 2018) with the modulation of psychophysical variables tracked across stimulus presentation, encoding and response execution phases.

### 4.2.2 Experimental approach

The experimental approach was designed to translate the working principles characterizing our framework into a task resembling everyday life activities in terms of motor and

perceptual demands. We designed a weight discrimination task in which participants had to compare the weight of two objects by lifting them and express their perceptual decision by deviating the lift on the horizontal plane; participants' arm/hand kinematics, grip force and muscular activity were recorded while performing the task.

#### 4.2.2.1 Participants

Sixteen healthy, right-handed participants (4 females; mean age ± SD: 23 ± 3.7 years) volunteered for this study. Each participant was compensated for the time spent in the laboratory ($10/h). All subjects had normal or corrected to normal vision. Subjects gave written informed consent to the procedures approved by the Office of Research Integrity and Assurance at Arizona State University, in conformity with the Declaration of Helsinki on the use of human subjects in research.

#### 4.2.2.2 Experimental set-up

The experimental set-up (Fig. 4.2A) consisted of a 3-D printed instrumented object, two haptic interfaces (Phantom, Geomagic), an inertial sensor, and a surface electromyography (EMG) system (9 Trigno wireless electrodes, Delsys). The instrumented object used to measure participants' grasp force consisted of two grasp surfaces with a force/torque sensor (Nano 25, ATI Industrial Automation, NC, USA) mounted at the object's center and connected to a PC (1 kHz sampling rate). An inertial sensor

(3-dof accelerometer, Delsys) was placed on top of the object base to track object acceleration, and the slot at the object base was used to place additional weights to increase object mass (Fig. 4.2A). Two 3-dof haptic devices (Phantom premium 1.5; Geomagic) were used to track (1 kHz sampling rate) hand position along the vertical (Z), horizontal (Y) and frontal (X) axis. Hand position was extracted by averaging the three-dimensional positions of the two endpoints of the device attached to the middle and proximal phalanx of the index finger and thumb, respectively. Time-varying endpoint positions were also used to render online visual feedback of the digits interacting with the object. A 3-D virtual object representing the actual object was also rendered and displayed on a monitor about 50 cm from the participant's eyes. A drape attached to the monitor between participants and the haptic interfaces prevented the view of the real object while allowing unhindered lifting movements. The virtual environment was rendered through a CHAI3d library included in a customized C++ code, which was used to control visual rendering during the different experimental stages and to store haptic device endpoint positions (500 Hz sampling rate). Surface electromyographic activity (EMG) from shoulder and arm muscles was recorded by a TRIGNO wireless system (Delsys, Boston). Communication between the base station and PC was established by analog output and then digitized by a data acquisition module (1 kHz sampling rate). We recorded nine shoulder and arm muscles: upper trapezius, middle deltoid, anterior deltoid, pectoralis major, biceps brachii, extensor carpi

radialis, flexor carpi radialis, opponens pollicis and first dorsal interosseous.

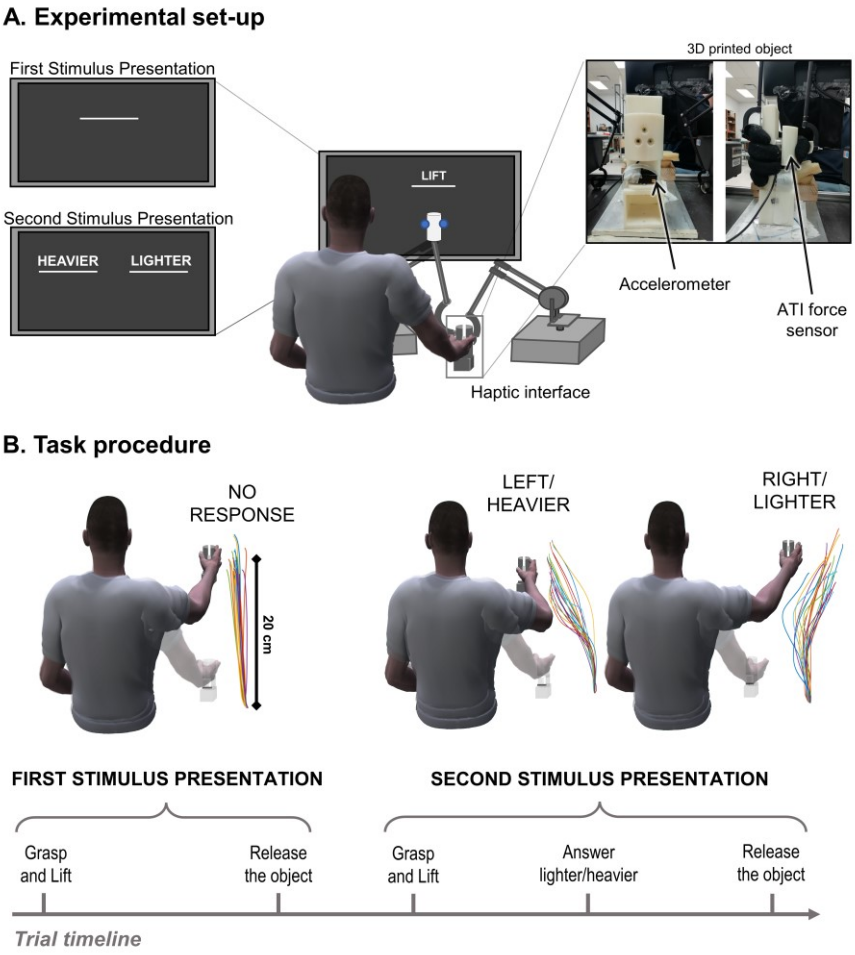



**Figure 4.2. A.** Experimental set-up. Participants sat in front of a monitor about 50cm from the participant's eyes, onto which a 3-D virtual representation of the object and participants' right index and thumb (as

spheres) was displayed. Two haptic interfaces (Phantom, Geomagic) were attached to the participant's index finger and thumb. One and two horizontal lines during the first and second lifts, respectively, were also displayed on the monitor to give participants visual cues on the lifting movement endpoint and whether they should respond (left). The 3-D printed object was concealed from participants' view by a drape attached to the base of the monitor; the object (right) was provided with an embedded force sensor (6-dof ATI sensor), and an accelerometer (3-dof, Delsys) was applied at its base. The object base was composed of a slot where weights could be applied to increase object mass. **B.** Task and procedure. During each trial, participants were asked to lift two objects of different weights sequentially. During the first stimulus presentation (left), participants were asked to lift the object vertically while estimating its weight. During the second stimulus presentation (right), participants were asked to judge whether the lifted object was heavier or lighter than the previous one. Participants expressed their perceptual decision by deviating their second lifting movement towards the right or the left.

#### 4.2.2.3 Task procedure

During each trial, two sequential lifting movements were performed on two objects of different weights (Fig. 4.2B). Participants started by placing their right hand in the resting position without touching the object. Then, instructions on the monitor informed the participants to grasp and lift the first object approximately 20 cm from the resting position while estimating and memorizing its weight (first stimulus presentation, Fig. 4.2B). Once the lifting movement was completed, participants were asked to drop the object and place their right hand in the resting position again. They were then cued to grasp and lift the second object (second stimulus

presentation, Fig. 4.2B). While lifting the second object, participants had to estimate its weight and judge whether the object being lifted was heavier or lighter than the first one. Participants expressed their perceptual decisions by deviating the trajectory of their hand horizontally towards the right or the left. Thus, on each trial, the hand's horizontal position (along the Y-axis) at the end of the second lifting movement determined the perceptual decision at the end of response execution, hereafter labeled as END. Participants were asked to drop the object without replacing it on the table at the end of each lifting movement; this was done to prevent subjects from experiencing the first object weight twice, i.e., during the lifting and replacing movements, whereas the second object's weight could only be experienced once before providing the perceptual response.

At the beginning of each lift, a 3-D virtual representation of the object was displayed along with two spheres representing online visual feedback of participants' right index and thumb (Fig. 4.2A). Visual feedback of the 3-D object and digits was extinguished when participants grasped the real object. Throughout the trial, one and two horizontal lines were displayed during the first and second lifts, respectively, to give participants visual cues on where the lifting movement should terminate (Fig. 4.2A, upper left inset). Participants were instructed to perform smooth lifting movements (single-peak vertical velocity profile) and to keep movement velocity consistent throughout a trial and not slower than 6.5 cm/s.

Movements too slow or uneven were classified as "invalid movement".

#### 4.2.2.4 Experimental design

The sequence and range of object weight presentation followed a constant stimuli method within a two alternative-forced choice (2AFC) design (Supplementary Table 3). On each trial, a standard and comparison object weight were presented either as the first or the second object to be lifted; the order of presentation was counterbalanced across trials. We used five comparison weights (CMP) ranging from 310 to 710 g (fixed interval of 100 g), and the standard weight (STD) was set to 510 g. Hereafter, we will refer to stimulus weights in terms of absolute or relative values (Δ Weight), i.e., the difference between each comparison weight minus the standard weight.

Before the experimental session, each participant performed 40 familiarization trials. During the first 20 trials, participants were asked to familiarize themselves with the motor component of the task: resting position, object grasping, lifting, and dropping phase. Participants were given 3D visual feedback of digit movements, and messages on the monitor informed them whether they performed an invalid movement (i.e., too slow or uneven). During the second set of 20 familiarization trials, participants were also asked to pay attention to the perceptual component of the task, hence providing a perceptual decision about object weight while performing the second lift. Participants were considered familiarized with the task - hence

ready to start the experimental session - when the percentage of correct responses was equal to or higher than 60% and the total number of invalid movements was equal to or less than 15. If these criteria were not met at the first familiarization session (40 trials), the participant was asked to complete additional 20 trials until all familiarization criteria were met.

The experimental session consisted of 50 lifts of each of the five standard/comparison combinations (250 trials) plus 48 confound trials (total of 298), comprising comparison weights for both the first and the second object lift. Confound trials were pseudo-randomly interleaved with the experimental trials and used to interfere with the subject's ability to infer upcoming standard/comparison combinations. The correspondence between movement direction and the response was counterbalanced across participants. Thus, half of the participants provided their answer 'heavier' and 'lighter' by deviating their lifting movements to the right and left, respectively, while for the other half, right and left deviations were associated with 'lighter' and 'heavier' answers, respectively. Trials where participants performed an invalid movement (see section 4.2.2.3) or did not execute an apparent response movement (e.g., the horizontal trajectory of the second lift did not deviate from the first) were excluded from further analysis.

### 4.2.3 Data-analysis and validation

This section outlines the set of analyses used to validate our experimental approach. This step also ensures the possibility of using our approach in other contexts of active perceptual decision-making beyond the one tested here. We first test the reliability of the kinematic analysis for detecting and classifying participants' decisions based on early components of response execution (section 4.2.3.1). Secondly, we demonstrate the ability of the proposed framework to quantify participants' perceptual performance at any time during the interaction rather than exclusively after response expression (section 4.2.3.2). Finally, we provide evidence of the ability of the proposed approach to link the evolution of motor patterns to decision formation (e.g., response time modulation) and perceptual performance (section 4.2.3.3). While EMG data were recorded throughout the experimental session, here we limited our analysis to kinematic and dynamic variables only.

#### 4.2.3.1 Response detection based on within-trial movement analysis

The initial step of our analysis aimed at estimating a response time approximating as closely as possible actual decision time, i.e., minimizing the delay between when a decision has been reached and when its overt expression is detected. We achieved this by devising a method to identify the earliest time perceptual decisions can be identified based on trial-by-trial analysis of the

participant's movements. Specifically, we defined a criterion to detect, within the interaction time, the earliest change in the kinematics of the lifting movement that represents the initiation of a response execution, i.e., a deviation either to the right or to the left from the vertical lifting motion (Fig. 4.2B). We found the modulation of horizontal hand velocity to be a reliable variable for detecting and classifying participants' responses (Fig. 4.3). Participants' horizontal and vertical hand velocity was obtained from the first derivative (2nd order Butterworth filter, 5 Hz cut-off) of the averaged endpoints' positions of the haptic devices attached to the participants' digits. For each trial, we extracted peak horizontal velocity on the first and second object lift ($v_{pk}$; green and red dots in Fig. 4.3, respectively).

