

**Classification:** Biological Sciences (Psychological and Cognitive Sciences)

**Title:** “Spatial frequency tuning of motor responses reveals differential contribution of dorsal and ventral systems to action comprehension”.

**Running title:** “Contribution of spatial frequencies to action comprehension”

**Lucia Amoruso<sup>1,2\*</sup> Alessandra Finisguerra<sup>3</sup> and Cosimo Urgesi<sup>1,3</sup>**

<sup>1</sup> *Laboratory of Cognitive Neuroscience, Department of Languages and Literatures, Communication, Education and Society, University of Udine, I-33100, Udine, Italy*

<sup>2</sup> *Basque Center on Cognition, Brain and Language, San Sebastian, 20009, Spain*

<sup>3</sup> *Scientific Institute, IRCCS E. Medea, Pasi di Prato, I-33037, Udine, Italy*

\*Author for correspondence:

Lucia Amoruso, PhD

Laboratory of Cognitive Neuroscience

DILL Department, University of Udine

Via Margreth, 3, 33100 Udine, Italy

Tel. +39-0432-249889

e-mail: l.amoruso@bcbl.eu

**Competing interest statement:** The authors declare no competing interests.

**Key words:** Action Comprehension, Spatial Frequency, Context, Transcranial magnetic stimulation

## **Abstract**

Understanding object-directed actions performed by others is central to everyday life. This ability is thought to rely on the interaction between the dorsal action observation network (AON) and a ventral object recognition pathway. On this view, the AON would encode action kinematics and the ventral pathway the most likely intention afforded by the objects. However, experimental evidence supporting this model is still scarce. Here, we aimed to disentangle the contribution of dorsal vs. ventral pathways to action comprehension by exploiting their differential tuning to low- (LSF) and high-spatial frequencies (HSF). We filtered naturalistic action images to contain only LSF or HSF and measured behavioral performance and corticospinal excitability (CSE) using TMS. Actions were embedded in congruent or incongruent scenarios as defined by the compatibility between grips and intentions afforded by the contextual objects. Behaviorally, participants were better at discriminating congruent actions in intact than LSF images. This effect was reversed for incongruent actions, with better performance for LSF than intact and HSF. These modulations were mirrored at the neurophysiological level, with greater CSE facilitation for congruent than incongruent actions for HSF and the opposite pattern for LSF images. Finally, only for LSF we observed CSE modulations according to grip kinematics. While results point to differential dorsal (LSF) and ventral (HSF) contributions to action comprehension for grip and context encoding, respectively, the negative congruency effect for LSF images suggests that object processing may influence action perception not only through ventral-to-dorsal connections, but also through a dorsal-to-dorsal route involved in predictive processing.

## **Significance Statement**

Understanding others' intentions via observing their actions is critical for effective social coping and survival. There is evidence that visual processing of movements and objects in context involves segregated dorsal and ventral pathways, which are tuned to distinct spatial frequencies (low and high, respectively). While this frequency tuning has been widely studied in object, face and scene

processing, little is known about its role in the recognition of naturalistic human actions. Here, we provide initial evidence for the contribution of dorsal and ventral pathways to action processing, with low and high spatial frequencies differentially encoding global grip kinematics and local object features. These preliminary findings inform current models on action comprehension, suggesting that frequency preferences can modulate this process.

## **Introduction**

Understanding object-directed actions in context is central to everyday life. However, the intentional states guiding these actions might differ in a way that can be potentially beneficial or harmful for the observer (i.e., someone may grasp a scalpel to cure or to hurt). Thus, detecting others' intentions is critical for social cognition and survival. The way in which the brain succeeds in exploiting relevant perceptual features about objects and movements to infer others' intentions has received wide attention in the last years. Yet, the role that contextual cues would play under these circumstances and how they would be integrated with action information remains poorly understood.

Classical views on visuo-motor control (1, 2) suggest that action and object processing occur along separate brain pathways, with reach-to-grasp movement information being mainly processed by a dorsal stream projecting from primary visual areas to parietal ones; and object recognition mostly relying on ventral projections from visual to temporal regions. Interestingly, a recent two-pathway model (3) tried to account for dorsal and ventral contributions during action comprehension via observation, highlighting the potential role that context would play to disambiguate action intentions. According to this model, the encoding of the concrete motor aspects (i.e., action kinematics) would take place through the classical dorsal action observation network (AON), while the underlying intention would be estimated from the context in areas beyond the AON, through a ventral gradient linking the middle temporal gyrus and the anterior part of the inferior frontal gyrus. The existence of an interaction between both pathways has been proposed for object recognition (4, 5). On this view, areas in the prefrontal cortex (PFC) would receive coarse, low-resolution information via fast dorsal projections and generate predictions about object identity. This prediction would be feedback to the temporal cortex, facilitating recognition by limiting the number of possible object candidates. This appealing view is based on evidence (6) indicating that dorsal and ventral streams are dominated by different neural contributions, with the former one primarily

receiving input from the magnocellular layers of the lateral geniculate nucleus of the thalamus, and the latter from the parvocellular ones. Neurons in magnocellular layers are sensitive to low-spatial frequencies (LSF), which rapidly carry coarse information about the global aspect of a stimulus; conversely, those from parvocellular layers preferentially respond to high-spatial frequencies (HSF) conveying local information about fine-grained details (7). Thus, a common experimental manipulation to bias processing towards dorsal and ventral pathways consists in filtering stimuli to contain only LSF or HSF, respectively. So far, these dissociable frequency preferences have been widely studied in face (8), object (4, 5) and scene categorization (9-11). However, little is known about LSF and HSF contributions to the processing of natural images implying object-directed actions in context. Investigating spatial frequency tuning is a compelling medium for examining the complex interactions between perceptual and cognitive processes (12) which, in turn, might have implications for understanding how we encode complex visual stimuli that we encounter in everyday life, such as context-embedded actions.

The current study builds on previous evidence showing that corticospinal excitability (CSE) in the motor system becomes differentially modulated depending on the context in which an action is observed (for a review see (13)). Briefly, in a series of studies we found that the observation of movement kinematics occurring in congruent contexts (i.e., reaching-to-grasp a cup full of coffee with a precision grip) increases motor CSE facilitation, while its observation in incongruent contexts (i.e., reaching-to-grasp an empty cup with a precision grip) results in decreased motor facilitation as compared to observing the same kinematics in neutral contexts. Of note, these congruency patterns were disrupted after interfering with activity in temporal and PFC via Transcranial magnetic stimulation (TMS) (14). While these findings seem to support current two-pathway models (3), suggesting an interplay between simulative motor responses in the AON and signals in temporal areas encoding the intention estimated from the context, they also point to the existence of alternative sources for the anticipatory encoding of intentional information (e.g., PFC).

These results fit well with neurophysiological evidence from neural recordings in monkeys showing that activity in the ventral PFC encodes object identity for action goal selection in a context-dependent manner (15, 16). Furthermore, the PFC exhibits dense connectivity with premotor regions, the inferotemporal cortex and orbitofrontal areas supporting the existence of different routes for processing actions in context beyond the classical AON (e.g., prefrontal-premotor route) (16, 17). Altogether, these studies provide functional and anatomical evidence for the existence of different routes impacting on action processing, carrying semantic information about objects in context that is further transformed into motor representations of potential actions afforded by them. Yet, their interaction in humans remains largely unexplored.

Here, we capitalized on the differential tuning to LSF and HSF shown by dorsal and ventral pathways to examine whether biasing perceptual processing toward one or the other modulated the ability to comprehend context-embedded actions and the relative (de)-activations of the observer's motor system. Pictures depicting an actor model grasping common objects in congruent or incongruent contexts were filtered to contain only LSF (i.e., magno-biased stimuli) or HSF information (i.e., parvo-biased stimuli). Figure 1A shows examples of stimuli and conditions used in the study. While grasp categorization relied on the global information provided by precision versus power grips directed toward the target object (i.e., either grasping a cup with a pincer grip or with a power grip, respectively), context categorization depended on the local elements present in the environment (i.e., biscuits and coffee inside a cup for a breakfast context, biscuit crumbs and an empty cup for a cleaning context). Considering evidence (18) indicating that images showing ongoing but incomplete actions can trigger simulative motor mapping in the observer's motor system, we selected those images displaying the action intermediate phase (i.e., before the model made full contact with the object), so that images depicted the pre-shaping of the hand configuration during the reaching-to-grasp phase of the movement.

In two different main experiments, one behavioral and the other employing the same paradigm but combined with electromyography and 'online' single-pulse TMS over the hand representation of the primary motor cortex (M1), participants were requested to observe the pictures and to identify, given grasping and contextual information, which was the most likely intention underlying the observed action (i.e., to drink vs. to clean).

We hypothesized that, if ventral and dorsal pathways are tuned to process object details and grasping movements, respectively (H1), only congruency effects, but not global grip distinctions should be triggered by HSF information (preferentially engaging the ventral parvocellular pathway); conversely, only global grip distinctions, but not congruency effects should be present for LSF stimuli, which are mainly processed by the dorsal magnocellular pathway. In addition, based on current views suggesting a key role of LSF in the generation of predictions about object identity (4) we hypothesized that, if the dorsal pathway uses LSF to constrain object recognition (H2), then both contextual and grip effects should be triggered by LSF stimuli. Furthermore, since the perceptual style of individuals with higher autistic traits is characterized by superior local processing, diminished global perception (19) and reduced magnocellular function (20), we predicted that participant's visual processing style, measured as the amount of attention to detail autistic trait (21), should be associated with greater performance and CSE facilitation for the parvo-biased HSF stimuli and lower performance and CSE facilitation for the magno-biased LSF stimuli. Overall, based on previous studies, we predicted better behavioral performance and higher CSE for congruent than incongruent contexts. We also expected contextual modulations (i.e., CSE facilitation for congruent and suppression for incongruent contexts) to be mainly reflected in the HSF condition and grip modulations in the LSF one (i.e., increase CSE facilitation for precision vs. power grips).

## Results

**Experiment 1: Spatial Frequency Contributions to Action Comprehension.** Sixteen volunteers took part in the behavioral study (11 women;  $M = 23.75$ ,  $SD = 6.51$ ). In a two-alternative forced choice (2AFC) task, participants observed pictures depicting an agent model performing different everyday actions in congruent and incongruent contexts and inferred the overarching intention that caused them. Trials started with a visual warning cue (5s), followed by an intact (i.e., broad-band), HSF or LSF picture presented for 300 or 500ms and, finally, by a frame with the verbal labels of the 2 possible intentions (e.g., “to drink” and “to clean”). To build-up their predictions, participants were instructed to carefully pay attention to both aspects of the scene: the agent’s hand pre-shaping and the contextual information in which the action was embedded. It is worth noting that, in keeping with previous studies (11-13), correct responses were defined by hand configurations in terms of precision vs. power grips, not by the context. Thus, when the participants observed a precision grip of a cup, “to drink” was the correct response irrespectively of the breakfast or cleaning scenario.

Proportion of correct responses (%) were calculated for each participant and condition and subjected to a repeated-measures ANOVA with Context (congruent, incongruent), Spatial Frequency (intact, HSF, LSF) and Grip (precision grip, power grip) as within-subjects variables. The analysis yielded a main effect of Context ( $F_{1, 15} = 23.7$ ,  $p = 0.0002$ ,  $\eta_p^2 = 0.61$ ), with overall higher performance in action recognition for congruent (mean = 87.71%, SEM = 2.65%) as compared to incongruent contexts (mean = 54.99%, SEM = 7.03%).

