

Pilot experiment

A *Pilot experiment* was performed to validate video stimuli, through three different steps. First, a sample of 20 participants was asked to rate, according to different dimensions, an initial set of experimental stimuli. This set of experimental stimuli comprised videos representing actions (eat, move) directed toward different types of objects (sweets, fruits) of different dimensions (small objects: bigné and strawberry; big objects: krapfen and apple) embedded in congruent, incongruent, and neutral contexts. Then, on a different sample of 20 participants, we tested the effects of contextual congruency on behavioural performance in an action prediction task. All participants provided written informed consent before taking part in the study. In light of the COVID-19 pandemic, the task sessions were completed remotely. Participants installed E-Prime 2 software subject station (Psychology Software Tools, Inc., Pittsburgh, PA, United States) on their own PCs and received instructions and assistance over the phone. Last, a kinematic analysis of the most significant videoframes was performed to assess the differences in the kinematics profiles for the two action intentions depending on the object size (big, small) and type (sweet, fruit). In the pilot experiment, videos were interrupted at three different time points, revealing progressively reliable kinematics.

Video stimuli

In addition to the stimuli used in the main experiment, showing the actor reaching and grasping a big (e.g., krapfen) or a small (i.e., cream puffs) sweet, here we also used videos showing the actor reaching and grasping a big (i.e., apple) or a small (strawberries) fruit object, again placed on a container and embedded in a breakfast-table context. Overall, a total of 96 videos (Object Type [2] x Object Size [2] x Action [2] x Context [3] x Actor [2] x Versions [2]) were originally recorded.

Supplementary Table 1 provides a full list of the combinations between objects, actions, grips and contexts which were manipulated during video recording.

| Object | Object size | Action | Grip | Congruent | Incongruent | Neutral |
|-------------|-------------|--------|-----------------|---------------|---------------|------------|
| Krapfen | Big | Eat | Whole-hand grip | Set table | Cleared table | Tray alone |
| | | Move | Precision grip | Cleared table | Set table | Tray alone |
| Cream puffs | Small | Eat | Precision grip | Set table | Cleared table | Tray alone |
| | | Move | Whole-hand grip | Cleared table | Set table | Tray alone |
| Apple | Big | Eat | Whole-hand grip | Set table | Cleared table | Tray alone |
| | | Move | Precision grip | Cleared table | Set table | Tray alone |
| Strawberry | Small | Eat | Precision grip | Set table | Cleared table | Tray alone |
| | | Move | Whole-hand grip | Cleared table | Set table | Tray alone |

Supplementary Table 1: List of object types and sizes, actions, prehension grips and contextual settings.

Stimuli rating task

20 participants (15 F; mean age = 21.8 years; SD = 2.33) were asked to watch all videos till the 23rd video-frame (i.e., seven frames before the contact with the object) and to evaluate the degree of “action plausibility”

and of “action positivity” by means of a Visual Analogue Scale (VAS) ranging from 0 to 100. Action plausibility referred to the strength of the association between a given action and the surrounding context, while action positivity refers to the valence of the observed action in each context. The rationale behind this preliminary step was to confirm the plausibility of the object-directed actions within each context and to control for any relation between the contextual manipulation and positive emotional factors. The VAS scores associated to action plausibility and action positivity collected in the stimuli validation study were entered into separate 2 (Object type: Sweet, Fruit) x 2 (Object size: Big, Small) x 3 (Context: Congruent, Incongruent, Neutral) x 2 (Action: Eat, Move) RM-ANOVAs.