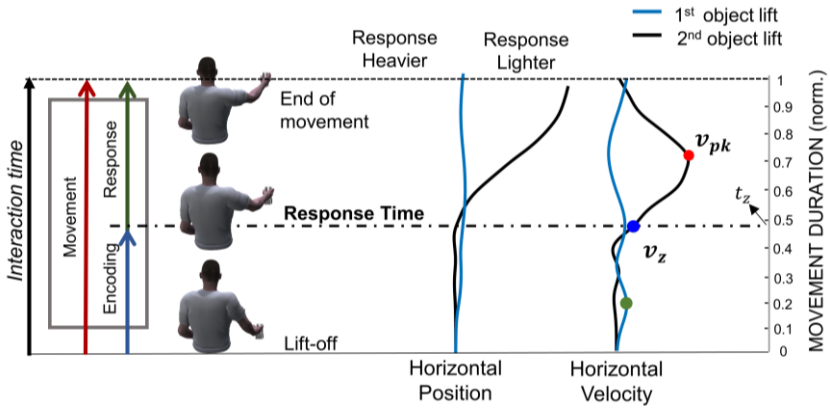



**Figure 4.3.** Detection of response time based on single trial kinematics. The earliest expression of response (i.e., response time) was defined as the time $t_z$ when the module of horizontal velocity exhibited during the second object lift (black profile) exceeds the peak of horizontal velocity shown during the first object lift (green dot on blue profile). Response

assignment (heavier/lighter) was then determined by considering the sign of horizontal velocity value at $t_z$, where positive values indicate a rightward response, negative otherwise. In the context of our framework (diagram on the left), $t_z$ sets the transition point within the interaction time when interaction movements underlying stimulus encoding start evolving into decision expression.

We then identified, on the second object lift, the first (absolute) value of hand horizontal velocity that exceeded the (absolute) value of peak horizontal velocity of the first lift and after which horizontal velocity continued to increase until reaching $v_{pk}$ ($v_z$; blue dot, Fig. 4.3). Consequently, the response time ($RT_Z$) was defined as the time $t_z$ at which $v_z$ occurred, corresponding to the earliest detection of the decision expression. Perceptual decisions were classified considering the sign of $v_z$ at $t_z$, positive and negative values indicating a rightward and leftward response, respectively (hence, heavier and lighter response, respectively, for half of the participants and vice-versa for the other half). Importantly, by comparing the second with the first object lift and choosing $v_z$ as the suprathreshold point closest in time to $v_{pk}$, we reduced misclassification errors that may occur due to the intrinsic variability of vertical lifting movements, i.e., unintentional deviations along the horizontal axis.

*Validation of RT detection method*

A fundamental step to ensure the reliability of the proposed detection and classification method was to exclude the possibility that the resulting RT distributions and associated

perceptual performance were misestimated. First, we addressed whether our method could correctly detect decision time during the early phase of response execution without overlooking crucial steps of decision formation. For this purpose, we compared the distribution of the median $RT_z$ as a function of Δ Weights with the median time to peak horizontal velocity during the second lift ($RT_{pk}$; Fig. 4.4A). $RT_{pk}$ is regarded as the reference reaction time distribution as both $RT_z$ and $RT_{pk}$ depend on the same decision strategies captured at different times of the same response execution kinematics, $v_z$ and $v_{pk}$, respectively. Thus, if our detection method based on $RT_z$ provided a reliable description of the decision-making process, modulation of $RT_z$ as a function of Δ Weights should be significantly correlated with $RT_{pk}$ medians modulations, with $RT_z$ distribution medians being characterized by shorter reaction time values.

To further validate our response classification method based on $v_z$, we compared perceptual responses, i.e., probability of responding comparison heavier than standard as a function of object weights, associated with $RT_z$, $RT_{pk}$ and end of response execution (END; Fig. 4.4A). For each participant, we obtained three psychometric curves by fitting a Probit function to probabilities of response based on hand position at END and the sign of horizontal velocity at $RT_z$ and $RT_{pk}$ (Fig. 4.4B). If the proposed response classification method can reliably assess participants' perceptual performance, the psychometric curve based on the $RT_z$ data should fit the END probabilities of judging the comparison heavier than the standard weight, i.e., p(CMP >

STD), not worse than the fit exhibited by the END psychometric curve. In this analysis, the END curve was considered the reference psychometric curve since it is obtained from the participants' final decision (i.e., end of overt response expression).

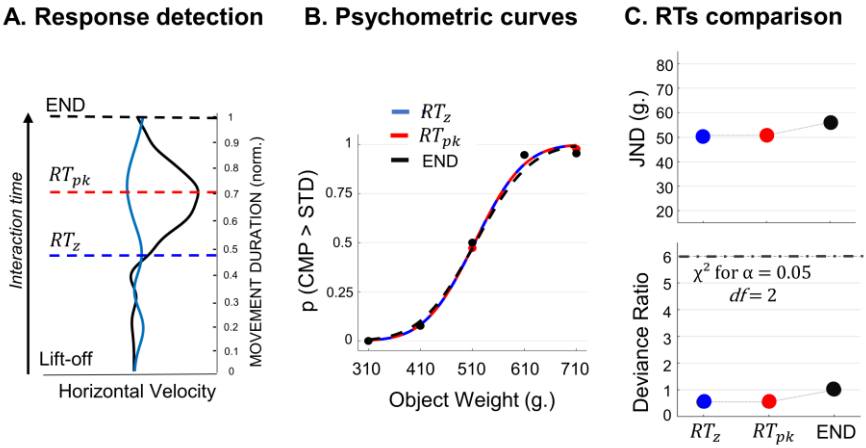



**Figure 4.4.** Validation analysis of the RT detection method. Data from a representative participant. **A.** Within each trial, three different time points were considered: the very early phase of response execution ($RT_z$), the middle of response execution ($RT_{pk}$) and response completion (END). **B.** Probability of responding comparison heavier than standard (CMP>STD) as a function of Δ Weight was calculated at each time point, based on the sign of horizontal velocity at $RT_z$ and $RT_{pk}$ and the sign of horizontal position at END. Three psychometric curves were fitted on these probabilities to describe the participant's perceptual decision-making. **C.** Reliability of the RT detection method to classify perceptual decisions was assessed by testing whether its associated psychometric curve fitted actual response probabilities not worse than the END curve (Deviance Ratio plot). JNDs resulting from the three psychometric curves were also tested for statistical differences (JND plot).

Quantification of these comparisons was obtained using the Deviance Ratio Test (Fig. 4.4C, lower plot; Toma & Santello, 2019), which considers the ratio between the fitting deviances resulting from the RT curve (either $RT_z$ or $RT_{pk}$) and the END curve:

$$\gamma = \frac{DEV_{RT}}{DEV_{END}},$$

where $\gamma$ is distributed as $\chi^2$ for 2 degrees of freedom, $DEV_{RT}$ is the deviance between RT psychometric curve (either $RT_z$ or $RT_{pk}$) and observed probabilities (i.e., probability of responding CMP > STD at each $\Delta$ Weight) based on hand position at END, and $DEV_{END}$ is the deviance obtained between the END psychometric curve and the same probabilities. Therefore, a $\gamma$ value lower than the critical value would indicate that the two curves were not statistically different in describing participants' perceptual behavior. Finally, we extended the comparison of the three curves by testing for statistical difference between their Just Noticeable Difference (JND) values (Fig. 4.4C, upper plot), calculated as the ratio between the critical value ($\vartheta_P$) and the slope ($\beta$) of the Probit function: $\vartheta_P$ / $\beta$, where $\vartheta_P$ = 0.645 for $p$ = 0.75 for the standardized normal probability distribution.

#### 4.2.3.2 Analysis of perceptual behavior as a function of time

An advantage of our approach is the possibility to explore dynamic aspects of perception, e.g., the modulation of perceptual behavior over time.

Fig. 4.5A shows how perceptual performance can be quantified at any given point within the interaction time. In this example, we assessed perceptual performance before, during, and after response initiation. Specifically, we extracted individual subject's probability of responding CMP > STD based on the sign of horizontal velocity at four time-points before $RT_z$, at $RT_z$, and at two time-points after response initiation (i.e., $RT_{pk}$ and END; Fig. 4.5A, left plot). Psychometric Probit functions were fitted on the resulting probabilities, thus obtaining seven curves describing the evolution of the perceptual decision throughout the interaction time (Fig. 4.5A, right plot). Response precision (JND) was calculated for each curve as $\vartheta_p$ / $\beta$, and the Deviance Ratio Test was used to quantify the goodness of fit of each curve to observed probabilities at END compared to the END curve. Single-subject curves and JND values were then averaged across subjects to assess overall participants' perceptual performance across interaction time.

The JND-Quantile (JQ; Fig. 4.5B) plot explores the modulation of perceptual decision-making over the range of decision times (Rank & Di Luca, 2015).

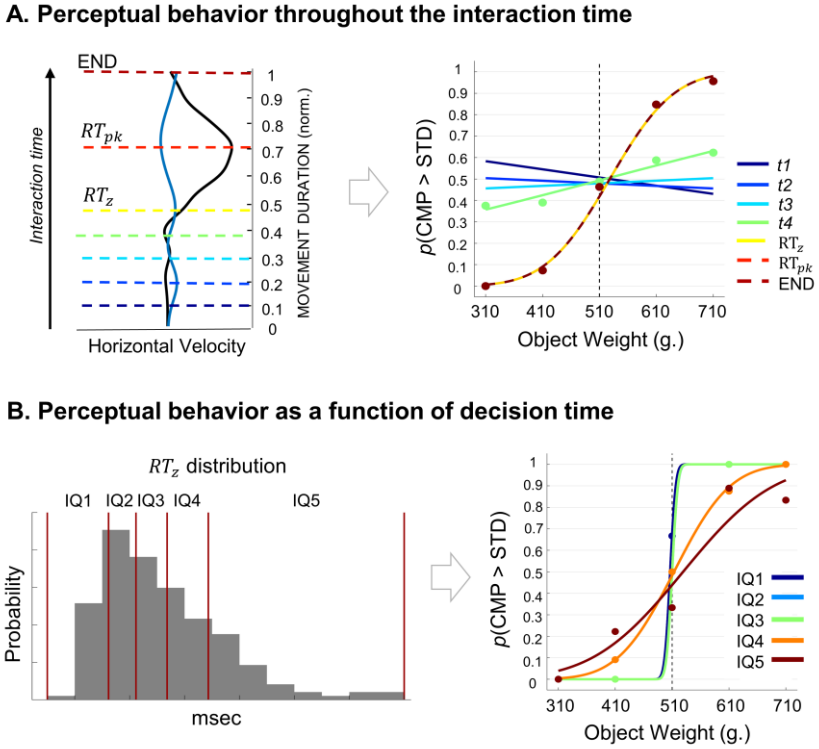



**Figure 4.5** Perceptual behavior as a function of time. Data from a representative participant. **A.** Perceptual behavior was evaluated throughout the interaction time by extracting probabilities of response at four time-points before decision time (*t1, t2, t3, t4*), at decision time ($RT_z$), and after ($RT_{pk}$ and END). Seven psychometric curves were fitted on the resulting probabilities describing the evolution of the perceptual decision. **B.** The $RT_z$ distribution (from all Δ Weights) was divided into five inter-quantile ranges (IQ), at 0%, 20%, 40%, 60%, 80% and 100% (vertical red lines) of the distribution. Heavier/lighter responses associated with the $RT_z$ within each IQ range were then used to calculate within-IQ p(CMP>STD). Five different psychometric curves (colored curves, left plot) were then fitted to each IQ observed probabilities of response.

For each participant, the whole $RT_z$ distribution was divided into five inter-quantile (IQ) intervals (Fig. 4.5B, left plot), i.e., 0% to 20% (IQ1), 20% to 40% (IQ2), 40% to 60% (IQ3), 60% to 80% (IQ4), and 80% to 100% (IQ5). Responses heavier/lighter associated with the $RT_z$ within each IQ range were then used to calculate p(CMP > STD), and a psychometric Probit function was fitted separately to the probabilities observed within each inter-quantile interval (Fig. 4.5B, right plot). Thus, each resulting curve captures participants' perceptual sensitivity in discriminating object weights across different phases of the interaction. At the group level, single-subject observed probabilities (i.e., p(CMP > STD)) were pooled across IQs defined independently for each participant. Psychometric curves were fitted on pooled data, resulting in five separate curves capturing the across-participant modulation of perceptual performance over time. The Deviance Ratio Test was used to quantify the goodness of fit of each IQ curve by comparing fitting curve deviance against the null fitting model (slope $\beta$=0); curves that were not significantly different from the null model were excluded for further analysis. JND and accuracy (PSE) values were extracted from each curve, where PSE was calculated as the ratio between the intercept ($\alpha$) and the slope of the Probit function: $-\alpha$ / $\beta$. JND and PSE confidence intervals were calculated using parametric bootstrapping (200 iterations).

#### 4.2.3.3 Receiver Operating Characteristic analysis

As mentioned, the present framework allows the in-depth investigation of the role of motor behaviors in perceptual decision-making. We pursued this objective by incorporating additional analytic steps to assess the contributions of kinematic, dynamic, and neurophysiological variables to perceptual decision formation. These analyses were based on a framework extensively used in sensorimotor control and perception research: the Receiver Operating Characteristic analysis (ROC; Green & Swets, 1966). We used ROC analysis to estimate how accurately an ideal observer could discriminate between across-trial modulation of a motor variable and to evaluate whether the timing and direction of this modulation resembled psychophysical performance (i.e., response times and perceptual decisions). As done in previous studies (Corneil et al., 2004; Pruszynski et al., 2008; Pruszynski et al., 2010), we performed ROC analysis separately for each participant and stimulus intensity (Δ Weight).