Significant Context x Spatial frequency ( $F_{2, 30} = 14.8$ ,  $p = 0.00003$ ,  $\eta_p^2 = 0.49$ ) and Spatial frequency x Grip ( $F_{2, 30} = 5.9$ ,  $p = 0.006$ ,  $\eta_p^2 = 0.28$ ) interactions were also observed.

Post-hoc comparisons on the Context x Spatial frequency interaction (MSE = 129, df = 30) indicated that congruency effects were significant for all spatial frequency conditions (all  $ps < 0.0001$ ; all Cohen's  $d > 0.77$ ). However, within congruent contexts, participants were less accurate



in the LSF condition (mean = 84.06%, SEM = 2.73%) as compared to the intact one (mean = 91.52%, SEM = 2.43%;  $p = 0.01$ , Cohen's  $d = 0.68$ ), while the HSF condition did not differ from the other two (mean = 87.56%, SEM = 2.52%, all  $ps > 0.16$ ; all Cohen's  $d < 0.3$ ). Conversely, within incongruent contexts, participants' performance was better for LSF (mean = 61.98%, SEM = 5.84%) as compared to intact (mean = 47.6%, SEM = 7.57%;  $p = 0.00008$ ; Cohen's  $d = -1.3$ ) and HSF conditions (mean = 55.4%, SEM = 7.3%;  $p = 0.02$ ; Cohen's  $d = -0.6$ ). In addition, recognition performance was better for HSF than for intact images ( $p = 0.01$ ; Cohen's  $d = 0.7$ ). See Fig. 2A. Post-hoc comparisons on the Spatial frequency x Grip interaction (MSE = 62.57, df = 30) showed grip differences in the LSF condition ( $p = 0.01$ ; Cohen's  $d = 0.43$ ), with higher accuracy for precision (mean = 75.78%, SEM = 5.9%) as compared to power grips (mean = 70.27%, SEM = 4.63%). No significant differences were observed between grips in the HSF (precision: mean = 69.53%, SEM = 7.56%; power: 73.43%, SEM = 5.93%;  $p = 0.08$ ; Cohen's  $d = -0.3$ ) or intact conditions (precision: mean = 69.07%, SEM = 8.35%; power: 70.05%, SEM = 7.46%;  $p = 0.6$ ; Cohen's  $d = -0.08$ ).

**Correlations between Perceptual Traits and Behavioral Results.** After the experiment, all participants completed the Italian version (22) of the Autism Spectrum Quotient (AQ) (21). The AQ is a self-report questionnaire that quantifies the degree to which individuals with normal intelligence have the traits associated to the autistic spectrum via assessing five different factors: *Social skills*, *Attention switching*, *Attention to detail*, *Communication* and *Imagination*; with higher scores endorsing more autistic-like behavior. In this particular case, we were interested in the *Attention to detail* subscale, which measures non-social aspects related to perceptual atypicalities in Autism but also present in the non-clinical population. Table 1 shows mean, standard deviation, and range for all AQ-subscale scores.

We ran Pearson correlation coefficients to investigate potential relationships between behavioral performance and participants' perceptual traits. Since the 3-way interaction between Context,

Spatial frequency and Grip was not significant, we collapsed grips and tested the correlation of the 5 AQ subscales with performance for congruent and incongruent contexts in the 3 spatial frequency conditions (total number of comparisons = 30), using a Bonferroni-adjusted alpha level of  $\alpha = 0.0016$  (i.e.,  $\alpha = 0.05/30$ ). Results showed significant positive correlations between accuracy in the HSF congruent condition and the *Attention to detail* AQ-subscale ( $r = 0.77$ ,  $p = 0.0004$ ; 95% CI = [0.58 0.92]), indicating that the more participants succeeded in identifying actions when only featural information was available, the higher their local processing style was (see Fig. 3). No other correlations survived Bonferroni correction. Since a relatively small sample size and the presence of potential outliers can lead to spurious correlations, we further estimated the robustness of the observed association by computing Pearson-skipped and percentage-bend correlations, which are known to provide better estimates of the true relationship between two variables. Importantly, this analysis further confirmed the existence of a significant association between action perception performance and perceptual style traits (Bend  $r = 0.81$ ,  $p = 0.0001$ , 95% CI = [0.55 0.94]; Pearson skipped  $r = 0.77$ ,  $t = 4.57$ , 95% CI = [0.55 0.91]).

**Experiment 2: Neurophysiological Correlates of Low- and High-Frequency Contributions to Action Comprehension.** Once we established that biasing perceptual processing toward dorsal and ventral pathways modulates action comprehension, in Experiment 2 we moved further to unveil the neurophysiological correlates of these effects. Sixteen volunteers (11 women;  $M = 22.18$ ,  $SD = 3.83$ ) took part in this experiment. Overall, we used the same paradigm (i.e., task, stimuli and experimental manipulation) as in Experiment 1, but here we combined it with the acquisition of electromyographic responses and 'online' single-pulse TMS. TMS-elicited motor-evoked potentials (MEPs) were recorded from the First Dorsal Interosseous (FDI) and from a control muscle, the Extensor Carpi Radialis (ECR). While both muscles are involved in reach-to-grasp movements, only the FDI is involved in determining the grasping kinematics that differentiates the two alternative actions in each pair (23). In addition, baseline CSE was assessed by acquiring 10 MEPs

before and 10 MEPs after the experimental task while participants passively watched a fixation cross. The mean raw MEP amplitudes recorded from the FDI and the ECR muscles in the different action observation conditions and baseline trials are reported in Table 2.

First, we run a 2-way ANOVA on the raw MEP amplitudes recorded during the baseline blocks before and after the experiment to rule out differences between FDI and ECR muscles not related to the experimental manipulation. No main effects of block ( $F_{1, 15} = 0.6, p = 0.45, \eta_p^2 = 0.03$ ) or muscle ( $F_{1, 15} = 4.19, p = 0.058, \eta_p^2 = 0.21$ ), neither a significant interaction between these effects ( $F_{1, 15} = 0.38, p = 0.54, \eta_p^2 = 0.02$ ) were observed, indicating that our stimulation parameters allowed reliable recording of comparable MEP amplitudes from both muscles throughout the experimental session.

The individual mean MEP amplitudes recorded during the action observation task were log-transformed and normalized by subtracting the spatially filtered conditions (HSF and LSF) to the intact condition (broad-band) in order to test the specific modulation of CSE according to the frequency left in the images. An omnibus RM-ANOVA was performed on the normalized values with Muscle (FDI, ECR), Context (congruent, incongruent), Spatial Frequency (HSF, LSF) and Grip (precision grip, power grip) as within-subjects variables. The analysis revealed significant 3-way interactions of Muscle x Context x Spatial frequency ( $F_{1, 15} = 8.73, p = 0.009, \eta_p^2 = 0.36$ ) and Context x Spatial frequency x Grip ( $F_{1, 15} = 6.43, p = 0.02, \eta_p^2 = 0.30$ ), which were further qualified by a significant 4-way interaction of Muscle x Context x Spatial frequency x Grip ( $F_{1, 15} = 4.76, p = 0.04, \eta_p^2 = 0.24$ ).

Post-hoc comparisons performed on the 4-way interaction ( $MSE = 0.0004, df = 15$ ) showed that congruency effects involved only the FDI muscle and those actions performed with a precision grip. No effects were found for the ECR muscle (all  $ps > 0.07$ ; all Cohen's  $d < 0.3$ ). Conversely, a congruency effect was found for the FDI with increased CSE facilitation for precision grip actions

observed in congruent as compared to incongruent contexts when they were presented in their HSF form ( $p = 0.04$ ; Cohen's  $d = 0.36$ ). Notably, however, this pattern was reversed for LSF stimuli, showing greater facilitation for incongruent than for congruent contexts ( $p = 0.002$ ; Cohen's  $d = -0.88$ ). No congruency effects were found for actions performed with power grips either in the HSF or LSF conditions ( $p = 0.21$ ; Cohen's  $d = 0.2$  and  $p = 0.07$ ; Cohen's  $d = 0.32$ , respectively).

Congruency effects are shown in Fig. 4 with solid grey lines.

When considering HSF and LSF conditions within each context, we found that action observation in the congruent condition resulted in increased FDI CSE facilitation for HSF stimuli ( $M = 0.01$ ,  $SEM = 0.01$ ) as compared to LSF ones ( $M = -0.023$ ,  $SEM = 0.01$ ;  $p = 0.001$ ; Cohen's  $d = 1.08$ ).

Conversely, a reversed pattern was found when the same actions were observed within an incongruent context, with increased FDI CSE facilitation for LSF stimuli ( $M = 0.021$ ,  $SEM = 0.01$ ) as compared to HSF ones ( $M = -0.0082$ ,  $SEM = 0.004$ ;  $p = 0.02$ ; Cohen's  $d = -0.94$ ). Spatial frequency effects are shown in Fig. 4 with solid black lines.

LSF stimuli revealed significantly greater FDI CSE facilitation for precision than for power grips, either in congruent (mean =  $-0.023$  and  $-0.0003$ , respectively;  $p = 0.01$ ; Cohen's  $d = -0.45$ ) or incongruent (mean =  $0.0211$  and  $-0.016$ , respectively;  $p = 0.0006$ ; Cohen's  $d = 0.73$ ) contexts. No differences between grips were found for HSF stimuli (all  $ps > 0.25$ ; all Cohen's  $d < 0.19$ ) suggesting that, when removing global information, M1 sensitivity to movement kinematics was reduced. Fig. 4 shows grip modulations with dotted lines.

**Correlations between Perceptual Traits and Neurophysiological Results.** We computed Pearson correlation coefficients to further explore whether the relationship with perceptual traits also kept up for FDI CSE modulations. Correlations were only tested for actions involving the FDI muscle and performed using a precision grip, since significant Context x Spatial Frequency effects were only observed for these conditions. Thus, we run a total of 20 comparisons (i.e., 5 subscales x 2 Contexts x 2 Spatial frequencies), using a Bonferroni-adjusted alpha level of  $\alpha = 0.0025$  (i.e.,  $\alpha =$

0.05/20). Results showed a negative correlation (Pearson  $r = -0.71$ ,  $p = 0.002$ ; 95% CI=[-0.90 - 0.37]) between the LSF congruent condition for the precision grip and the *Attention to detail* subscale, indicating that the higher the participant's local processing style, the lower the level of motor facilitation for actions observed in congruent contexts in their LSF form (see Fig. 5). No other correlations survived the Bonferroni correction. Importantly, also in this case, robust correlation analysis further confirmed the observed association between CSE facilitation and perceptual style traits (Bend  $r = -0.69$   $p = 0.002$  95% CI=[-0.93 -0.31]; Pearson skipped  $r = -0.71$ ,  $t = -3.84$ , 95% CI = [-0.90 -0.42]).

**Supplementary Control Experiment 1.** We ran a supplementary control experiment on a new sample of 16 participants (11 women;  $M = 26.43$  years,  $SD = 5.7$ ) in order to test whether spatial frequency modulations were similar while using dynamic videos instead of static pictures implying movement and thus, our results could be also extended to actions unfolding in time. To the end, we used the original videos from where snapshots showing the intermediate phase of the movement were taken and employed in Experiment 1-2. Videos lasted 500ms and their ending was matched so that the final frame was the same frame shown to participants in Experiment 1-2.