We first performed kinematic-based ROC analysis to quantify the extent to which the first and second object lift could be discriminated based on the difference in horizontal hand velocity profiles across all trials characterized by the same Δ Weight (Fig. 4.6A). Each velocity profile was considered from the time of object liftoff to the participant's median distribution of time-to-peak horizontal velocity. From each sample n (1 ms) of horizontal velocity profiles, we obtained the ROC curve and the associated Area Under the Curve (AUC) value, which is

equivalent to the probability of correctly discriminating between the first 'lift only' and the second 'lift plus response' movements. The resulting AUC time-series profile represented the ideal observer kinematic-based discrimination performance over time (Fig. 4.6A, bottom plot). AUC values of 0.5 indicate discrimination performance at a chance level (no difference between the first and the second lift velocity profiles). Conversely, AUC greater than 0.5 indicates classification of horizontal velocity exhibited during the second object lift as higher than the velocity exhibited during the first movement.

Due to the direct relation between changes in hand horizontal velocity and response execution, our method allows the interpretation of the AUC output as response detections based on cross-trial kinematic changes, i.e., deviations of hand velocity. Therefore, following the same procedure used in Pruszynski et al. (2010), we extracted from each AUC time series a discrimination time (DT) defined as the time when the AUC profile was < 0.33 or > 0.66 for at least ten consecutive samples, and indicating an above chance level discrimination of lift-only velocity profiles from lift+response movements (Fig. 4.6A). If multiple discrimination times were identified within the same time series, the last one was chosen as the actual DT. As a result, we obtained five discrimination times for each participant, one for each weight, which can be directly compared with $RT_z$ (Fig. 4.6B). Such comparison allows evaluating to what extent within-trial identification of participants' response time ($RT_z$) resembles

cross-trial (ROC-based) output, thus providing a measure of the robustness of the kinematic-based response time estimation.

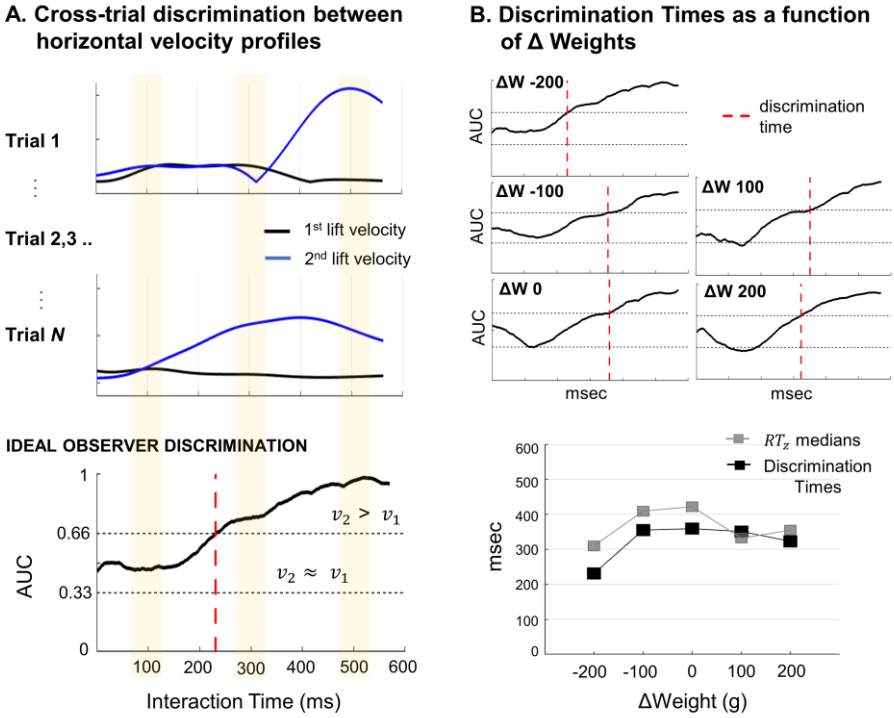



**Figure 4.6.** ROC analysis on horizontal hand velocity profiles. Data from a representative participant. **A.** The ideal observer discrimination analysis was performed across all trials with the same Δ Weight. A ROC curve and associated AUC (Area Under the Curve) value were obtained for each sample n (1 ms) of velocity profiles representing the probability of correctly discriminating between first and second lift, i.e., lift only and lift+response movements. AUC values of 0.5 indicate discrimination performance at the chance level; AUC equal to 1 indicates perfect classification performance. A threshold was set at >0.66 and <0.33 to identify above-chance level performance. **B.** From the

five AUC time-series (one per Δ Weight), five DTs were obtained and compared with $RT_z$ modulation as a function of Δ Weight. Discrimination Time (DT, vertical dashed red line) was defined as the time when the AUC profile was < 0.33 or > 0.66 for at least ten consecutive samples (10 ms).

The same analysis was performed on grip force profiles to quantify the accuracy of an ideal observer in discriminating between grip force exerted while lifting the standard and the comparison object across all trials characterized by the same Δ Weight (Fig. 4.7A). Moreover, we also assessed whether and to what extent the modulation of grip force correlates with participants' perceptual performance. The procedure to obtain AUC time series and DT values was the same as the one described above, except grip force profiles were grouped by type of stimulus presented, i.e., comparison versus standard object weight, independently of the order of presentation. AUC values of 0.5 indicate an ideal observer discrimination performance at chance level, indicating that grip force modulation during comparison weight lift ($GF_{cmp}$) is indistinguishable from grip force exerted during standard weight lift ($GF_{std}$). Conversely, AUC values greater (less) than 0.5 indicate an increase (decrease) in grip force while lifting the comparison weight relative to the standard (Fig. 4.7A).

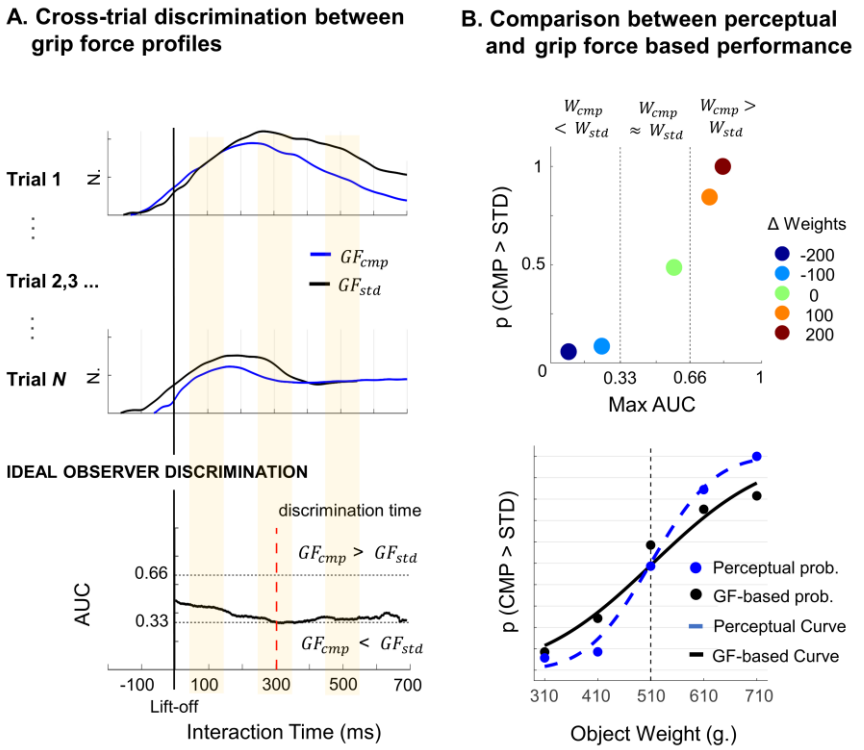



**Figure 4.7.** ROC analysis on grip force profiles. Data from a representative participant. **A.** The ideal observer discrimination analysis was performed across all trials with the same Δ Weight. A ROC curve and associated AUC (Area Under the Curve) value were obtained for each sample n (1 ms) of grip force profiles representing the probability of correctly discriminating between grip force exerted while lifting the standard and the comparison object weight. AUC values of 0.5 indicate discrimination performance at a chance level; AUC equal to 1 indicates perfect classification performance. A threshold was set at >0.66 and <0.33 to identify above-chance level performance. **B.** In the upper plot, the probability of perceiving the comparison weight heavier than the standard weight for each Δ Weight is plotted as a function of Max AUC, defined as the highest absolute value of the AUC time-series

profile after centering on 0.5. Vertical dashed lines represent the threshold of above-chance discrimination. AUC values < 0.33 are interpreted as 'lighter' response ($W_{cmp} < W_{std}$), whereas AUC values > 0.66 are interpreted as 'heavier' response ($W_{cmp} > W_{std}$). In the lower plot, the psychometric blue curve and probabilities obtained from p(CMP > STD), i.e., perceptual probabilities, are compared with grip force (GF) -derived curve and probabilities (black).

AUC values have been formally shown to correspond to the performance expected of an ideal observer in a 2AFC psychophysical paradigm (Green & Swets, 1966); they can thus be interpreted, in this instance, as the probability of classifying comparison objects as heavier than the standard relying on cross-trial grip force modulation. This approach allows translating AUC values in perceptual response terms and hence comparing grip force-based probabilities to the overt, $RT_z$-based perceptual decisions, i.e., p(CMP > STD) (Fig. 4.7B, upper plot). Grip force-based probabilities, here max AUC, were obtained by identifying the most extreme value of the AUC time series relative to the chance level probability of 0.5. Notably, this comparison permits the estimation of the extent to which grip force modulation can account for perceptual performance. We performed this quantification by fitting a psychometric curve on the AUC-based probabilities (Fig. 4.7B, lower plot) and then extracting the amount of perceptual variance ($R^2$) explained by the grip force-based curve. $R^2$ was computed as

$$1 - \frac{SSE}{SST},$$

where SSE is the sum of squared residual errors between $RT_z$-based observed probabilities and the probabilities predicted by the grip force-based curve, and SST is the sum of the squared difference between $RT_z$-based observed probabilities and their average. The Deviance Ratio value was also calculated as the ratio between the deviances of the grip force-based curve fitting actual observed p(CMP > STD) and the $RT_z$-based curve fitting the same probabilities. Medians and CIs of $R^2$ and Deviance Ratio were calculated through parametric bootstrap (1000 iterations). To identify the subgroup of participants whose grip force modulation could account for perceptual performance, we first excluded participants whose grip force-based curve was significantly worse than the perceptual curve in explaining perceptual responses, i.e., whose Deviance Ratio ± 95% CI was > 6 (critical $\chi^2$ for 2 df). Secondly, we selected only those participants with an $R^2$ ± 95% CI higher than 0.7, that is, whose grip force modulation explained more than 70% of the perceptual variance. Psychometric perceptual and grip force-based curves were averaged across this subset of subjects, and associated JND and PSE were tested for statistical differences.

## 4. 3 Results

All sixteen participants completed the experimental task with an average accuracy of 82.39%. The overall percentage of invalid trials that were excluded from further analyses (e.g., slow vertical lifting movements, no apparent response movement) was 16.40%; the number of invalid movements appeared to be modulated by weight discrimination difficulty (Supplementary Fig. S2A), as harder-to-discriminate stimulus pairs were associated with a higher number of invalid movements (13%, 17.63%, 22.38%, 14.64%, 14.37% excluded trials for -200, -100, 0, 100, 200 Δ Weight, respectively).

The percentage of fast response initiations (i.e., $RT_z$ < 100 ms) was also determined to estimate the number of possible prediction-driven decisions, guesses or other possible stimulus-unrelated responses. The 100 ms criterion was chosen based on previous works reporting ≈100 ms as the time necessary for the Central Nervous System to encode proprioceptive feedback and drive a related motor response (Pruszynski et al., 2011; Scott, 2012; Crevecoeur et al., 2016). Seven out of 16 participants did not exhibit any fast responses, and the overall percentage was less than 5% of all trials; the number of fast responses was modulated as a function of Δ Weights (Supplementary Fig. S2B) as well, as easier-to-discriminate stimulus pairs were associated with a higher number of these responses (4.31%, 3.04%, 2.25%, 1.47%, 2.63% for -200, -100, 0, 100, 200 Δ Weight, respectively).

This modulation and the above-chance level percentage of correct answers associated with fast responses suggest that they reflected prediction-driven decisions, that is, expectations about the second object weight based on the object weight perceived during the first lift.

Overall, these statistics show that participants engaged in an active perceptual task and modulated their motor behavior relative to stimulus characteristics, supporting the validity of the proposed protocol.

### 4.3.1 Validation of the response detection and classification method

To validate our early response detection method, we first compared the estimated $RT_z$ distribution with the reference reaction time $RT_{pk}$. Participants' $RT_z$ and $RT_{pk}$ medians ± 95% CI are shown in Fig. 4.8A as a function of Δ Weights. Critically, the modulation of the $RT_z$ distribution as a function of discrimination difficulty is significantly correlated with the modulation observed at $RT_{pk}$ (Pearson's r = 0.92, $p < 0.01$). This finding provides empirical evidence for the validity of the proposed response detection method in capturing decision-making strategies even at the early stages of decision expression.