The RM-ANOVA performed on participant's accuracy during action video recognition yielded a main effect of Context ( $F_{1, 15} = 33.53$ ,  $p = 0.00003$ ,  $\eta_p^2 = 0.69$ ) and Frequency ( $F_{1, 15} = 9$ ,  $p = 0.0008$ ,  $\eta_p^2 = 0.37$ ) with overall higher performance in action recognition for congruent (mean = 93.33%, SEM = 2.25%) as compared to incongruent contexts (mean = 59.76%, SEM = 7.03%) and for the LSF (mean = 80.95%, SEM = 4.79%) condition as compared to the HSF (mean = 74.33%, SEM = 6.77%;  $p = 0.001$ ) and the intact (mean = 74.37%, SEM = 8.07%;  $p = 0.001$ ) ones, which in turn did not differ ( $p = 0.98$ ). Furthermore, a significant Context x Spatial frequency interaction ( $F_{2, 30} = 29.21$ ,  $p < 0.00001$ ,  $\eta_p^2 = 0.66$ ) was observed. As in Experiment 1, post-hoc comparisons (MSE = 111.46, df = 30) indicated that congruency effects were significant for all spatial frequency conditions (all  $ps < 0.0001$ ; all Cohen's  $d > 0.73$ ). Within congruent contexts, a similar pattern was

also found, with participants being less accurate in the LSF condition (mean = 89.99%, SEM = 2.71%) as compared to the intact one (mean = 97.46%, SEM = 1.24%;  $p = 0.01$ , Cohen's  $d = 0.72$ ), while the HSF condition did not differ from the other two (mean = 92.56%, SEM = 2.18%, all  $ps > 0.07$ ; all Cohen's  $d < 0.24$ ). Within incongruent contexts, participants' performance was better for LSF (mean = 71.9%, SEM = 5.36%) as compared to intact (mean = 51.27%, SEM = 7.87%;  $p = 0.00006$ ; Cohen's  $d = -1.99$ ) and HSF conditions (mean = 56.1%, SEM = 6.76%;  $p = 0.0001$ ; Cohen's  $d = -1.52$ ). No differences were observed between HSF and intact conditions ( $p = 0.07$ ; Cohen's  $d = 0.46$ ). See Fig. 2B.

Finally, we performed a RM-ANOVA including Experiment as a between-subject variable to test whether the two experiments (i.e., the one using static pictures and the one using dynamic videos) statistically differed. No effects including the factor Experiment yielded significance (all  $Ps > 0.13$ , all  $Fs < 2.37$ ), ensuring that stimuli used in both experiments were comparable. Nevertheless, since frequentist null-hypothesis significance testing does not allow collecting evidence in favor of the null hypothesis and thus, drawing conclusions from negative results, we used the software JASP (19) to implement a Bayesian approach and calculated inclusion Bayes Factors ( $BF_{\text{incl}}$ ) for the RM-ANOVA effects that included the between-subject factor experiment. Briefly,  $BF_{\text{incl}}$  can be interpreted as the evidence in the data for including a given predictor against the (matched) models without it (24). In our particular case, we were interested in the interaction terms including the within-subject variables context and/or spatial frequency and the between-subject factor Experiment. This analysis indicated that, for the interaction terms involving Experiment,  $BF_{\text{incl}}$  did not exceed 0.137, meaning that the data were more than 7 times more likely under the models without the predictor Experiment than under the models including it, thus providing substantial evidence for its irrelevance in the observed outcome.

It is worth mentioning that the frequentist RM-ANOVA yielded a significant Spatial frequency x Grip interaction ( $F_{2,60} = 4.9$ ,  $p = 0.01$ ,  $\eta_p^2 = 0.14$ ), with only the LSF condition differentiating

between grips ( $p = 0.0003$ , Cohen's  $d = 0.47$ ; HSF:  $p = 0.59$ , Cohen's  $d = -0.07$  ; intact:  $p = 0.11$ , Cohen's  $d = 0.2$ ), as in Experiment 1. Based on this result, we performed planned comparisons in the control Experiment between precision and power grips in the LSF, HSF and intact conditions. In line with Experiment 1, significant differences were found in the LSF ( $p = 0.04$ ) but not in the intact ( $p = 0.11$ ) and the HSF ( $p = 0.57$ ) conditions. Overall, the control analysis indicates that both types of stimuli (i.e., static pictures implying an ongoing movement and videos) showed comparable results suggesting that picture-related modulations observed in our TMS experiment could be also extended to the observation of dynamic action videos.

**Supplementary Control Experiment 2.** Finally, following the same logic of the behavioral effect replication, we ran a preliminary control experiment in an independent sample of 9 participants (6 women;  $M = 25.88$  years,  $SD = 5.01$ ) aiming to explore whether the CSE modulation effects could be replicated using dynamic actions in videos rather than in static images. Methods paralleled those used in Experiment 2, but with the 500ms videos implemented in Supplementary Control Experiment 1. spTMS pulses were delivered at the end of video presentation (i.e., 467-533ms from stimulus onset). Giving the preliminary nature of this experiment, we used a Bayesian analysis approach to evaluate evidence in favor of a muscle-selective CSE modulation according to context, spatial frequency (and grip). The results (Fig. 6) provided substantial evidence that the data were modeled by an interaction between muscle, context, and spatial frequency ( $BF_{incl} = 3.2$ ). They also provided anecdotal evidence for a 4-way interaction including grip ( $BF_{incl} = 1.9$ ). Bayesian pairwise comparisons provided anecdotal evidence ( $BF_{+0} = 2.05$ ) for greater FDI CSE facilitation during observation of precision grip actions embedded in congruent as compared to incongruent contexts in the HSF condition, and substantial evidence ( $BF_{+0} = 6.93$ ) for the opposite pattern in the LSF condition. Furthermore, there was substantial evidence ( $BF_{+0} = 3.69$ ) for greater FDI CSE facilitation during observation of precision grips embedded in congruent contexts for the HSF than LSF condition. All other comparisons were not supported ( $0.2 < BF_{+0} < 1$ ). Thus, the Bayesian

analysis on the data of this independent sample of participants provided evidence supporting the main results of Experiment 2. This also suggests that using action videos yields the same patterns of muscle-selective modulation of contextual congruency in the two spatial frequency bands as using static images.

## **Discussion**

In the present study, we aimed to investigate whether biasing perceptual processing toward ventral and dorsal pathways modulated the ability to recognize context-embedded actions and their encoding in the observer's motor system. To this aim, we capitalized on the differential preference for LSF and HSF shown by dorsal and ventral pathways, respectively. While these frequency preferences have been widely studied in other domains such as face (8) scene (9) and object processing (4), to the best of our knowledge this is the first study in examining HSF and LSF contributions to the processing of natural scenes containing object-directed actions performed by others.

In line with previous evidence from face (8), object (4) and scene categorization (11) studies, congruent action stimuli were overall better recognized when presented in their intact full-spectrum form as compared to the spatially-filtered one, although this difference only reached significance when compared to the LSF condition. Conversely, the reverse pattern was true when actions were presented in incongruent contexts, with higher accuracy for LSF as compared to intact and HSF conditions. Since performance was codified based on kinematics (i.e., grip configurations recognition), this reflects that removing local detail information reduced the overall impact of the context, favoring the mere discrimination of kinematics based on LSF. In particular, this suggests that the behavioral interference caused by incongruent contexts was mainly driven by local information present in the scene (i.e., object details) conveyed by HSF.



In addition, global distinctions between grips were mainly captured by LSF. Indeed, when only low-resolution information was presented and, thus, action recognition mostly relied on mere global kinematics encoding, higher recognition accuracy for precision vs. power grips was observed. In keeping with previous findings (25), this effect might suggest that, at least when ambiguous information is presented, it is easier to detect a precision grip than a power grip, since the former deviates more and earlier than the latter from the hand starting position. In other words, it may be easier to detect when the index finger flexion deviates from the other fingers to perform a precision grip than to exclude such a deviation in a power grip.

These behavioral results were mirrored at the neurophysiological level (Experiment 2), with greater motor facilitation (i.e., higher MEP amplitudes) for congruent contexts when action images contained only HSF, and a suppression of this facilitation when they contained only LSF. Likewise, the motor inhibition triggered by incongruent contexts (i.e., lower MEP amplitudes) was present only for HSF stimuli, while it was reversed into a facilitation for LSF stimuli. In addition, CSE also differentiated between grip configurations, with larger FDI MEPs during precision as compared to power grips. As shown by previous studies, this difference might reflect that distinct cortical circuits (26) and pools of motor neurons are selectively recruited depending on grip type, thus leading to differences in CSE modulation (27). Furthermore, evidence from monkeys (28) and humans (29) points to a larger cortical representation of precision than power grips, an aspect that might have contributed to the increase CSE observed for actions performed using precision grips.

In sum, both the behavioral and the neurophysiologic results suggest that the facilitation for congruent and the inhibition for incongruent contexts reported in previous studies (14, 25, 30) mostly rely on HSF information, which is likely conveyed by the ventral/parvocellular system. Conversely, kinematic encoding of grip aperture mostly relies on LSF information, which is likely conveyed by the dorsal/magnocellular one.

Finally, correlational analyses showed that both behavioral performance and motor CSE facilitation in response to actions embedded in congruent contexts were predicted by the *Attention to detail* subscale, which measures a perceptual bias toward superior local processing (19). Indeed, the more participants tended to pay attention to local information the more they were able to recognize actions using parvo-biased stimuli (HSF), but the less their motor system was facilitated by the observation of global kinematics in magno-biased one (LSF). Overall, these results highlight a key contribution of participants' visual processing style to the observed effects, by showing that those individuals with a higher local bias were more affected by the presence/elimination of fine-grained object information.

It is worth mentioning that the observed CSE modulations are unlikely owing to a general effect of spatial attention. First, posture and object-related information were distributed across the scene in a balanced fashion and the proximity between them, implying an imminent hand-object interaction, occurred always at the center of the image. Second, the muscle specificity of the CSE effects and the condition selectivity of the behavioral effects would speak against it. Indeed, irrespectively of context and spatial frequency information, motor CSE modulations were specifically observed for the FDI, a hand muscle that is differently facilitated by the observation of precision versus power grips, while no effects were observed for the ECR, an arm muscle that does not show any differential activation for the observation/execution of either precision or power grips (31, 32).

Nevertheless, as suggested by our correlational results, attention was indeed playing a key role in the observed modulations, but this effect was rather related to the observer's focused attention style as shown by trait measures.

Collectively, our findings align well with H1, suggesting that ventral (parvocellular) and dorsal (magnocellular) streams differently contribute to the processing of context-embedded actions, with the former mainly encoding action intentions based on object features, and the latter action grasping kinematics. This is in keeping with the two-pathway model of action recognition (3), in which the

generation of prior expectations about the most likely intention of an action is mediated by the ventral pathway, while the concrete motor implementation is supported by the dorsal AON.

However, it is also true that the elimination of fine-grained local information did not completely suppress the congruency effects since, in contrast with what was observed for HSF and intact stimuli, the LSF condition showed greater CSE facilitation to incongruent than congruent contexts. This opens the possibility, in line with H2, that coarse LSF information about the objects present in the scene may have been used at earlier stages to form context-based initial guesses, namely predictions of the most likely grip based on the affordances triggered by the object in a given context. Interestingly, recent evidence from monkey recordings (33) aligns well with this view by showing that neural activity in F6, an area bridging PFC and premotor regions, supports a similar mechanism for the representation of object affordance. This mechanism would allow observers to anticipate others' actions by recruiting the same motor representation that would be activated if they were to act upon that object in a similar context.

Nonetheless, the specific effect of these early coarse representations on the motor system could not be detected in our study since we recorded motor facilitation as late as 300ms after stimulus onset. Studies investigating the time-course of predictive top-down signals during single object or scene categorization have reported prefrontal activation triggered by LSF images around ~130ms (4) and ~140-160ms (9) after stimulus presentation, respectively. Thus, it is likely that LSF information may have been used to generate context-based prediction signals and modulated motor activity but earlier in time. Future studies exploring this early time-window (~130-150ms) prone to top-down influences potentially engaging the magnocellular system are clearly required to clarify this issue.

Noteworthy, not only we failed at finding a positive compatibility effect on CSE for LSF stimuli, namely a greater facilitation for congruent than incongruent contexts, but, an opposite pattern of facilitation was unexpectedly observed, with greater CSE for incongruent than congruent contexts. This negative compatibility effect of context on action intention encoding resembles negative

priming effects reported in motor control studies for either symbolic (34, 35) or action primes presented in conditions of low visibility (36). According to classical views on negative compatibility effects in the motor domain (34, 35), a motor representation automatically activated by a visual stimulus that is interrupted (e.g., briefly displayed or masked) is promptly inhibited, leading to greater responses for incongruent than congruent prime-target pairs. Of note, as suggested by EEG recordings over the motor cortex, this effect depends also on timing: while positive compatibility effects appear earlier in time, within 200ms after prime onset, the reversal to a negative compatibility effect occurs between 300-400ms. Interestingly, this mechanism (i.e., active inhibition) has been recently hypothesized for unchosen initial guesses during visual object recognition (37). According to this perspective, a coarse object representation quickly conveyed by LSF information through the magnocellular pathway could resemble, and therefore prime, multiple object candidates (e.g., a drill, a hairdryer, a gun). However, only one representation would be ultimately selected as the correct one as more detailed HSF evidence is accumulated, with the unselected candidates being actively suppressed. Likewise, an action representation that is activated by a LSF-based initial guess may undergo a similar fate (i.e., inhibited) when it does not encounter further disambiguating HSF information, as it was the case in our experiment.

Altogether, our findings provide preliminary support for two-pathway models of action comprehension (3), suggesting that congruency effects are mediated by object-related HSF information processed in the ventral pathway and integrated with the grasping LSF information processed in the dorsal AON. A limitation of the current study, however, is the relatively small sample size, which prevents from drawing strong conclusions on the generalizability of these effects. Notwithstanding, these initial findings outline avenues for future research suggesting the existence of a third pathway to action comprehension, in which a coarse LSF representation of objects in context would be rapidly projected to PFC and feedback not only to temporal regions (4, 5, 38) but also dorsally to the AON (i.e., premotor), ultimately providing a prediction signal on the

expected action kinematics afforded by them (see Fig. 7). This latter aspect has been robustly demonstrated in monkeys (16, 33, 39) and the existence of a similar mechanism in humans seems highly possible. Future studies are needed to disentangle whether these prediction signals would modulate AON activity via a dorsal-dorsal route (i.e., prefrontal-premotor connection) or indirectly through a dorsal-ventral-dorsal network (i.e., prefrontal-temporal-premotor connections). While evidence from monkey studies (15, 17) seems to support the first possibility, its existence in humans awaits direct testing.

## **Methods**

**Participants.** 32 individuals recruited at the University of Udine were randomly allocated to Experiment 1 or 2. In addition, 16 different individuals took part in the Supplementary Control Experiment 1, and 9 in the Supplementary Control Experiment 2, for a total of 57 participants. All individuals were all right-handed according to the Standard Handedness Inventory (40), had normal or corrected-to-normal visual acuity and were free from any contraindication to TMS (41). None of the participants reported history of neurological, psychiatric, or other major medical problems. They all gave their written informed consent prior to experimentation and received course credits for their contribution. The experimental procedures were approved by the local Ethics Committee (Comitato Etico Regionale Unico, Friuli Venezia Giulia, Italy) and were carried out in accordance with the revised Helsinki declaration (World Medical Association General Assembly 2008). All participants were naïve to the purpose of the experiment and a detailed debriefing was provided only after the whole experiment was completed. The sample size required for our 2 x 3 x 2 repeated-measures ANOVA design (Context x Spatial frequency x Grip) was determined with the G\*power software (42), using the “as in SPSS” option for estimating effect size from partial eta-squared ( $\eta_p^2$ ). The expected effect size was set at 0.25 based on previous studies of contextual modulation to observed actions (25, 30), alpha-level at 0.05, and desired power (1-beta) at 80%.

**Stimuli and task.** Stimuli were snapshots selected from original videos used in a set of previous studies (14, 25, 30, 43) taken with a Canon EOS 550D digital camera. All pictures were further edited with the Adobe Photoshop 7 (Adobe Systems, San Jose, CA) software and converted to grayscale. Pictures depicted everyday-life actions involving the grasping of four different objects (i.e., a bottle, a cup, a glass, and a spray cleaner). All actions were performed by a woman model (aged 31 years) and with the same effector (right hand). Depending on the kinematics (precision vs. power grips), each object could be grasped to perform either one of two possible actions. For instance, in the case of the object “bottle”, the two possible actions were a) to pour and b) to place, each of them performed with the correspondent kinematics: reaching-to-grasp and pour using a power grip and reaching-to-grasp and lift using a precision grip (44). Actions were shot in two different contextual settings: congruent and incongruent. In the congruent condition, the action suggested by the context was compatible with the action suggested by the movement kinematics (i.e., reaching-to-grasp an open bottle located near an empty glass with a power grip). Conversely, in the incongruent condition, the context interfered with the perception of the movement kinematics by cueing to the opposite action (i.e., reaching-to-grasp a bottle with a cork located near a glass already full of water with a power grip). For a complete description of objects, action labels, grip types, contexts, and their possible combinations, please refer to (25, 30). Importantly, stimuli were previously validated (25, 30), confirming the appropriate manipulation of action plausibility, with actions performed in incongruent contexts judged as less plausible than those performed in congruent ones. Moreover, to ensure that motor modulations were triggered by our contextual manipulation and not by differences in the movement kinematic profiles of the same action across scenes, we performed a frame-based analysis, which showed that kinematics were comparable across contexts (25, 30).

Based on previous studies manipulating the spatial frequency content of images (4), we used a Gaussian blur filter with a 19 pixel kernel for low-pass filtering (resulting in images low-pass

filtered at  $\approx 6$  cycles per image, LSF) and the high-pass filter application set to a radius of 0.3 pixels for the high-pass filtering (resulting in images filtered at  $\approx 30$  cycles per image, HSF). The mean contrast values were matched across images. Furthermore, average stimulus luminance did not differ according to image type (mean luminance for Intact: 115.56, for HSF: 117.56 and for LSF: 116.46 on a 256 gray-level scale;  $X^2 = 1.75, p = 0.41$ ); neither did it differ between congruent and incongruent conditions (mean: 116.57 and 116.79, respectively;  $X^2 = 0.50, p = 0.47$ ). A total of 48 stimuli were created (4 action pairs embedded in 2 different contexts, presented in their original intact form or containing only low-spatial or high-spatial frequency information). Each image was randomly presented 4 times for a total of 192 trials.

**Procedure.** Each participant was tested in a single experimental session lasting approximately  $\sim 60$  min. They sat in a comfortable armchair in a dimly lit room  $\sim 1$  m away from a 24 inch monitor (resolution:  $1920 \times 1080$  pixels, refresh frequency: 60 Hz). Images appeared at the center of the screen on a neutral background and subtended approximately  $15.96^\circ \times 11.97^\circ$  of visual angle. Before starting the experiment, participants were familiarized with intact action video examples. More specifically, they observed broad-band grayscale videos of the eight different actions performed in either congruent or incongruent contexts. This procedure was aimed at facilitating the recognition of contextual and grip kinematics cues as well as their relation, rather than at familiarizing with the spatial frequency content of the stimuli.

Trials started with a visual warning cue lasting for 5 s (the Italian word “*attendi*”, in English “wait”) and it was followed by image presentation. Then a frame with the verbal descriptors of the two possible intentions (e.g., “*versare*” and “*spostare*”, in English “to pour” and “to place”, respectively; one located on the left and the other on the right) written in black on a white background was presented. The location of the two descriptors was counterbalanced, ensuring that in half of the trials one of the descriptors was presented on the left side of the screen and, in the other half, it was presented on the right. This procedure enabled us to prevent participants from

planning their response in advance on the basis of the descriptor spatial location. This frame remained on the screen until a response was recorded.

In the Experiment 1 and the Supplementary control Experiment 1, participants provided their responses by pressing with the index finger the computer keys “z” (for left choices) or “m” (for right choices). The response keys were covered with white stickers in order to facilitate localizing their position on the keyboard (a QWERTY keyboard was used). In the Experiment 2 and the Supplementary control Experiment 2 the descriptors were located one up and one down in a counterbalanced order during the task. Participants were requested to verbalize their responses (by saying “*su*” or “*giù*”, in English “up” or “down”, respectively) to prevent that peripheral muscular contraction artifacts resulting from button press contaminated MEPs. Importantly, verbal responses were required only after the TMS pulse was delivered, thus reducing the possibility of interfering with the hand MEPs (45). Even though verbal responses in the TMS experiment were not further analyzed, we requested them in order to control that participants were paying attention to the stimuli and engaged in the task. TMS pulses were delivered at stimulus offset. The interval between TMS pulses was at least 8,500 ms to avoid cumulative effects of the stimulation. Stimulus-presentation timing, EMG recording, and TMS triggering, as well as randomization of stimuli in a block, were controlled using E-prime V2 software (Psychology Software Tools Inc., Pittsburgh, PA) running on a PC.

**Electromyography (EMG) recording and TMS.** In Experiment 2, single-pulse TMS was applied to the left M1 using a Magstim 200 stimulator (maximum output = 2 T at coil surface, pulse duration = 250  $\mu$ sec, rise time = 60  $\mu$ sec; The Magstim Company, Carmarthenshire, Wales, UK) connected to a 70-mm figure-of-eight coil (Magstim polyurethane-coated coil). Motor-evoked potentials (MEPs) were recorded simultaneously from the FDI and from the ECR of the right hand. Surface Ag/AgCl disposable electrodes (1 cm diameter) were placed in a belly-tendon montage for



each muscle. The EMG signal was amplified, filtered (band-pass 5 Hz to 20 kHz) and recorded with the Biopac MP-36 system (BIOPAC Systems, Inc., Goleta, CA) at a sampling rate of 50 kHz.

The coil was positioned tangentially on the scalp, with the handle pointing backward and approximately 45° lateral from the midline, perpendicular to the line of the central sulcus (46). This orientation was chosen based on the finding that the lowest motor threshold is achieved when the induced electric current in the brain is flowing perpendicular to the central sulcus (47, 48). The optimal scalp position (OSP) for inducing MEPs in the right FDI and ECR muscles was detected by moving the coil in 1-cm steps over the left M1 and by delivering TMS pulses at constant intensity until the largest MEPs for both muscles were found. Then, the position was marked with a pen on a tight-fitting bathing cap worn by participants. The coil was held on the scalp by a coil holder with an articulated arm, and its position with respect to the mark was checked continuously to compensate for small movements of the participants' head during data collection.

The TMS intensity during the recording session was adjusted to 120% of the motor threshold at rest (rMT), which is defined as the minimum intensity able to evoke MEPs with  $\geq 50 \mu\text{V}$  peak-to-peak amplitude in the higher threshold muscle (namely, the ECR), in 5 out of 10 consecutive pulses (49). The rMT ranged from 36% to 48% ( $M = 41.12\%$ ,  $SD = 4.03\%$ ) in Experiment 2 and from 35% to 54% in the Supplementary control Experiment 2 ( $M = 44.55\%$ ,  $SD = 6.48\%$ ) of the maximum stimulator output. To ensure that there was no unwanted background EMG activity before the magnetic pulse, the signal from both muscles was continuously monitored, and when voluntary contractions were detected, stimulus presentation was suspended, and participants were encouraged to fully relax their muscles. Furthermore, before starting the experiment, participants received acoustic (via loudspeakers) and visual (via oscilloscope) feedback of their muscle activity in order to help them to get relaxed. MEPs' peak-to-peak amplitudes (in millivolts) were collected and stored in a computer for off-line analysis.