Moreover, perceptual decisions evaluated at early response execution stages ($RT_z$) proved to be consistent with mid-response execution ($RT_{pk}$) and response completion (END) data (Fig. 4.8B).

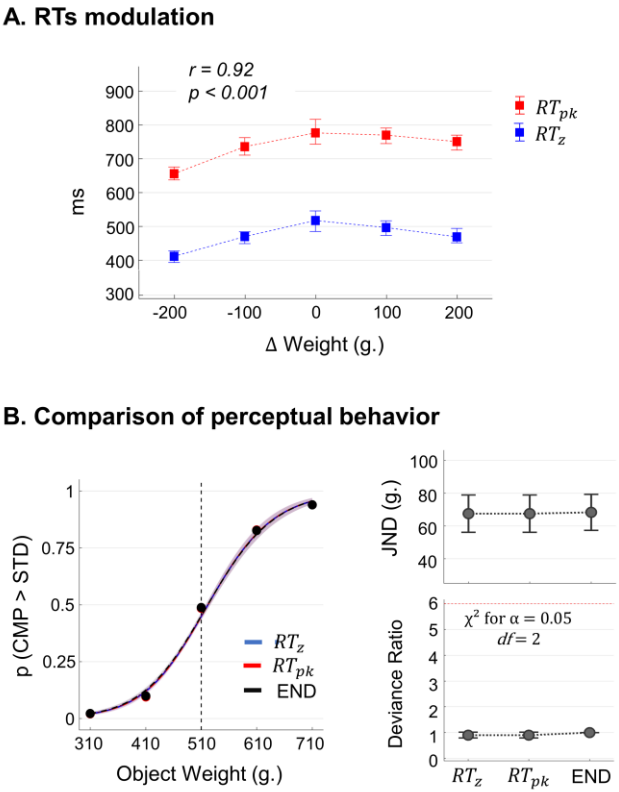



**Figure 4.8.** Validation of the response detection and classification method. **A.** Median $RT_z$ (blue squares) and $RT_{pk}$ (red squares) are plotted as a function of Δ Weights. The two distributions are significantly correlated, indicating that the response time modulation observed in the mid-response execution ($RT_{pk}$) can be reliably detected earlier in the interaction ($RT_z$). Error bars represent the 95% confidence intervals of the median. **B.** Three psychometric curves were fitted on subjects' probabilities of response associated with $RT_z$, $RT_{pk}$ and END and averaged across subjects. JND and Deviance Ratio values were obtained from single-subject psychometric curves, then averaged across subjects. JND (upper plot) values based on the early interaction dataset ($RT_z$) are not statistically different from the $RT_{pk}$ and the END

curves parameters. Deviance ratio (lower plot) values indicate that the difference between the three curves in explaining perceptual variance is not statistically different. Error bars represent the 95% confidence intervals of the mean.

Our early response classification method provides a description of perceptual sensitivity (JND) that is not statistically different from the JND values obtained from the reference curves $RT_{pk}$ and END (Repeated Measures ANOVA: $F_{(2,30)}$ = 1.65, $p$ = 0.21). Similarly, the Deviance Ratio associated with $RT_z$ was lower than the critical value (Fig. 4.8B), providing evidence that the $RT_z$ curve explained perceptual variance, i.e., hand position at the end of the movement, not worse than the reference END curve.

Overall, these results support the validity of the proposed kinematic-based early-response detection and classification method in reliably describing participants' perceptual decisions.

### 4.3.2 Exploration of active perception mechanisms throughout the interaction time

The present methodological framework enables the exploration of the dynamics of perceptual behavior from the beginning of stimulus encoding and active sampling to the end of the interaction. Perceptual performance can be quantified at any given time within the interaction time, thus allowing assessment of the evolution of the decision process (Fig. 4.9A). This analysis revealed that, on average, participants needed a minimum time

of 480 ± 208 ms (average $RT_z$) to build and start expressing a perceptual decision about object weight (Fig. 4.9A, left plot). The psychometric curves describing perceptual behavior before $RT_z$ explained perceptual decision-making statistically worse than the END reference curve, as evidenced by the associated Deviance Ratio values being higher than the critical value (Fig. 4.9A, lower right plot), resulting in high JND variability (Fig. 4.9A, upper right plot). This evidence suggests that perceptual decision-making before $RT_z$ is characterized by null or minimum discrimination sensitivity for object weight differences, supporting the identification of $RT_z$ as the earliest time, during the interaction, when perceptual behavior resembles actual decision-making.

To further explore the modulation of perceptual decision-making as a function of response time distribution, psychometric curves were fitted on data from different inter-quantile ranges within the $RT_z$ distribution (Fig. 4.9B), where IQ1 and IQ5 represent the fastest and slowest set of responses, respectively, and IQ3 includes responses centered around the distribution median. This analysis allows the exploration of the extent to which perceptual accuracy (PSE) and sensitivity (JND) depend on the time taken to reach a decision. A one-way ANOVA (IQ as factor, 5 levels) on bootstrapped JND values revealed an effect of response time (IQ) on response sensitivity ($F_{(4,995)} = 560.51$, $p < 0.01$), suggesting that slower responses were associated with worse performance. On the other hand, PSE values were

consistently higher than the standard object weight (510g), suggesting an overall tendency to underestimate object weights.

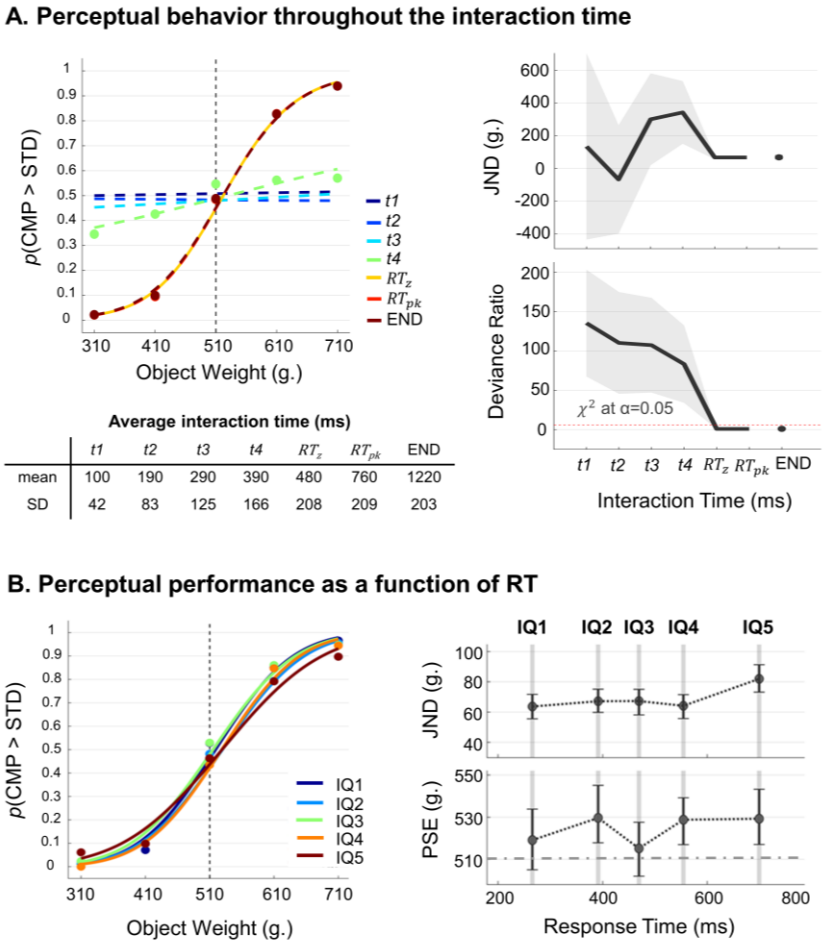



**Figure 4.9.** Perceptual behavior as a function of time. **A.** The psychometric curves describe perceptual decision-making before (blue, cyan and green curves), during (yellow curve) and after (red curves) response initiation. Each psychometric curve was fitted to each

subject's response probabilities and then averaged across participants. The legend reports the times, across participants average ± pooled standard deviation, at which response probabilities were sampled. The right-side plots represent JND and Deviance Ratio values associated with each psychometric curve; values were extracted from single-subject curves and then averaged across subjects. Shaded areas represent the 95% confidence intervals of the mean. **B.** The psychometric curves represent the overall modulation of participants' perceptual behavior observed within the response time distribution. A psychometric curve was fitted on responses pooled across subjects for each of the four inter-quantile intervals. JND and PSE values were extracted from each curve and plotted as a function of response time. Error bars represent 95% confidence intervals computed using a parametric bootstrap procedure (200 iterations).

### 4.3.3 Role of motor behaviors in perceptual decision-making

We incorporated ROC analysis into our approach to determine whether and to what extent kinematic and dynamic motor variables contribute to perceptual decision formation. Specifically, we tested across-trial discriminability of horizontal hand velocity and grip force profiles and quantified to what extent the modulation of ROC output was related to psychophysical performance exhibited over the interaction time.

#### 4.3.3.1 Correlation between motor variables and response time

ROC analysis on horizontal hand velocity resulted in five AUC time series representing the probability at each time point of an

ideal observer classifying the second lift velocity as higher than the first lift velocity (AUC > 0.5) or vice versa (AUC < 0.5). As participants were asked to express the perceptual decision only during the second object lift, the AUC profiles mainly show modulations towards 1. Fluctuations toward the 0 were sometimes observed at the beginning of AUC profiles (e.g., Fig. 4.6B), which probably reflected timing execution differences between the first and second movement: participants were more likely to perform first object lifts faster than second lifts, resulting in higher horizontal position variability at early time points in first object lifts compared to second object lifts.

From the AUC time series, five discrimination times for each subject were identified, representing the time when the hand horizontal velocity during the first and the second lifts starts differing. Discrimination times were compared with response times identified based on within-trial movement analysis (Fig. 4.10A). Correlation analysis revealed similar modulation of median discrimination times and $RT_z$ medians as a function of Δ Weights (Pearson's $r = 0.95$, $p < 0.01$). Moreover, 2x5 repeated measures ANOVA (Time Estimates and Δ Weight as within-subject factors) showed an effect of Δ Weight ($F_{(4,60)} = 14.54$, $p < 0.01$) and Time Estimates ($F_{(1,15)} = 19.08$, $p < 0.01$). Notably, there was no significant interaction between Δ Weights and Time Estimates ($F_{(4,60)} = 0.38$, $p = 0.82$), indicating that the modulation of $RT_z$ medians and discrimination times as a function of Δ Weight was not statistically different. In the context of our experimental validation, this result supports the robustness of our response

detection approach based on kinematic analysis. ROC analysis further demonstrates that the action involved in object lift, which underlies active stimulus sampling, can be distinguished from response expression based on kinematic changes during the interaction.

ROC analysis enables the investigation of the relationship between response time modulations and any other motor continuous variables which may play a role in decision formation. Accordingly, we used ROC analysis to assess the discriminability of grip force profiles exhibited while lifting standard and comparison objects. Four discrimination times were identified based on the AUC time series, representing the time when cross-trial grip force profiles exhibited during standard lift start differing from comparison object profiles. As expected, the AUC time series for Δ Weight 0 never passed the above-chance threshold since the same object weight (510 g) was presented in standard and comparison object lifts. Correlation analysis showed that grip force-derived discrimination times were modulated as a function of weights similarly to kinematic-based response time modulation (Fig. 4.10B; Pearson's correlation $r = 0.69$, $p < 0.01$). Moreover, grip force-derived discrimination times tended to anticipate kinematic-based response time estimates: one-tailed paired sample t-tests showed discrimination time to be statistically shorter than $RT_z$ medians for ±200 g Δ Weight ($t_{(13)} = -4.16$, $p$-corrected < 0.01, $t_{(15)} = -4.52$, $p$-corrected < 0.001) but not statistically different for ±100 g ($t_{(8)} = -0.21$, $p$ -corrected = 0.42, $t_{(9)} = -1.39$, $p$ -corrected = 0.10).

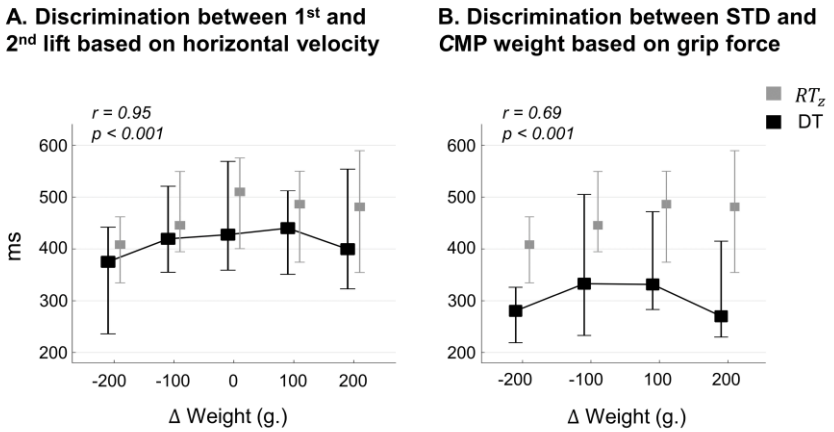



**Figure 4.10.** Comparison between ROC-derived Discrimination Times and $RT_z$. **A.** Discrimination times were obtained by performing ROC analysis on hand horizontal velocity exhibited during the first and second lifting movements. Across subjects medians of discrimination times and $RT_z$ are plotted in black and grey, respectively. Statistically significant correlation indicates comparable modulations of $RT_z$ and discrimination times medians as a function of Δ Weight. **B.** Discrimination times were obtained by performing ROC analysis on grip force exhibited while lifting comparison and standard object weight. Statistically significant correlation indicates comparable modulations of $RT_z$ and discrimination times medians as a function of Δ Weight. Δ Weight 0 has been omitted because its relative AUC values never passed the threshold for any subjects, indicating chance-level discrimination probability. Error bars represent the 95% confidence intervals of the median.

The observation of lower discrimination times indicates that grip force modulation occurs before (as for ±200 g) or in parallel (as for ±100 g) to kinematic response initiation, hence excluding the

possibility that grip force changes might be a by-product of the response-dependent change in hand velocity.

#### 4.3.3.2 Correlation between motor variables and perceptual performance

To quantify the extent to which grip force modulation can account for perceptual performance, we fitted psychometric curves on single-subject AUC-based probabilities derived from ROC analysis on grip force profiles. From grip force-based curves, we extracted $R^2$ values and estimated the Deviance Ratio to quantify how well grip force-based psychometric curves fitted actual response probabilities compared to the kinematic-based psychometric curve. These two measures were used to identify a subgroup of participants whose grip force modulation accounted for more than 70% of the perceptual variance ($R^2$ ±95% confidence intervals > 0.7) and whose grip force-based curve explained perceptual performance not worse than the kinematic-based perceptual curve. About 60% of participants (10 out of 16) satisfied both criteria (Fig. 4.11A).

The finding of a subgroup of participants whose grip force explained a significant amount of perceptual variance suggests that grip force can be one of the motor variables - among others - correlating with perceptual decision behavior during object weight estimation. To quantify the role of grip force in perceptual decision-making, we first averaged the grip force-based psychometric curves of those participants that fulfilled both the $R^2$ and Deviance Ratio criteria (n = 10; Fig. 4.11B).

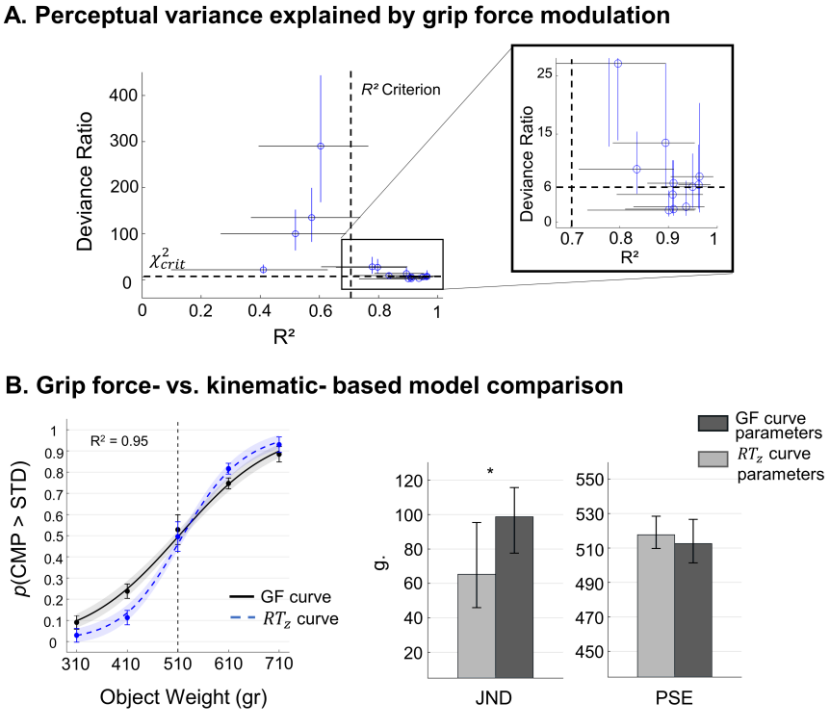



**Figure 4.11.** Contribution of grip force to perceptual behavior **A.** Single-subject $R^2$ values, quantifying perceptual variance explained by the grip force-based psychometric curve, are plotted against Deviance Ratio, which quantifies how well grip force-based psychometric curves fit actual response probabilities. Vertical and horizontal dashed lines represent the $R^2$ and Deviance Ratio-based threshold criteria, i.e., 0.7 and 6, respectively. These criteria were used to select participants whose grip force explained a significant amount of perceptual variance. Error bars represent the 95% confidence intervals of the medians estimated through parametric bootstrapping (1000 iterations). **B.** Grip force-based and kinematic-based psychometric curves were averaged across subjects who satisfied both the $R^2$ and the Deviance Ratio-based criteria. Bar plots depict median JND and PSE parameters associated with grip force-based (dark grey) and kinematic response-based (light

grey). Paired t-tests were performed on both parameters. * $p < 0.01$. Error bars and shaded areas represent the 95% confidence intervals.

The Deviance Ratio test comparing the averaged grip force-based curve with the null model confirmed that the grip force curve is statistically different from the null model in describing observed response probabilities (Deviance ratio = 60.05, $p < 0.001$). The averaged grip force curve explained approximately 95% of the perceptual variance ($RT_z$-based observed probabilities). However, the grip force-based curve described across-participant perceptual behavior worse than the averaged $RT_z$-based curve (Deviance ratio = 15, $p < 0.001$). Further analysis of curve parameters (Fig. 4.11B) revealed that the perceptual accuracy derived from grip force modulation was not statistically different from kinematic-based PSE ($t_{(9)} = 2.03$, $p = 0.07$). Conversely, we found a statistically significant difference between grip force- and $RT_z$-derived JND values, with the grip force model accounting for lower perceptual sensitivity ($t_{(9)} = -4.87$, $p < 0.01$; Fig. 4.11B).

Despite the variability across participants and the different degree of perceptual variance explained by the grip force-based model, these results demonstrate the potential afforded by our approach for exploring motor correlates of active perception using cross-trial ROC analysis.

## 4. 4 Discussion

In the present study, we designed and validated a novel experimental and analytical framework to investigate perceptual decision-making in active interaction contexts. The proposed approach is devised to overcome previous protocols limitations, primarily the lack of an action continuum allowing thorough investigation of the interplay between action, perception and decision-making. To test the validity of our approach, we collected data from a 2AFC weight discrimination task; the task was designed so that response execution was a direct continuation of sensory encoding, thus ensuring that pre- and post-decision interaction movements occurred along the same continuum. Motor, perceptual, and physiological variables were continuously recorded throughout the interaction to assess the extent to which motor components contribute to perceptual decisions.

First, we designed and implemented a method based on within-trial analysis of hand movement kinematics for detecting and classifying participants' decisions at early stages of response execution; the method proved to be reliable in estimating participants' perceptual behavior (see Fig. 4.8 and section 4.3.1). Continuous tracking of hand kinematics enabled us to assess perceptual performance throughout the interaction time (see Fig. 4.9 and section 4.3.2). By extracting hand velocity at different times during the interaction, we were able to differentiate

between pre- and post-decision motor behavior and establish that perceptual behavior before response time is characterized by low sensitivity for weight discrimination. Finally, we employed ROC analysis to establish whether and to what extent patterns of motor behavior are involved in perceptual decision formation, revealing a significant contribution of grip force to perceptual behavior (see section 4.3.3).

### 4.4.1 Early detection of perceptual responses

A relevant issue in investigating perceptual decision-making mechanisms through behavioral paradigms is the estimation of actual decision time (Lin & Tadin, 2021); since the only way to know a decision has been reached is to "wait" for its overt expression, decision time can only be approximated. Thus, the observed response time also depends on the time necessary to translate a decision into action (Gallivan et al., 2018); the earlier the detection of response initiation is, the more valid the assumption that response time approximates actual decision time.

To address this issue, participants' response times were identified based on within-trial changes in movement kinematics, namely horizontal hand velocity. Kinematic-based response times ($RT_z$) showed the same modulation as a function of discrimination difficulty as the reference reaction times based on peak velocity. Notably, perceptual performance assessed at $RT_z$ was not statistically different from that assessed at the end of

the lifting movements, suggesting that the proposed response detection method captured participants' decision-making at the very beginning of response execution. One of the advantages of our response detection method is that perceptual decision can be decoded before the completion of decision expression, i.e., arrival at the target position; this allows a better approximation of the actual decision time by reducing the time necessary to complete the response movement once the decision is made. Moreover, velocity has the advantage of signaling instant changes in movement directionality independently of actual position or distance from target positions. Indeed, our velocity-based investigation of perceptual behavior across interaction time (see sections 4.2.3.1 and 4.3.2) revealed a significant difference between pre-$RT_z$ psychometric curves and the reference END curve in describing perceptual behavior; this evidence shows that $RT_z$ is the earliest time when perceptual behavior resembles the performance exhibited at response completion.

Together, these findings support the validity of our response detection and classification method in identifying decision expression at very early stages of response execution, hence providing a more precise estimate of the actual decision time.

### 4.4.2 Evolution of decision-making through time

A relevant novelty of the proposed approach is the opportunity to investigate the mechanisms driving the modulation of

perceptual behavior as a function of time. Continuous tracking of the kinematic, dynamic, and neurophysiological behavior allows explorative analysis to provide insights into the link between response times, perceptual performance modulation and motor patterns characterizing the interaction. We implemented two types of analysis to explore this relation, assessing the evolution of perceptual behavior throughout the interaction time and the modulation of perceptual performance as a function of response time.

Within-interaction analysis of perceptual behavior entailed evaluating participants' performance at multiple time points throughout the interaction. Drawing from parallel approaches to decision-making, this methodology comprises the continued assessment of action dynamics to track the underlying decision-formation process (Lepora & Pezzulo, 2015; Gordon et al., 2021). Importantly, it also allows differentiating, within the action continuum, between pre-decision (i.e., active sensing) and post-decision (i.e., response execution) motor behaviors. Indeed, we found a clear dissociation in performance before and after response time, with null or minimum discrimination sensitivity for weight differences before $RT_z$. Modulation of psychometric curves and associated JND values as a function of time provided a description of the evolution of perceptual behavior, revealing the time and the amount of sampled information necessary to build a perceptual decision (Plaisier et al., 2019). Before the sampled information threshold, weight discrimination performance was low but increased with accumulated sensory

evidence until passing the sampled information threshold and reaching a plateau.

Modulation of perceptual performance as a function of response time was examined using JND-Quantile plots (Rank & Di Luca, 2015), which capture perceptual sensitivity at different delays across the distribution of response times. The JQ plot illustrates the pattern of speed-accuracy tradeoffs (Luce, 1986), i.e., the relation between responses accuracy and the time taken to generate them. Speed-accuracy tradeoffs, namely faster error responses than correct ones, are generally observed in time-constrained tasks, while slower error responses are observed when task instructions emphasize accuracy (Swensson, 1972; Luce, 1986). Our results showed that JND values increased with longer $RT_z$, indicating that slower response times were generally associated with a higher error rate. This phenomenon has been explained within the framework of drift-diffusion models by implementing urgency signals, which decrease the amount of evidence needed to trigger a decision as the end of the response time window approaches (Ratcliff et al., 2016). Urgency signals amplify the accumulated evidence as time passes, lowering the sampled information threshold and producing slow error responses (Ditterich, 2006). In the context of our task, urgency signals would become more influential as participants approach the end of the vertical movement without having expressed a decision, thus leading to a higher number of slow erroneous responses.

### 4.4.3 Linking motor control to perceptual decision-making

ROC analysis was incorporated into our framework to enable precise quantifications of what and to what extent features of motor behavior play a role in the formation of perceptual decisions formation. ROC technique has been successfully used to provide insights into how sensory and motor signals are integrated for motor control (Pruszynski et al., 2008; 2010; Crevecoeur et al., 2012; 2016). Specifically, our goal was to quantify the extent to which timing and direction of the modulation of a given motor variable resemble the modulation of psychophysical performance exhibited over the interaction time, i.e., response times and perceptual performance.