**Data Analysis.** Percent of correct responses (accuracy) was calculated for each individual and experimental condition of the behavioral experiments. Latencies of responses were not considered since they were prompted only after picture/video offset in order to ensure comparable stimulus exposure across conditions. Individual mean peak-to-peak (in mV) amplitudes of MEPs recorded from the FDI and ECR muscles were calculated separately for each condition. Since background EMG is known to modulate MEP amplitude, it was assessed in each participant by calculating the mean rectified EMG signal across a 100ms interval prior to TMS. MEPs with preceding background EMG deviating from the mean by  $> 2$  SD were removed from further analysis. Trials were also excluded if the peak-to-peak amplitude of the MEP was  $\pm 2$  SDs beyond the participant's mean calculated in each experimental condition. The total percentage of excluded MEPs (15%) was not significantly different across conditions and muscles (all  $F$ s  $< 2.4$ , all  $P$ s  $> 0.05$ ).

In all RM-ANOVAs, post-hoc analysis was carried out using the Duncan test correction for multiple comparisons. Estimates of the effect size were obtained using the partial eta-squared for main effects and interactions. For post-hoc contrasts we report Cohen's  $d$  effects. The  $\alpha$  value for all statistical tests was set at 0.05. All frequentist tests were implemented in Statistica software v.10 (Statsoft, Tulsa, OK). The Bayesian RM-ANOVA was performed using the JASP software (50). Finally, correlation analyses were computed using the Robust Correlation Toolbox (51) implemented in MatlabR2012B.

### **Data Availability**

The data that support the findings of this study are available in the Dryad Digital Repository (doi:10.5061/dryad.h44j0zpgh).

### **Acknowledgments**

This work was supported by grants from the European Commission (MCSA-H2020-NBUCA, grant N. 656881; to L.A.), from the Italian Ministry of University and Research (PRIN 2017, Prot.

2017N7WCLP; to C.U.), and from the Italian Ministry of Health (Ricerca Corrente 2020, Scientific Institute, IRCCS E. Medea; to A.F.)

## References

1. Goodale MA & Milner AD (1992) Separate visual pathways for perception and action. *Trends in neurosciences* 15(1):20-25.
2. Goodale MA, Milner AD, Jakobson LS, & Carey DP (1991) A neurological dissociation between perceiving objects and grasping them. *Nature* 349(6305):154-156.
3. Kilner JM (2011) More than one pathway to action understanding. *Trends in cognitive sciences* 15(8):352-357.
4. Bar M, *et al.* (2006) Top-down facilitation of visual recognition. *Proceedings of the National Academy of Sciences of the United States of America* 103(2):449-454.
5. Kveraga K, Boshyan J, & Bar M (2007) Magnocellular projections as the trigger of top-down facilitation in recognition. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 27(48):13232-13240.
6. Denison RN, Vu AT, Yacoub E, Feinberg DA, & Silver MA (2014) Functional mapping of the magnocellular and parvocellular subdivisions of human LGN. *NeuroImage* 102 Pt 2:358-369.
7. Merigan WH & Maunsell JH (1993) How parallel are the primate visual pathways? *Annual review of neuroscience* 16:369-402.
8. Vuilleumier P, Armony JL, Driver J, & Dolan RJ (2003) Distinct spatial frequency sensitivities for processing faces and emotional expressions. *Nature neuroscience* 6(6):624-631.
9. Peyrin C, *et al.* (2010) The neural substrates and timing of top-down processes during coarse-to-fine categorization of visual scenes: a combined fMRI and ERP study. *Journal of cognitive neuroscience* 22(12):2768-2780.
10. Peyrin C, *et al.* (2005) Hemispheric specialization of human inferior temporal cortex during coarse-to-fine and fine-to-coarse analysis of natural visual scenes. *NeuroImage* 28(2):464-473.
11. Berman D, Golomb JD, & Walther DB (2017) Scene content is predominantly conveyed by high spatial frequencies in scene-selective visual cortex. *PloS one* 12(12):e0189828.
12. Morrison DJ & Schyns PG (2001) Usage of spatial scales for the categorization of faces, objects, and scenes. *Psychonomic bulletin & review* 8(3):454-469.
13. Amoruso L, Finisguerra, A. (2019) Low or High-Level Motor Coding? The Role of Stimulus Complexity. *Frontiers in human neuroscience* 13:332.
14. Amoruso L, Finisguerra A, & Urgesi C (2018) Contextualizing action observation in the predictive brain: Causal contributions of prefrontal and middle temporal areas. *NeuroImage* 177:68-78.
15. Bruni S, Giorgetti V, Bonini L, & Fogassi L (2015) Processing and Integration of Contextual Information in Monkey Ventrolateral Prefrontal Neurons during Selection and Execution of Goal-Directed Manipulative Actions. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 35(34):11877-11890.
16. Borra E, Gerbella M, Rozzi S, & Luppino G (2011) Anatomical evidence for the involvement of the macaque ventrolateral prefrontal area 12r in controlling goal-directed actions. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 31(34):12351-12363.
17. Battaglia-Mayer A & Caminiti R (2019) Corticocortical Systems Underlying High-Order Motor Control. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 39(23):4404-4421.
18. Urgesi C, *et al.* (2010) Simulating the future of actions in the human corticospinal system. *Cerebral cortex* 20(11):2511-2521.
19. Dakin S & Frith U (2005) Vagaries of visual perception in autism. *Neuron* 48(3):497-507.

20. Sutherland A & Crewther DP (2010) Magnocellular visual evoked potential delay with high autism spectrum quotient yields a neural mechanism for altered perception. *Brain : a journal of neurology* 133(Pt 7):2089-2097.
21. Baron-Cohen S, Wheelwright S, Skinner R, Martin J, & Clubley E (2001) The autism-spectrum quotient (AQ): evidence from Asperger syndrome/high-functioning autism, males and females, scientists and mathematicians. *Journal of autism and developmental disorders* 31(1):5-17.
22. Ruta L, Mazzone D, Mazzone L, Wheelwright S, & Baron-Cohen S (2012) The Autism-Spectrum Quotient--Italian version: a cross-cultural confirmation of the broader autism phenotype. *Journal of autism and developmental disorders* 42(4):625-633.
23. Fadiga L, Fogassi L, Pavesi G, & Rizzolatti G (1995) Motor facilitation during action observation: a magnetic stimulation study. *Journal of neurophysiology* 73(6):2608-2611.
24. van den Bergh D, van Doorn, J., Marsman, M., Draws, T., van Kesteren, E.-J., Derks, K., Dablander, F., Gronau, Q. F., Kucharsky, S., Komarlu Narendra Gupta, A. R., Sarafoglou, A., Voelkel, J. G., Stefan, A., Ly, A., Hinne, M., Matzke, D., & Wagenmakers, E.-J. (2020) A tutorial on conducting and interpreting a Bayesian ANOVA in JASP. *L'Année Psychologique/Topics in Cognitive Psychology*.
25. Amoruso L, Finisguerra A, & Urgesi C (2016) Tracking the Time Course of Top-Down Contextual Effects on Motor Responses during Action Comprehension. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 36(46):11590-11600.
26. Federico P & Perez MA (2017) Distinct Corticocortical Contributions to Human Precision and Power Grip. *Cerebral cortex* 27(11):5070-5082.
27. Hilt PM, et al. (2017) Action observation effects reflect the modular organization of the human motor system. *Cortex; a journal devoted to the study of the nervous system and behavior* 95:104-118.
28. Fluet MC, Baumann MA, & Scherberger H (2010) Context-specific grasp movement representation in macaque ventral premotor cortex. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 30(45):15175-15184.
29. Pistohl T, Schulze-Bonhage A, Aertsen A, Mehring C, & Ball T (2012) Decoding natural grasp types from human ECoG. *NeuroImage* 59(1):248-260.
30. Amoruso L & Urgesi C (2016) Contextual modulation of motor resonance during the observation of everyday actions. *NeuroImage*.
31. Urgesi C, Candidi M, Fabbro F, Romani M, & Aglioti SM (2006) Motor facilitation during action observation: topographic mapping of the target muscle and influence of the onlooker's posture. *The European journal of neuroscience* 23(9):2522-2530.
32. Urgesi C, Moro V, Candidi M, & Aglioti SM (2006) Mapping implied body actions in the human motor system. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 26(30):7942-7949.
33. Livi A, et al. (2019) Agent-based representations of objects and actions in the monkey pre-supplementary motor area. *Proceedings of the National Academy of Sciences of the United States of America* 116(7):2691-2700.
34. Eimer M & Schlaghecken F (2003) Response facilitation and inhibition in subliminal priming. *Biological psychology* 64(1-2):7-26.
35. Eimer M & Schlaghecken F (1998) Effects of masked stimuli on motor activation: behavioral and electrophysiological evidence. *Journal of experimental psychology. Human perception and performance* 24(6):1737-1747.
36. Mattiassi AD, Mele S, Ticini LF, & Urgesi C (2014) Conscious and unconscious representations of observed actions in the human motor system. *Journal of cognitive neuroscience* 26(9):2028-2041.
37. Tal A & Bar M (2014) The proactive brain and the fate of dead hypotheses. *Frontiers in computational neuroscience* 8:138.
38. Bar M (2007) The proactive brain: using analogies and associations to generate predictions. *Trends in cognitive sciences* 11(7):280-289.
39. Bruni S, Giorgetti V, Fogassi L, & Bonini L (2017) Multimodal Encoding of Goal-Directed Actions in Monkey Ventral Premotor Grasping Neurons. *Cerebral cortex* 27(1):522-533.

40. Briggs GG & Nebes RD (1975) Patterns of hand preference in a student population. *Cortex; a journal devoted to the study of the nervous system and behavior* 11(3):230-238.
41. Rossi S, Hallett M, Rossini PM, & Pascual-Leone A (2009) Safety, ethical considerations, and application guidelines for the use of transcranial magnetic stimulation in clinical practice and research. *Clinical neurophysiology : official journal of the International Federation of Clinical Neurophysiology* 120(12):2008-2039.
42. Faul F, Erdfelder E, Buchner A, & Lang AG (2009) Statistical power analyses using G\*Power 3.1: Tests for correlation and regression analyses. *Behavior research methods* 41(4):1149-1160.
43. Amoroso L, Finisguerra A, & Urgesi C (2018) Autistic traits predict poor integration between top-down contextual expectations and movement kinematics during action observation. *Scientific reports* 8(1):16208.
44. De Stefani E, Innocenti A, Bernardi NF, Campione GC, & Gentilucci M (2012) The bottle and the glass say to me: "pour!". *Experimental brain research* 218(4):539-549.
45. Gentilucci M, Bernardis P, Crisi G, & Dalla Volta R (2006) Repetitive transcranial magnetic stimulation of Broca's area affects verbal responses to gesture observation. *Journal of cognitive neuroscience* 18(7):1059-1074.
46. Di Lazzaro V, et al. (1998) Comparison of descending volleys evoked by transcranial magnetic and electric stimulation in conscious humans. *Electroencephalogr Clin Neurophysiol* 109(5):397-401.
47. Brasil-Neto JP, et al. (1992) Optimal focal transcranial magnetic activation of the human motor cortex: effects of coil orientation, shape of the induced current pulse, and stimulus intensity. *J Clin Neurophysiol* 9(1):132-136.
48. Mills KR, Boniface SJ, & Schubert M (1992) Magnetic brain stimulation with a double coil: the importance of coil orientation. *Electroencephalogr Clin Neurophysiol* 85(1):17-21.
49. Rossini PM, et al. (1994) Non-invasive electrical and magnetic stimulation of the brain, spinal cord and roots: basic principles and procedures for routine clinical application. Report of an IFCN committee. *Electroencephalogr Clin Neurophysiol* 91(2):79-92.
50. Team J (2019) JASP (Version 0.11.1)[Computer software].
51. Pernet CR, Wilcox R, & Rousselet GA (2012) Robust correlation analyses: false positive and power validation using a new open source matlab toolbox. *Frontiers in psychology* 3:606.
52. Wetzels R, et al. (2011) Statistical Evidence in Experimental Psychology: An Empirical Comparison Using 855 t Tests. *Perspectives on psychological science : a journal of the Association for Psychological Science* 6(3):291-298.