ROC analysis of hand velocity profiles revealed across trials above-chance discriminability between the first and the second lifting movement. Notably, discrimination times, i.e., the time at which first and second lift velocities start differing according to the ideal observer performance, closely approximated actual response times, further supporting the validity of our response detection method. Similarly, cross-trial analysis of grip force discriminability reached above-chance values and grip force-derived discrimination times were significantly correlated with response times. Moreover, the psychometric curve fitted on grip force discrimination probabilities explained a large amount of perceptual performance variance in ~60% of participants. This result suggests that, at least in a subgroup of participants, grip

force might contribute to the final decision, consistent with previous works relating grip force and weight perception (van Polanen & Davare, 2015; Ellis & Lederman, 1999; Flanagan & Bandomir, 2000).

Overall, these findings highlight the potential of employing ROC analysis to identify and quantify correlations between modulations of motor variables and the evolution of decision formation and decision expression as a function of stimulus features.

### 4.4.4 Advantages of the proposed approach to the exploration of active perception mechanisms

By integrating and advancing previous approaches to the study of perceptual decision-making, the proposed approach offers several advantages to investigating the interplay between action, perception, and decision processes.

Firstly, the proposed response detection and classification method provides a means to more closely approximate decision time, identifying response movements at the earliest stages of their execution rather than relying on movement completion. Additionally, early response detection facilitates reliably differentiating within the action continuum between the sensory-motor signals driving the decision (i.e., pre-response, sampling movements) and the motor patterns underlying decision expression (i.e., post-decision, response execution). The capability of operating this distinction, coupled with the

inclusion of an active sampling component in the experimental paradigm, permits assessing the influences from action dynamics to the decision vs. sensory encoding processes (Hagura et al., 2017; Ozbagci et al., 2021). Therefore, it allows for a comprehensive analysis of the interactions between the decision layer, the motor components underlying stimulus encoding and the motor components involved in response execution.

The implementation of an action continuum from stimulus encoding to response execution and completion grants the possibility of extracting psychophysical, motor and physiological variables at any point in time during the interaction. Consequently, our approach allows the exploration of often-overlooked dynamic aspects of perception (Rank & Di Luca, 2014), such as the time required to build up a specific percept (Plaisier et al., 2019) and the modulation of perceptual performance over time (Rank & Di Luca, 2014; 2015).

### 4.4.5 Possible applications and future directions

Based on the results of the present study, we believe our analytical approach is suited for implementation in diverse contexts of active perceptual decision-making and could be expanded beyond the range of analyses described. In particular, the use of ROC analysis is not limited to the motor variables tested here but can be performed on any kinematic, dynamic, and neurophysiological variable continuously recorded

throughout the interaction, allowing precise temporal estimation of the influence of multiple motor components on decision. Indeed, previous works have used EMG (Selen et al., 2012; Servant et al., 2015) and EEG (for a review, see O'Connell & Kelly, 2021) to study the neurophysiological correlates of decision variables, identifying signals reflecting the evolution of perceptual decisions (Hauser & Salinas, 2014). In the present study, we limited our investigation to kinematic and dynamic variables to assess the role of motor behavior in perceptual performance; however, ROC analysis of physiological measures can provide a window into the underlying neurophysiological correlates of perceptual decision-making.

Moreover, we suggest that our methodological approach is well suited for investigating decision-making within the framework of the Embodied Choice model (Lepora & Pezzulo, 2015), incorporating the notions of changes of mind, motor cost and commitment. Recent studies suggest that changes of mind not only reflect purely deliberative processes but also depend on the state of the motor system, and the probability of their occurrence is influenced by the degree of commitment and the motor cost associated with changing decisions (Burk et al., 2014; Cos et al., 2021). However, it is still unclear what and how much sensory evidence is needed to trigger changes of mind and under which conditions they are more likely to occur (Gordon et al., 2021). Another open question in the Embodied Choice framework concerns at what stage of the decision process commitment effects emerge (Gordon et al., 2021) and whether they reflect

cognitive aspects of the choice or variability in the state of the motor system (Lepora & Pezzulo, 2015; Cos et al., 2021). These issues could be addressed in the context of our approach, which allows assessment of the contribution of different motor and neurophysiological variables to changes of mind and modeling commitment effects throughout the interaction, from object contact to response execution stages.

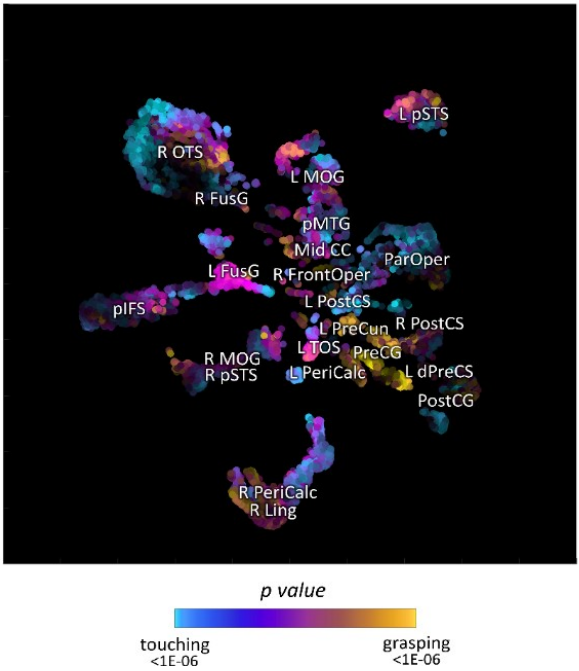
The proposed approach also offers the possibility of clarifying the role of predictions and sensorimotor memory, i.e., how previous experience is combined with sensory evidence to influence decision formation and motor planning (Johansson & Westling, 1984; Summerfield & de Lange, 2014). As discussed in the Results section, response times lower than 100 ms were interpreted as predictions, meaning that decision was possibly driven by expectations about the object to be lifted based on the object weight detected in the previous lift. Grip force exerted between object contact and liftoff has been shown to reflect expectations about the object weight, scaling according to the predicted weight based on memory about the previously applied force (van Polanen & Davare, 2015; 2019). While we focused our analyses on the lifting phase, analysis of fingertip forces and grip force rate variations before liftoff could disentangle the effect of sensorimotor memory from active sensing mechanisms driving decision, quantifying the influence of predictions on motor and perceptual performance.

In this chapter, we presented a novel experimental and analytical framework aimed at assessing the spatiotemporal relationship between action, perception, and decision-making. The results of our validation and analytical procedure suggest that the proposed approach is capable of describing the modulation of perceptual decision formation through time and determining the time and amount of information necessary to reach a decision in interactive contexts. Moreover, it allows the assessment of correlations between patterns of motor behavior and decision processes and the exploration of the neurophysiological correlates underlying decision-making mechanisms. Overall, the evidence presented in the present study supports the validity of the proposed methodological approach for investigating perceptual decision-making in different contexts of active interaction.

# Appendix

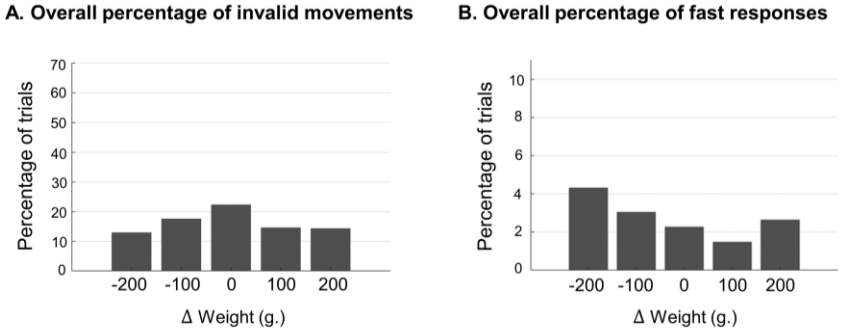
## Supplementary Figures

### Supplementary figure S1. 2D representation of voxel tuning to grasping and touching dimensions




Dimensions describing the second set of stimuli were mapped using the *p*-value associated with *d'* onto the space defined by the t-SNE. The relative contribution of action features from grasping to touching was mapped with a color scale ranging from yellow to magenta to cyan; the maximal saturation of each channel reflects an uncorrected $p < 1E\text{-}06$.

## Supplementary figure S2. Percentages of invalid movements and fast responses as a function of Δ Weight




**A.** Percentage of invalid movements is computed separately for each Δ Weight by pooling data from all participants (N=16). Invalid movements were defined as slow vertical lifting movements, multiple-peaks vertical velocity profiles, absence of a clear response movement, or trials in which either the experimenter's errors or hardware failures resulted in invalid datasets. Percentages are modulated as a function of Δ Weight, reflecting task difficulty: harder-to-discriminate stimulus pairs were associated with more invalid movements. **B.** Percentage of fast responses is computed separately for each Δ Weight by pooling all participants' data (N=16). Fast responses were defined as trials in which $RT_z$ < 100 ms; these responses are likely to reflect stimulus-unrelated decisions, e.g., predictions or guesses, since the response time was too short to allow for stimulus encoding, decision formation, and transmission of the decision computation to the motor system. Percentages are modulated as a function of Δ Weight, as easier-to-discriminate stimulus pairs were associated with a higher number of these responses.

# Supplementary Tables

## Supplementary Table 1.

**Table 1.** *Experimental stimuli in Experiment 1*

| *Animate target* | | *Inanimate target* | |
|---|---|---|---|
| Animals | Body parts | Natural | Artificial |
| 1. Chick | 1. Elbow | 1. Apple | 1. Cup |
| 2. Fish | 2. Foot | 2. Banana | 2. Drill |
| 3. Lobster | 3. Hand | 3. Eggplant | 3. Mouse |
| 4. Rabbit | | 4. Fennel | 4. Phone |
| 5. Shrimp | | 5. Onion | 5. Scissors |
| 6. Snail | | | |
| 7. Turtle | | | |

Participants were presented with movie clips depicting a hand-performed transitive action. The target objects of the actions differed in the animacy dimension (ten animate and ten inanimate exemplars) and their category (i.e., animals, body parts, natural objects, such as vegetables and fruits, artificial objects).

## Supplementary Table 2.

**Table 2.** *Experimental stimuli in Experiment 2*

| *Transitive* | | *Symbolic* | |
|---|---|---|---|
| 1. Cup | 10. Eyeglasses | 1. Ok | 10. Not-okay |
| 2. Penholder | 11. Keys | 2. Waving | 11. Come |
| 3. Scissors | 12. Wristwatch | 3. More-or-less | 12. Stop |
| 4. Paperknife | 13. Glass | 4. No | 13. Horns |
| 5. Plate | 14. Bottle | 5. Blessing | 14. Hello |
| 6. Ducktape | 15. Pen | 6. Victory sign | 15. Cut |
| 7. Highlighter | 16. Post-it | 7. Hitchhiking | 16. Tighten |
| 8. Eraser | 17. Knife | 8. Later | 17. Fine |
| 9. Stapler | 18. Spoon | 9. Money | 18. What |

*Note.* Transitive actions included grasping and touching performed on the same target object

Participants were presented with a continuous video depicting an actor performing a movement with the upper limb. The actor performed 18 grasping and 18 touching movements directed to 18 target objects, and 18 symbolic and 18 nonsense gestures.

## Supplementary Table 3.

**Table 3.** *Experimental stimuli and procedure*

| *Standard* | *Comparison* | *Δ Weight* |
|---|---|---|
| | 310 g | -200 g |
| | 410 g | -100 g |
| 510 g | 510 g | 0 g |
| | 610 g | 100 g |
| | 710 g | 200 g |

On each trial, a standard and comparison object weight were presented either as the first or the second object to be lifted; the order of presentation was counterbalanced across trials. We used five comparison weights (CMP) ranging from 310 to 710 g (fixed interval of 100 g), and the standard weight (STD) was set to 510 g. Hereafter, we will refer to stimulus weights in terms of absolute or relative values (Δ Weight), i.e., the difference between each comparison weight minus the standard weight.

# References


Abdollahi, R. O., Jastorff, J., & Orban, G. A. (2013). Common and segregated processing of observed actions in human SPL. *Cerebral cortex*, *23*(11), 2734-2753.

Aflalo, T., Zhang, C. Y., Rosario, E. R., Pouratian, N., Orban, G. A., & Andersen, R. A. (2020). A shared neural substrate for action verbs and observed actions in human posterior parietal cortex. *Science Advances*, *6(*43), eabb3984.

Andric, M., Solodkin, A., Buccino, G., Goldin-Meadow, S., Rizzolatti, G., & Small, S. L. (2013). Brain function overlaps when people observe emblems, speech, and grasping. *Neuropsychologia*, *51*(8), 1619-1629.

Avants, B. B., Tustison, N., & Song, G. (2009). Advanced normalisation tools (ANTS). *Insight j*, *2*(365), 1-35.

Badre, D., & D'esposito, M. (2009). Is the rostro-caudal axis of the frontal lobe hierarchical?. *Nature Reviews Neuroscience*, *10*(9), 659-669.

Baldauf, D., & Desimone, R. (2014). Neural mechanisms of object-based attention. *Science*, *344*(6182), 424-427.

Bedini, M., & Baldauf, D. (2021). Structure, function and connectivity fingerprints of the frontal eye field versus the inferior frontal junction: A comprehensive comparison. *European Journal of Neuroscience*, *54*(4), 5462-5506.

Benjamini, Y., & Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Annals of statistics*, 1165-1188.

Bischoff, M., Zentgraf, K., Pilgramm, S., Stark, R., Krüger, B., & Munzert, J. (2014). Anticipating action effects recruits audiovisual

movement representations in the ventral premotor cortex. *Brain and cognition, 92*, 39-47.

Bozzacchi, C., Giusti, M. A., Pitzalis, S., Spinelli, D., & Di Russo, F. (2012). Awareness affects motor planning for goal-oriented actions. *Biological psychology, 89*(2), 503-514.

Buccino, G., Binkofski, F., Fink, G. R., Fadiga, L., Fogassi, L., Gallese, V., ... & Freund, H. J. (2001). Action observation activates premotor and parietal areas in a somatotopic manner: an fMRI study. *European journal of neuroscience*, *13*(2), 400-404.

Budisavljevic, S., Dell'Acqua, F., & Castiello, U. (2018). Cross-talk connections underlying dorsal and ventral stream integration during hand actions. *Cortex, 103*, 224-239.

Burk, D., Ingram, J. N., Franklin, D. W., Shadlen, M. N., & Wolpert, D. M. (2014). Motor effort alters changes of mind in sensorimotor decision making. *PloS one, 9*(3), e92681.

Caramazza, A., Anzellotti, S., Strnad, L., & Lingnau, A. (2014). Embodied cognition and mirror neurons: a critical assessment. *Annual review of neuroscience, 37*, 1-15.

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.

Cavina-Pratesi, C., Connolly, J. D., Monaco, S., Figley, T. D., Milner, A. D., Schenk, T., & Culham, J. C. (2018). Human neuroimaging reveals the subcomponents of grasping, reaching and pointing actions. *Cortex*, *98*, 128-148.

Chen, G., Saad, Z. S., Nath, A. R., Beauchamp, M. S., & Cox, R. W. (2012). FMRI group analysis combining effect estimates and their variances. *Neuroimage*, *60*(1), 747-765.

Cisek, P., & Kalaska, J. F. (2005). Neural correlates of reaching decisions in dorsal premotor cortex: specification of multiple direction choices and final selection of action. *Neuron, 45*(5), 801-814.

Coles, M. G., Gratton, G., Bashore, T. R., Eriksen, C. W., & Donchin, E. (1985). A psychophysiological investigation of the continuous flow model of human information processing. *Journal of Experimental Psychology: Human Perception and Performance, 11*(5), 529.

Connolly, A. C., Guntupalli, J. S., Gors, J., Hanke, M., Halchenko, Y. O., Wu, Y. C., ... & Haxby, J. V. (2012). The representation of biological classes in the human brain. *Journal of Neuroscience*, *32*(8), 2608-2618.

Corneil, B. D., Olivier, E., & Munoz, D. P. (2004). Visual responses on neck muscles reveal selective gating that prevents express saccades. *Neuron, 42*(5), 831-841.

Cos, I., Pezzulo, G., & Cisek, P. (2021). Changes of Mind after movement onset depend on the state of the motor system. *Eneuro, 8*(6).

Cox, R. W. (1996). AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. *Computers and Biomedical research, 29*(3), 162-173.

Creem-Regehr, S. H., & Kunz, B. R. (2010). Perception and action. *Wiley Interdisciplinary Reviews: Cognitive Science, 1*(6), 800-810.

Crevecoeur, F., Kurtzer, I., & Scott, S. H. (2012). Fast corrective responses are evoked by perturbations approaching the natural variability of posture and movement tasks. *Journal of neurophysiology, 107*(10), 2821-2832.

Crevecoeur, F., Thonnard, J. L., Lefèvre, P., & Scott, S. H. (2016). Long-latency feedback coordinates upper-limb and hand muscles during object manipulation tasks. *Eneuro, 3*(1).

Culham, J. C., & Valyear, K. F. (2006). Human parietal cortex in action. *Current opinion in neurobiology, 16*(2), 205-212.

Decroix, J., & Kalénine, S. (2019). What first drives visual attention during the recognition of object-directed actions? The role of kinematics and goal information. *Attention, Perception, & Psychophysics, 81*, 2400-2409.

Delis, I., Dmochowski, J. P., Sajda, P., & Wang, Q. (2018). Correlation of neural activity with behavioral kinematics reveals distinct sensory encoding and evidence accumulation processes during active tactile sensing. *Neuroimage, 175*, 12-21.

Ditterich, J. (2006). Stochastic models of decisions about motion direction: behavior and physiology. *Neural networks, 19*(8), 981-1012.

Donnarumma, F., Costantini, M., Ambrosini, E., Friston, K., & Pezzulo, G. (2017). Action perception as hypothesis testing. *Cortex, 89*, 45-60.

Ellis, R. R., & Lederman, S. J. (1999). The material-weight illusion revisited. *Perception & psychophysics, 61,* 1564-1576.

Fabbri, S., Strnad, L., Caramazza, A., & Lingnau, A. (2014). Overlapping representations for grip type and reach direction. *Neuroimage, 94*, 138-146.

Filimon, F., Nelson, J. D., Hagler, D. J., & Sereno, M. I. (2007). Human cortical representations for reaching: mirror neurons for execution, observation, and imagery. *Neuroimage, 37*(4), 1315-1328.

Filimon, F. (2010). Human cortical control of hand movements: parietofrontal networks for reaching, grasping, and pointing. *The Neuroscientist, 16*(4), 388-407.

Filimon, F., Rieth, C. A., Sereno, M. I., & Cottrell, G. W. (2015). Observed, executed, and imagined action representations can be decoded from ventral and dorsal areas. *Cerebral Cortex, 25*(9), 3144-3158.

Flanagan, J. R., & Bandomir, C. A. (2000). Coming to grips with weight perception: effects of grasp configuration on perceived heaviness. *Perception & Psychophysics, 62*(6), 1204-1219.

Fonov, V. S., Evans, A. C., McKinstry, R. C., Almli, C. R., & Collins, D. L. (2009). Unbiased nonlinear average age-appropriate brain templates from birth to adulthood. *NeuroImage,* (47), S102.

Freeman, J. B., Dale, R., & Farmer, T. A. (2011). Hand in motion reveals mind in motion. *Frontiers in psychology*, *2*, 59.

Freud, E., Plaut, D. C., & Behrmann, M. (2016). 'What' is happening in the dorsal visual pathway. *Trends in Cognitive Sciences*, *20*(10), 773-784.

Gallivan, J. P., McLean, D. A., Valyear, K. F., & Culham, J. C. (2013). Decoding the neural mechanisms of human tool use. *Elife*, *2*, e00425.

Gallivan, J. P., & Culham, J. C. (2015). Neural coding within human brain areas involved in actions. *Current opinion in neurobiology*, *33*, 141-149.

Gallivan, J. P., Chapman, C. S., Wolpert, D. M., & Flanagan, J. R. (2018). Decision-making in sensorimotor control. *Nature Reviews Neuroscience*, *19*(9), 519-534.

Gazzola, V., & Keysers, C. (2009). The observation and execution of actions share motor and somatosensory voxels in all tested subjects: single-subject analyses of unsmoothed fMRI data. *Cerebral cortex*, *19*(6), 1239-1255.

Giese, M. A., & Poggio, T. (2003). Neural mechanisms for the recognition of biological movements. *Nature Reviews Neuroscience*, *4*(3), 179-192.

Giese, M. A., & Rizzolatti, G. (2015). Neural and computational mechanisms of action processing: interaction between visual and motor representations. *Neuron*, *88*(1), 167-180.

Goldenberg, G. (1995). Imitating gestures and manipulating a mannikin—the representation of the human body in ideomotor apraxia. *Neuropsychologia*, *33*(1), 63-72.

Goodale, M. A., & Milner, A. D. (1992). Separate visual pathways for perception and action.

Gordon, J., Maselli, A., Lancia, G. L., Thiery, T., Cisek, P., & Pezzulo, G. (2021). The road towards understanding embodied decisions. *Neuroscience & Biobehavioral Reviews*, 131, 722-736.

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.

Gratton, G., Coles, M. G., Sirevaag, E. J., Eriksen, C. W., & Donchin, E. (1988). Pre-and poststimulus activation of response channels: a psychophysiological analysis. *Journal of Experimental Psychology: Human perception and performance, 14*(3), 331.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics* (Vol. 1, pp. 1969-2012). New York: Wiley.

Grill-Spector, K., & Weiner, K. S. (2014). The functional architecture of the ventral temporal cortex and its role in categorization. *Nature Reviews Neuroscience, 15*(8), 536-548.

Hagura, N., Haggard, P., & Diedrichsen, J. (2017). Perceptual decisions are biased by the cost to act. *Elife, 6*, e18422.

Han, Z., Bi, Y., Chen, J., Chen, Q., He, Y., & Caramazza, A. (2013). Distinct regions of right temporal cortex are associated with biological and human–agent motion: functional magnetic resonance imaging and neuropsychological evidence. *Journal of Neuroscience*, *33*(39), 15442-15453.

Handjaras, G., Bernardi, G., Benuzzi, F., Nichelli, P. F., Pietrini, P., & Ricciardi, E. (2015). A topographical organization for action representation in the human brain. *Human brain mapping*, *36*(10), 3832-3844.

Hardwick, R. M., Caspers, S., Eickhoff, S. B., & Swinnen, S. P. (2018). Neural correlates of action: Comparing meta-analyses of imagery, observation, and execution. *Neuroscience and Biobehavioral Reviews*, *94*, 31-44.

Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science, 293*(5539), 2425-2430.

Haxby, J. V., Guntupalli, J. S., Connolly, A. C., Halchenko, Y. O., Conroy, B. R., Gobbini, M. I., ... & Ramadge, P. J. (2011). A common, high-dimensional model of the representational space in human ventral temporal cortex. *Neuron, 72*(2), 404-416.

Haxby, J. V., Connolly, A. C., & Guntupalli, J. S. (2014). Decoding neural representational spaces using multivariate pattern analysis. *Annual review of neuroscience, 37*, 435-456.

Hauser, C.K., Salinas, E. (2014). Perceptual Decision Making. In: Jaeger, D., Jung, R. (eds) *Encyclopedia of Computational Neuroscience*. New York: Springer.

Hodgson, V. J., Lambon Ralph, M. A., & Jackson, R. L. (2022). The cross-domain functional organization of posterior lateral temporal cortex: insights from ALE meta-analyses of 7 cognitive domains spanning 12,000 participants. *Cerebral Cortex*.

Huth, A. G., Nishimoto, S., Vu, A. T., & Gallant, J. L. (2012). A continuous semantic space describes the representation of thousands of object and action categories across the human brain. *Neuron*, *76*(6), 1210-1224.

Huth, A. G., De Heer, W. A., Griffiths, T. L., Theunissen, F. E., & Gallant, J. L. (2016). Natural speech reveals the semantic maps that tile human cerebral cortex. *Nature*, *532*(7600), 453-458.

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.

Jastorff, J., Begliomini, C., Fabbri-Destro, M., Rizzolatti, G., & Orban, G. A. (2010). Coding observed motor acts: different organizational

principles in the parietal and premotor cortex of humans. *Journal of neurophysiology, 104*(1), 128-140.

Johansson, R. S., & Westling, G. (1984). Roles of glabrous skin receptors and sensorimotor memory in automatic control of precision grip when lifting rougher or more slippery objects. *Experimental brain research, 56*, 550-564.

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.

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.

Kemmerer, D. (2021). What modulates the mirror neuron system during action observation?: Multiple factors involving the action, the actor, the observer, the relationship between actor and observer, and the context. *Progress in Neurobiology, 205*, 102128.

Kilner, J. M. (2011). More than one pathway to action understanding. *Trends in cognitive sciences*, *15*(8), 352-357.

Koul, A., Cavallo, A., Cauda, F., Costa, T., Diano, M., Pontil, M., & Becchio, C. (2018). Action observation areas represent intentions from subtle kinematic features. *Cerebral Cortex, 28*(7), 2647-2654.

Kravitz, D. J., Saleem, K. S., Baker, C. I., & Mishkin, M. (2011). A new neural framework for visuospatial processing. *Nature Reviews Neuroscience*, *12*(4), 217-230.

Kriegeskorte, N., Goebel, R., & Bandettini, P. (2006). Information-based functional brain mapping. *Proceedings of the National Academy of Sciences, 103*(10), 3863-3868.

Kriegeskorte, N., Mur, M., & Bandettini, P. A. (2008a). Representational similarity analysis-connecting the branches of systems neuroscience. *Frontiers in systems neuroscience, 2*, 4.

Kriegeskorte, N., Mur, M., Ruff, D. A., Kiani, R., Bodurka, J., Esteky, H., ... & Bandettini, P. A. (2008b). Matching categorical object representations in inferior temporal cortex of man and monkey. *Neuron, 60*(6), 1126-1141.