## Figure Legends

**Fig.1. Examples of Stimuli and Experimental Task.** (A) Stimuli depicted ongoing but incomplete everyday actions performed in congruent or incongruent scenarios using different grips. Stimuli could be presented either in their intact form or only containing low- (LSF) or high-spatial frequency (HSF) information. Congruency was manipulated in terms of compatibility between observed kinematics (i.e., precision vs. power grips) and the motor intention suggested by the context (i.e., to drink or to clean). (B) In a two-alternative forced choice task (2AFC), participants

observed the snapshots or videos, depending on the experiment, and were requested to predict action unfolding by choosing one of 2 possible overarching intentions. Each trial began with the word “wait” on the screen for 5000ms, followed by the picture/video. Stimuli duration varied, such that half of the pictures were displayed for 300ms, and the other half for 500ms in a random order. After picture/video presentation, a frame with the verbal descriptors of the overarching intention remained on the screen until a response was recorded.

**Fig. 2. Behavioral Results.** Participants' performance in predicting the course of the observed actions under the different spatial frequency conditions (i.e., intact, low- and high-spatial frequency) during the observation of (A) static pictures and (B) dynamic action videos in congruent and incongruent contexts. Asterisks indicate significant comparison ( $p < 0.05$ ). Error bars represent SEM.

**Fig. 3. Correlation between Perceptual Traits and Recognition Performance.** Relationship between perceptual traits and behavioral results indicating that the more participants were able to identify congruent actions when only featural information (i.e., HSF) was available, the higher their local processing style.

**Fig. 4. MEP Results.** Boxplots showing individual data points for MEP amplitudes recorded from the FDI and ECR muscles during the action observation in the different experimental conditions. Amplitudes are expressed as the difference between the spatially filtered conditions (HSF and LSF) and the intact condition (broad-band images). Asterisks indicate significant comparison ( $p < 0.05$ ). End points of the whiskers represent the minimum and maximum value in the data.

**Fig. 5. Correlation between Perceptual Traits and MEPs.** Relationship between perceptual traits and neurophysiological results indicating that the higher the participant's local processing style, the lower the level of motor facilitation for LSF-based actions observed in typical contexts.

**Fig. 6. Supplementary MEP Analysis.** Boxplots showing individual data points for MEP amplitudes recorded from the FDI and ECR muscles during the observation of action videos in an independent sample of participants. Bayes Factors (BF) quantify the strength of empirical one-tailed evidence, with values between 1-3 providing anecdotal evidence and values between 3-10 providing substantial evidence for the alternative hypothesis (52).

**Fig. 7. 3-way Model for Action Comprehension.** The human Action Observation Network (AON), including the superior temporal sulcus, the superior parietal and the premotor areas is shown in grey arrows. The ventral pathway connecting occipital with temporal areas is shown in light-blue arrows. The red solid arrow represents the magnocellular route linking occipital areas with PFC structures where top-down predictions (i.e., initial guesses) are formed based on LSF information. Red dotted arrows back-projecting from PFC to temporal regions depict the shortcut previously suggested by Bar and colleagues (4, 5, 28) for object recognition. Red dotted arrows back-projecting from PFC to premotor regions depict the route we proposed for the comprehension of naturalistic context-embedded actions.

**Table 1.** Mean (SD) scores and range for the AQ subscales in the different experiments.

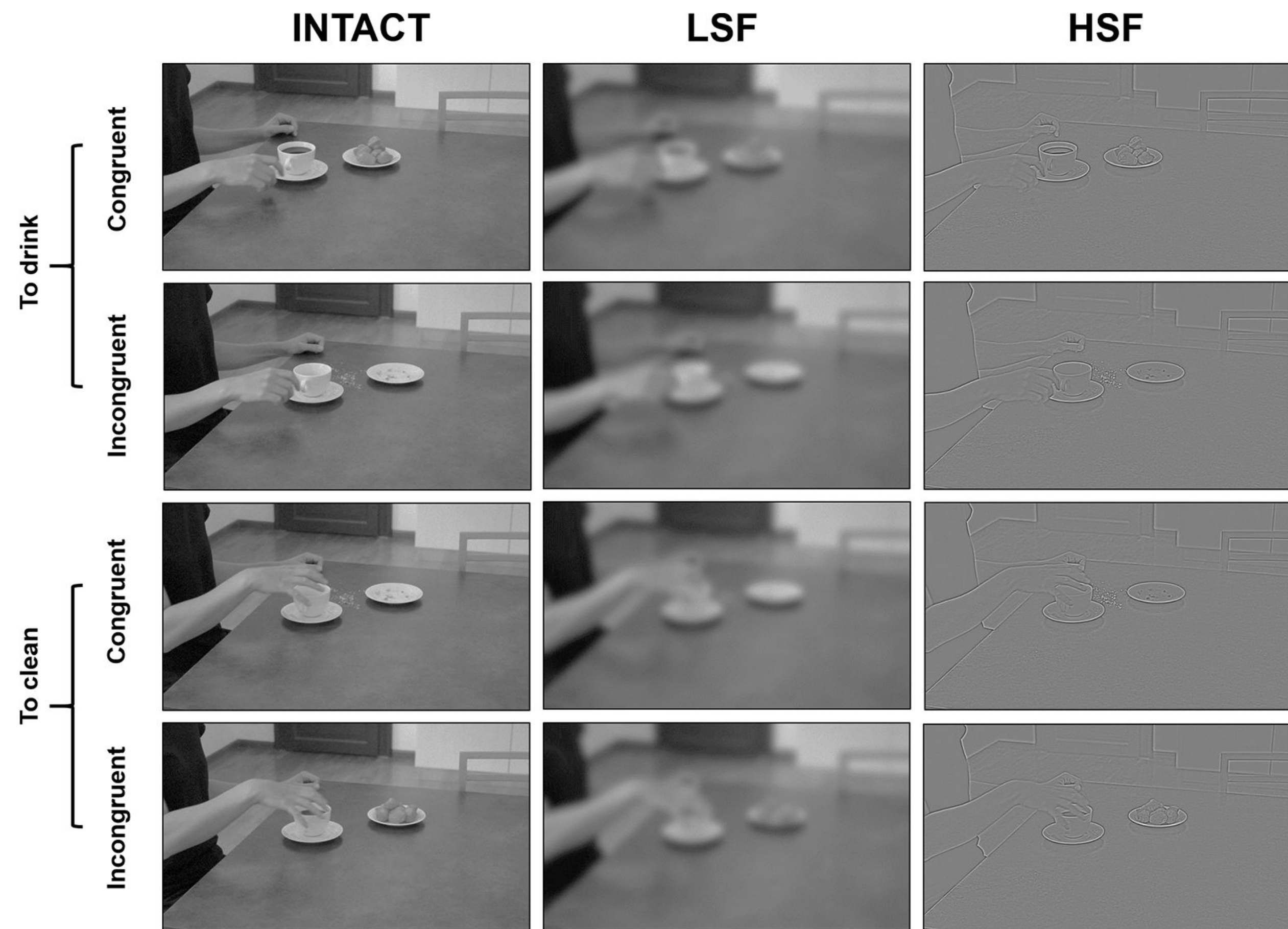
	<b>Experiment 1</b>		<b>Experiment 2</b>	
	Mean (SD)	Range	Mean (SD)	Range
<b>AQ total</b>	18.25 (4.87)	7-25	14.93 (5.84)	5-26
<b>Social skills</b>	2.75 (1.52)	0-5	2 (1.67)	0-6
<b>Attention switching</b>	4.93 (1.94)	2-9	4.37 (2.33)	0-9
<b>Attention to detail</b>	6.18 (2.13)	2-9	4.81 (2.5)	1-8
<b>Communication</b>	2.5 (2)	0-6	1.93 (2.01)	0-6
<b>Imagination</b>	1.87 (1.02)	0-4	1.81 (1.68)	0-6

**Table 2.** Mean raw MEP amplitudes for each experimental condition.

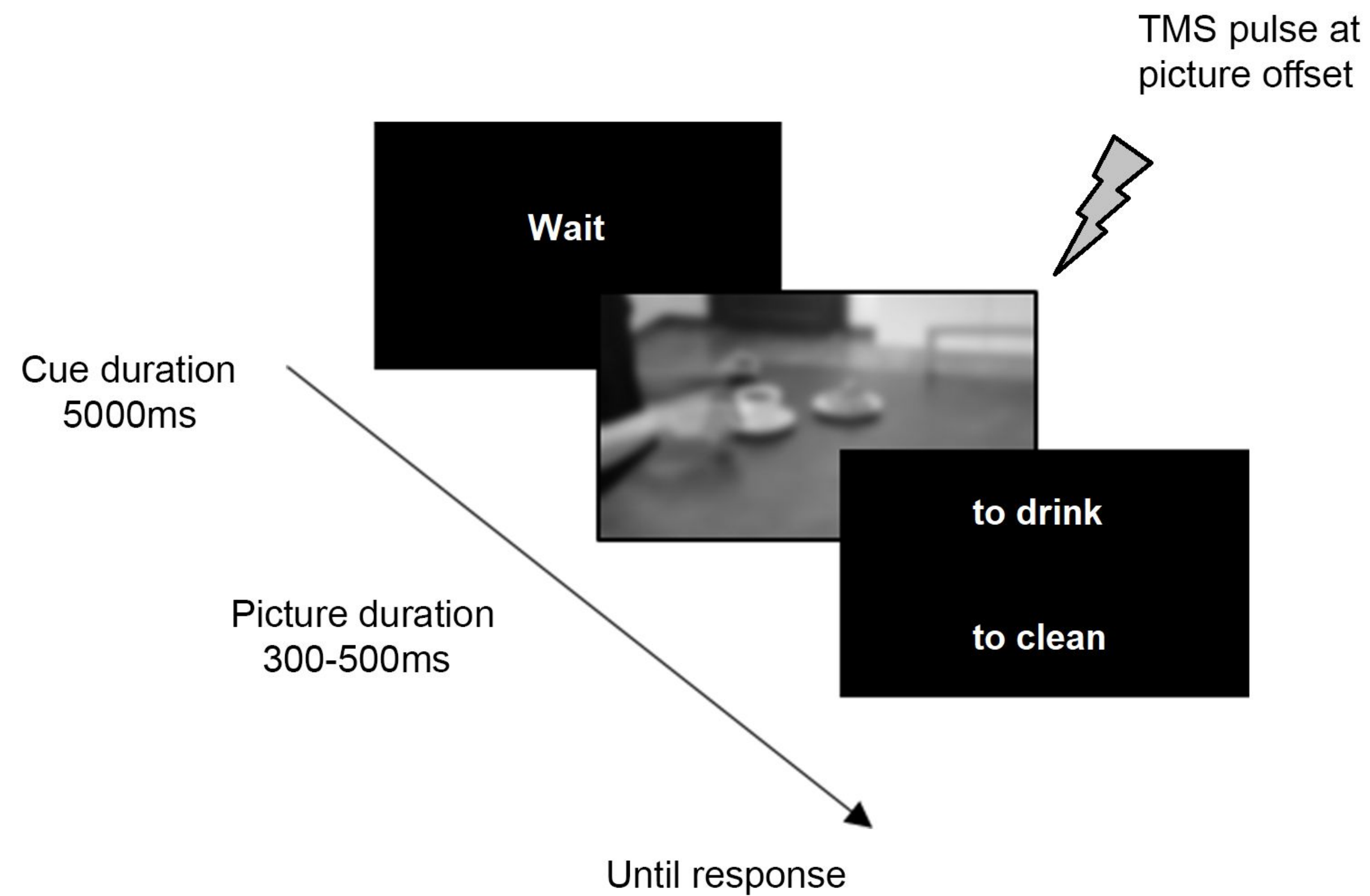
		<b>FDI</b>		<b>ECR</b>	
		<b>Congruent</b>	<b>Incongruent</b>	<b>Congruent</b>	<b>Incongruent</b>
	Intact	0.947±0.26	0.912±0.23	0.499±0.09	0.503±0.09
<b>Precision grip</b>	HSF	1.006±0.28	0.855±0.20	0.519±0.11	0.494±0.10
	LSF	0.819±0.21	1.027±0.27	0.518±0.10	0.528±0.10
	Intact	0.891±0.21	0.934±0.25	0.497±0.09	0.519±0.10
<b>Power grip</b>	HSF	0.894±0.21	0.858±0.21	0.509±0.10	0.511±0.09
	LSF	0.913±0.24	0.810±0.17	0.519±0.10	0.471±0.08
<b>Baseline</b>	Block 1	0.52±0.12		0.39±0.08	
	Block 2	0.51±0.14		0.35±0.07	



(A)

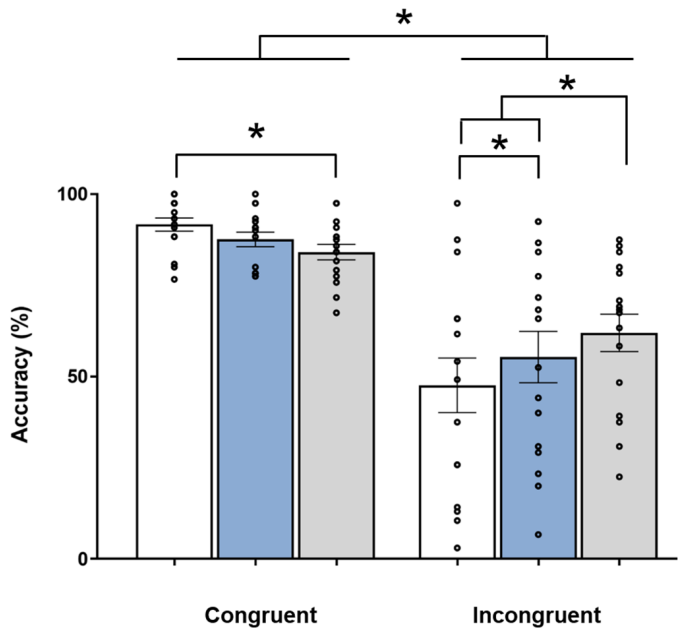


(B)