Króliczak, G., & Frey, S. H. (2009). A common network in the left cerebral hemisphere represents planning of tool use pantomimes and familiar intransitive gestures at the hand-independent level. *Cerebral Cortex, 19*(10), 2396-2410.

Kruskal, J. B. (1964). Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika, 29*(1), 1-27.

Kubiak, A., & Króliczak, G. (2016). Left extrastriate body area is sensitive to the meaning of symbolic gesture: evidence from fMRI repetition suppression. *Scientific Reports, 6*(1), 1-13.

Lepora, N. F., & Pezzulo, G. (2015). Embodied choice: how action influences perceptual decision making. *PLoS computational biology, 11*(4), e1004110.

Lesourd, M., Osiurak, F., Baumard, J., Bartolo, A., Vanbellingen, T., & Reynaud, E. (2018). Cerebral correlates of imitation of intransitive gestures: An integrative review of neuroimaging data and brain lesion studies. *Neuroscience & biobehavioral reviews*, *95*, 44-60.

Liepmann, H. (1988). *Apraxie, Originally Published in ''Ergebnisse der Gesamten Medizin'' Vol. 1, pp. 516–543, 1920*. In J. S. Brown (Ed.), *Agnosia and apraxia: Selected papers of Liepmann, Lange and Potzl*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Lin, Y., & Tadin, D. (2021). Estimating decision time in perceptual decision making. *Journal of Vision, 21*(9), 2694-2694.

Lloyd, S. (1982). Least squares quantization in PCM. *IEEE transactions on information theory, 28*(2), 129-137.

Luce, R. D. (1986). *Response times: Their role in inferring elementary mental organization* (No. 8). Oxford University Press on Demand.

Lui, F., Buccino, G., Duzzi, D., Benuzzi, F., Crisi, G., Baraldi, P., ... & Rizzolatti, G. (2008). Neural substrates for observing and imagining non-object-directed actions. *Social neuroscience, 3*(3-4), 261-275.

Maaten, L. V. D., & Hinton, G. (2008). Visualizing data using t-SNE. *Journal of machine learning research, 9*(Nov), 2579-2605.

Macmillan, N. A., and Creelman, C. D. (2004). *Detection theory: A user's guide*. Psychology press.

Mah, Y. H., Husain, M., Rees, G., & Nachev, P. (2014). Human brain lesion-deficit inference remapped. *Brain, 137*(9), 2522-2531.

Marcos, E., Cos, I., Cisek, P., Girard, B., & Verschure, P. F. (2013). Biomechanical costs of reaching movements bias perceptual decisions. *BMC Neuroscience, 14*, 1-2.

McKinstry, C., Dale, R., & Spivey, M. J. (2008). Action dynamics reveal parallel competition in decision making. *Psychological Science, 19*(1), 22-24.

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.

Möttönen, R., Farmer, H., & Watkins, K. E. (2016). Neural basis of understanding communicative actions: Changes associated with knowing the actor's intention and the meanings of the actions. *Neuropsychologia, 81,* 230-237.

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.

Newell, A., & Simon, H. A. (1972). *Human problem solving* (Vol. 104, No. 9). Englewood Cliffs, NJ: Prentice-hall.

Nishimoto, S., Vu, A. T., Naselaris, T., Benjamini, Y., Yu, B., & Gallant, J. L. (2011). Reconstructing visual experiences from brain activity evoked by natural movies. Current biology, 21(19), 1641-1646.

O'Connell, R. G., Dockree, P. M., & Kelly, S. P. (2012). A supramodal accumulation-to-bound signal that determines perceptual decisions in humans. *Nature neuroscience, 15*(12), 1729-1735.

O'Connell, R. G., & Kelly, S. P. (2021). Neurophysiology of human perceptual decision-making. *Annual Review of Neuroscience, 44*, 495-516.

Ozbagci, D., Moreno-Bote, R., & Soto-Faraco, S. (2021). The dynamics of decision-making and action during active sampling. *Scientific Reports, 11*(1), 23067.

Papeo, L., Agostini, B., & Lingnau, A. (2019). The large-scale organization of gestures and words in the middle temporal gyrus. *Journal of Neuroscience, 39*(30), 5966-5974.

Pazzaglia, M., Pizzamiglio, L., Pes, E., & Aglioti, S. M. (2008). The sound of actions in apraxia. *Current biology*, *18*(22), 1766-1772.

Peuskens, H., Vanrie, J., Verfaillie, K., & Orban, G. A. (2005). Specificity of regions processing biological motion. *European Journal of Neuroscience, 21*(10), 2864-2875.

Pitcher, D., & Ungerleider, L. G. (2021). Evidence for a third visual pathway specialized for social perception. *Trends in Cognitive Sciences*, *25*(2), 100-110.

Plaisier, M. A., Kuling, I. A., Brenner, E., & Smeets, J. B. (2019). When does one decide how heavy an object feels while picking it up? *Psychological science, 30*(6), 822-829.

Pruszynski, J. A., Kurtzer, I., & Scott, S. H. (2008). Rapid motor responses are appropriately tuned to the metrics of a visuospatial task. *Journal of neurophysiology, 100*(1), 224-238.

Pruszynski, J. A., King, G. L., Boisse, L., Scott, S. H., Flanagan, J. R., & Munoz, D. P. (2010). Stimulus-locked responses on human arm muscles reveal a rapid neural pathway linking visual input to arm motor output. *European Journal of Neuroscience, 32*(6), 1049-1057.

Pruszynski, J. A., Kurtzer, I., Nashed, J. Y., Omrani, M., Brouwer, B., & Scott, S. H. (2011). Primary motor cortex underlies multi-joint integration for fast feedback control. *Nature, 478*(7369), 387-390.

Power, J. D., Barnes, K. A., Snyder, A. Z., Schlaggar, B. L., & Petersen, S. E. (2012). Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. *Neuroimage, 59*(3), 2142-2154.

Pulvermüller, F. (2005). Brain mechanisms linking language and action. *Nature reviews neuroscience*, *6*(7), 576-582.

Rank, M., & Di Luca, M. (2014). Response time-dependent force perception during hand movement. In *Haptics: Neuroscience, Devices, Modeling, and Applications: 9th International Conference, EuroHaptics 2014, Versailles, France, June 24-26, 2014, Proceedings, Part I 9* (pp. 85-92). Springer Berlin Heidelberg.

Rank, M., & Di Luca, M. (2015). Speed/accuracy tradeoff in force perception. *Journal of experimental psychology: human perception and performance, 41*(3), 738.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological review, 85*(2), 59.

Ratcliff, R., Smith, P. L., Brown, S. D., & McKoon, G. (2016). Diffusion decision model: Current issues and history. *Trends in cognitive sciences, 20*(4), 260-281.

Resulaj, A., Kiani, R., Wolpert, D. M., & Shadlen, M. N. (2009). Changes of mind in decision-making. *Nature, 461*(7261), 263-266.

Reynaud, E., Navarro, J., Lesourd, M., & Osiurak, F. (2019). To watch is to work: a review of neuroimaging data on tool use observation network. *Neuropsychology Review, 29*, 484-497.

Savaki, H. E., Kavroulakis, E., Papadaki, E., Maris, T. G., & Simos, P. G. (2022). Action observation responses are influenced by movement kinematics and target identity. *Cerebral Cortex, 32*(3), 490-503.

Scott, S. H. (2012). The computational and neural basis of voluntary motor control and planning. *Trends in cognitive sciences, 16*(11), 541-549.

Selen, L. P., Shadlen, M. N., & Wolpert, D. M. (2012). Deliberation in the motor system: reflex gains track evolving evidence leading to a decision. *Journal of Neuroscience, 32*(7), 2276-2286.

Servant, M., White, C., Montagnini, A., & Burle, B. (2015). Using covert response activation to test latent assumptions of formal decision-making models in humans. *Journal of Neuroscience, 35*(28), 10371-10385

Sokolov, A. A., Zeidman, P., Erb, M., Ryvlin, P., Friston, K. J., & Pavlova, M. A. (2018). Structural and effective brain connectivity underlying biological motion detection. *Proceedings of the National Academy of Sciences*, *115*(51), E12034-E12042.

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.

Summerfield, C., & De Lange, F. P. (2014). Expectation in perceptual decision making: neural and computational mechanisms. Nature *Reviews Neuroscience, 15*(11), 745-756.

Swensson, R. G. (1972). The elusive tradeoff: Speed vs accuracy in visual discrimination tasks. *Perception & Psychophysics, 12*(1), 16-32.

Tarhan, L., De Freitas, J., & Konkle, T. (2021). Behavioral and neural representations en route to intuitive action understanding. *Neuropsychologia*, *163*, 108048.

Thura, D., & Cisek, P. (2014). Deliberation and commitment in the premotor and primary motor cortex during dynamic decision making. *Neuron, 81*(6), 1401-1416.

Toma, S., & Santello, M. (2019). Motor modules account for active perception of force. *Scientific reports, 9*(1), 1-16.

Tucciarelli, R., Turella, L., Oosterhof, N. N., Weisz, N., & Lingnau, A. (2015). MEG multivariate analysis reveals early abstract action

representations in the lateral occipitotemporal cortex. *Journal of Neuroscience, 35*(49), 16034-16045.

Tucciarelli, R., Wurm, M., Baccolo, E., & Lingnau, A. (2019). The representational space of observed actions. *elife*, *8*, e47686.

Turella, L., Rumiati, R., & Lingnau, A. (2020). Hierarchical Action Encoding Within the Human Brain. *Cerebral Cortex*.

Urgen, B. A., Pehlivan, S., & Saygin, A. P. (2019). Distinct representations in occipito-temporal, parietal, and premotor cortex during action perception revealed by fMRI and computational modeling. *Neuropsychologia, 127*, 35-47.

Urgen, B. A., & Orban, G. A. (2021). The unique role of parietal cortex in action observation: Functional organization for communicative and manipulative actions. *NeuroImage, 237*, 118220.

van Polanen, V., & Davare, M. (2015). Sensorimotor memory biases weight perception during object lifting. *Frontiers in human neuroscience, 9*, 700.

van Polanen, V., & Davare, M. (2019). Dynamic size-weight changes after object lifting reduce the size-weight illusion. *Scientific Reports, 9*(1), 1-10.

Vannuscorps, G., F Wurm, M., Striem-Amit, E., & Caramazza, A. (2019). Large-scale organization of the hand action observation network in individuals born without hands. *Cerebral Cortex*, *29*(8), 3434-3444.

Villarreal, M., Fridman, E. A., Amengual, A., Falasco, G., Gerscovich, E. R., Ulloa, E. R., & Leiguarda, R. C. (2008). The neural substrate of gesture recognition. *Neuropsychologia, 46*(9), 2371-2382.

Ward, E., Brownsett, S. L., McMahon, K. L., Hartwigsen, G., Mascelloni, M., & de Zubicaray, G. I. (2022). Online transcranial magnetic stimulation reveals differential effects of transitivity in left inferior parietal cortex but not premotor cortex during action naming. *Neuropsychologia*, *174*, 108339.

Wiener, M., Zhou, W., Bader, F., & Joiner, W. M. (2019). Movement improves the quality of temporal perception and decision-making. *eneuro, 6*(4).

Winkler, A. M., Ridgway, G. R., Douaud, G., Nichols, T. E., & Smith, S. M. (2016). Faster permutation inference in brain imaging. *Neuroimage, 141*, 502-516.

Wurm, M. F., & Lingnau, A. (2015). Decoding actions at different levels of abstraction. *Journal of Neuroscience*, *35*(20), 7727-7735.

Wurm, M. F., Ariani, G., Greenlee, M. W., & Lingnau, A. (2016). Decoding concrete and abstract action representations during explicit and implicit conceptual processing. *Cerebral Cortex, 26*(8), 3390-3401.

Wurm, M. F., Caramazza, A., & Lingnau, A. (2017). Action categories in lateral occipitotemporal cortex are organized along sociality and transitivity. *Journal of Neuroscience, 37*(3), 562-575.

Wurm, M. F., & Caramazza, A. (2019). Distinct roles of temporal and frontoparietal cortex in representing actions across vision and language. *Nature communications, 10(*1), 289.

Wurm, M. F., & Caramazza, A. (2022). Two 'what' pathways for action and object recognition. *Trends in cognitive sciences*, *26*(2), 103-116.

Xu, J., Gannon, P. J., Emmorey, K., Smith, J. F., & Braun, A. R. (2009). Symbolic gestures and spoken language are processed by a common neural system. *Proceedings of the National Academy of Sciences, 106*(49), 20664-20669.

Yang, J., Andric, M., & Mathew, M. M. (2015). The neural basis of hand gesture comprehension: a meta-analysis of functional magnetic resonance imaging studies. *Neuroscience & Biobehavioral Reviews*, 57, 88-104.

Zeki, S. (2015). Area V5—a microcosm of the visual brain. *Frontiers in integrative neuroscience, 9*, 21.