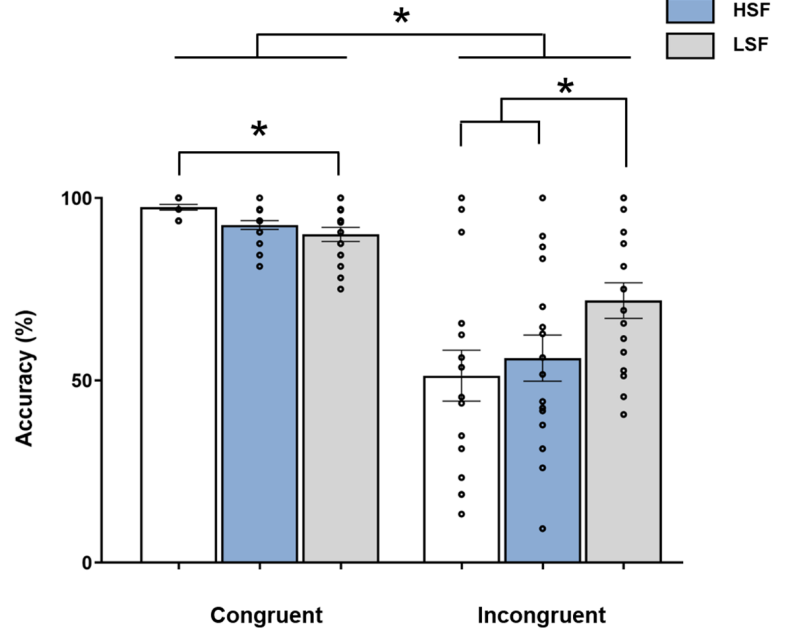


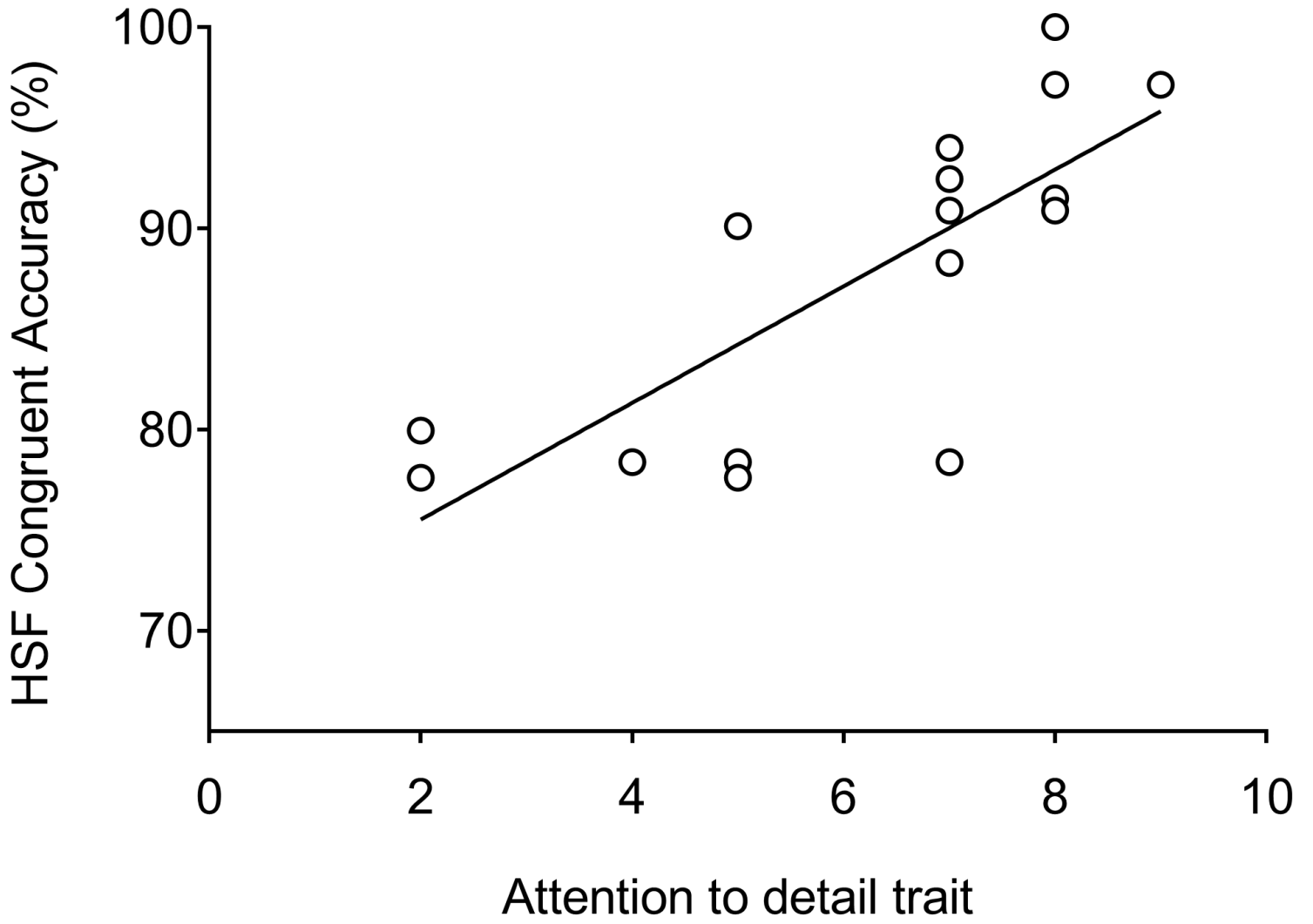
Context x Spatial Frequency

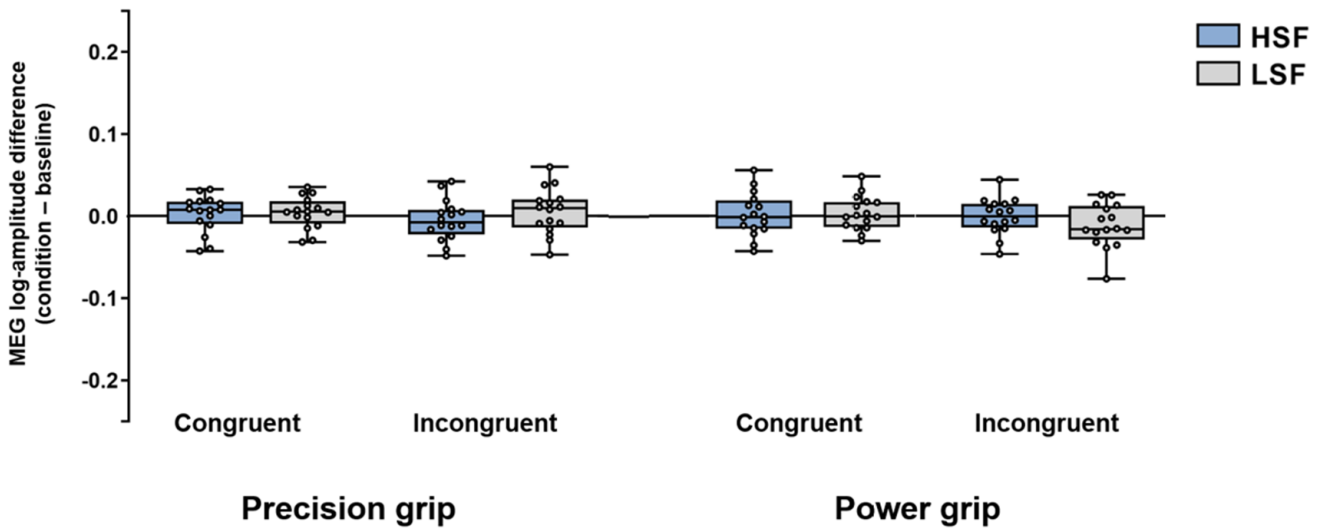
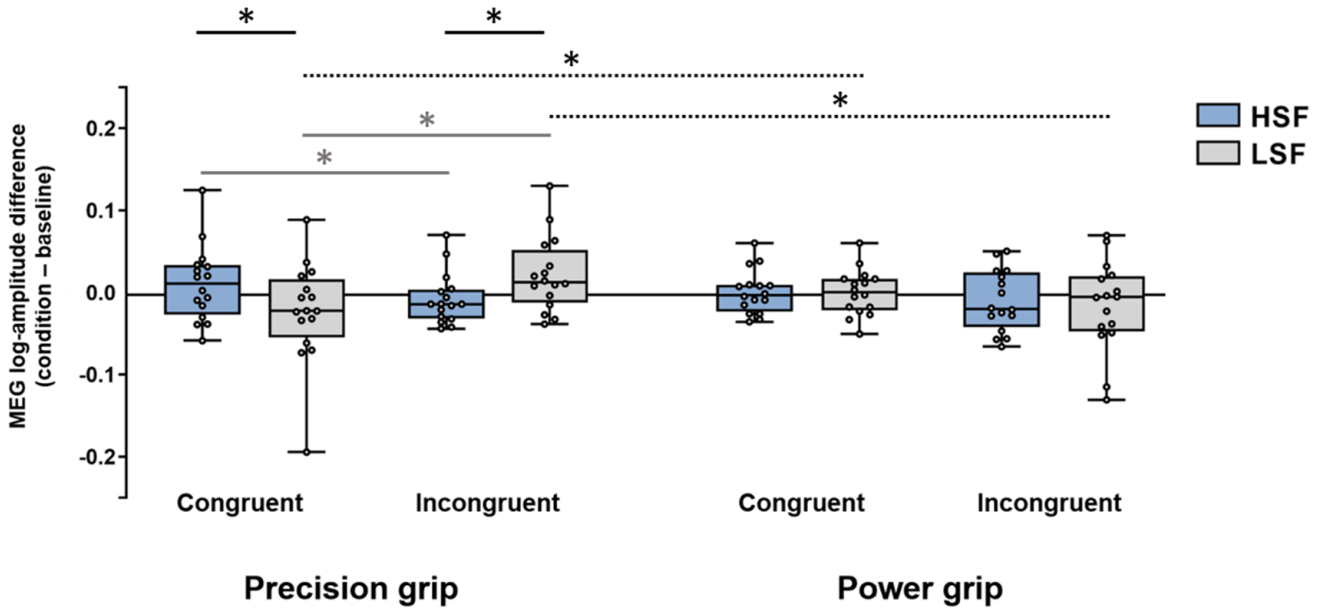
(A) Static Pictures

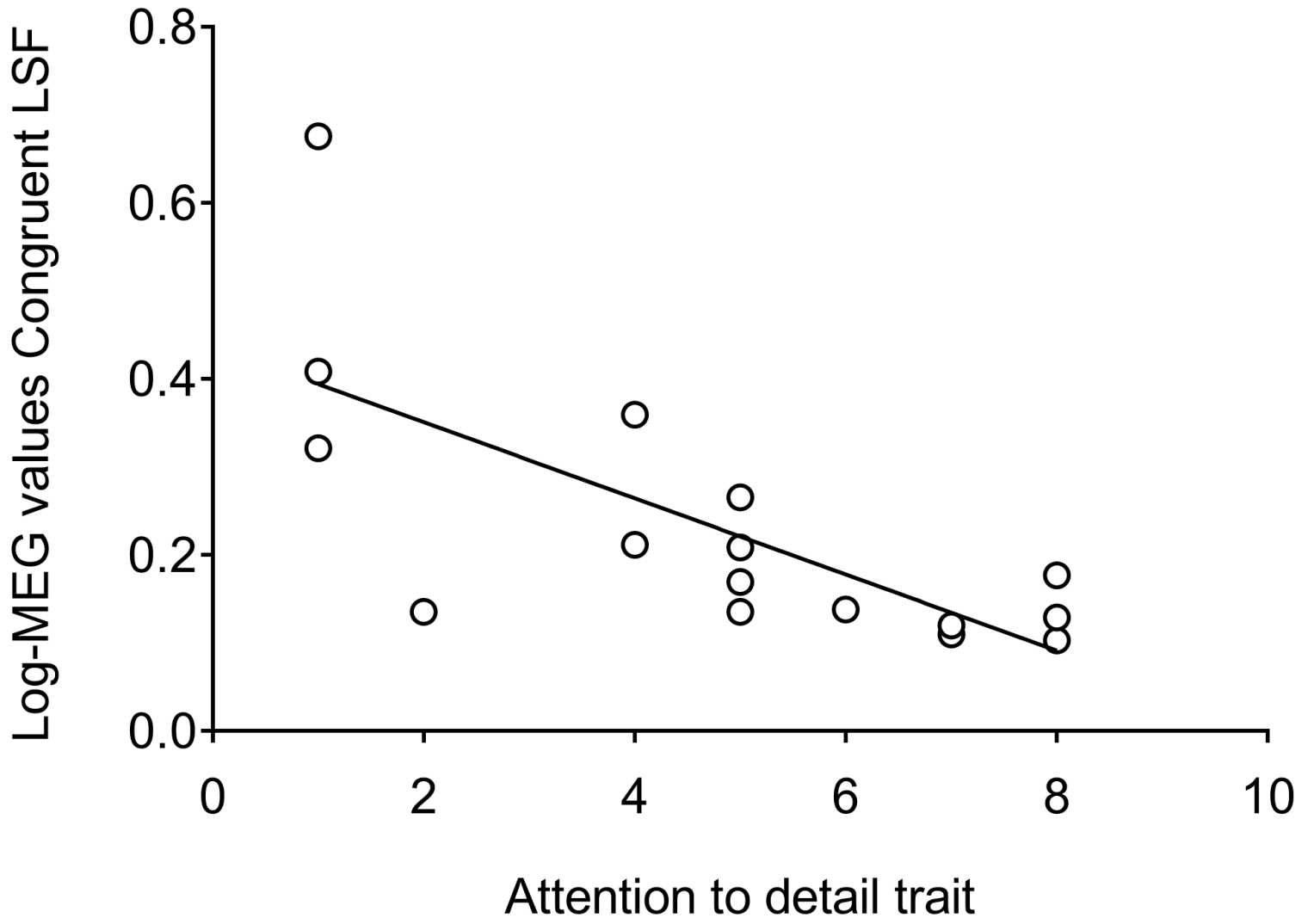


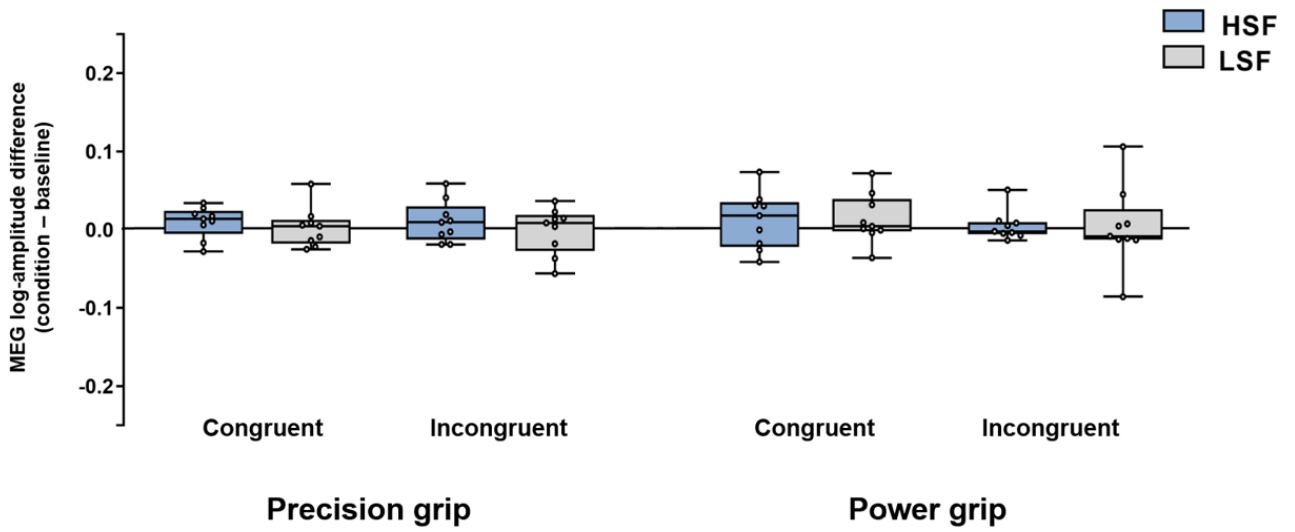
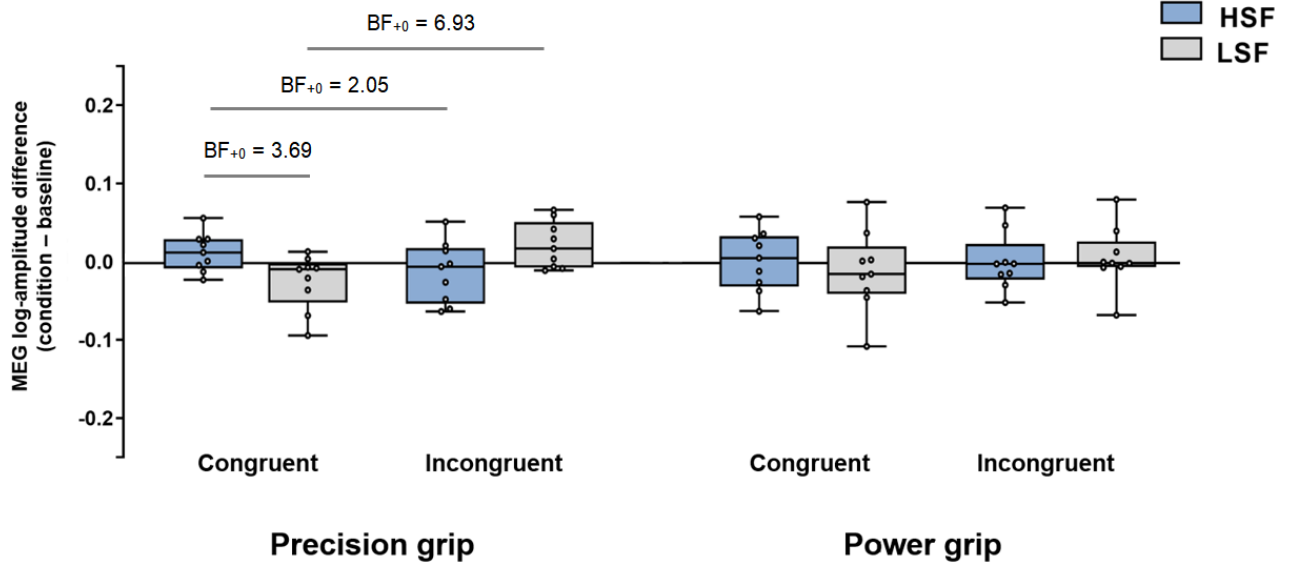
(B) Dynamic Videos

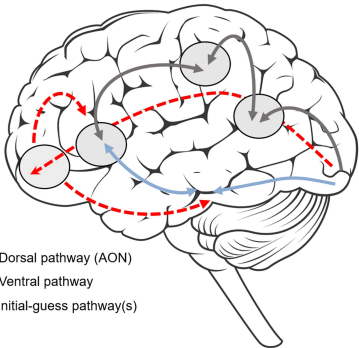












- ▶ Dorsal pathway (AON)
- ▶ Ventral pathway
- - -▶ Initial-guess pathway(s)