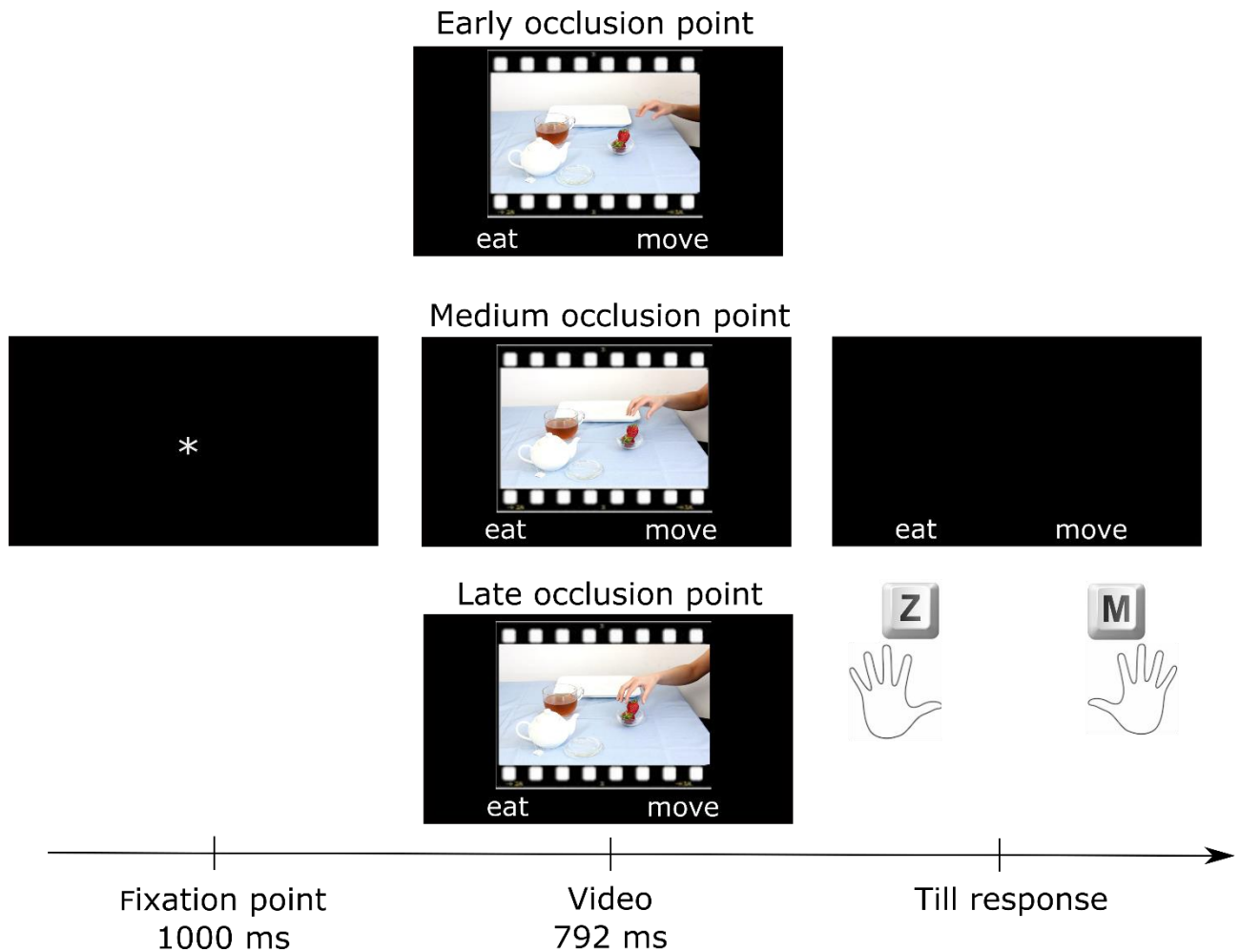
Action prediction task

20 participants (19 F; mean age = 22.3 years; SD = 4.70) volunteered for this study. As in the main experiment, we used a two-alternative forced choice (2AFC) task, in which participants were instructed to observe the videos and predict which one of the two possible intentions (i.e., to eat or to move) drove the actor’s movements. Participants performed two repetitions of the task in two different daily sessions, remotely. At the beginning of the first session, they received experimental instructions on the phone. They were asked to stay in a quiet room and to seat at 60 cm from their computer screen (refresh rate 60, Hz). The size of the videos was adapted to fill approximately a 11° x 6° area at the centre of the monitor for all computer screens.

Each trial started with a white fixation point at the centre of the screen lasting for 1,000 ms, followed by the presentation of the experimental video. For video presentation, each frame was presented for 2 refresh cycles (i.e., 33.33 ms). The videos were interrupted at an early (frame 13), middle (frame 18) or late (frame 23) action phase. As in the main experiment, to match the duration of the videos across the different interruption points. The duration of the first frame was lengthened to 400, 233.33, or 66.67 ms, for early, middle or late interruption time, respectively. This video duration was 800 ms for all interruption times. The response prompt consisted of two verbal descriptors at the left and right bottom of the screen, showing the two possible action alternatives (“mangia”, i.e., to eat and “sposta”, i.e., to move). The response prompt was already present at the onset of the video and remained on the screen until participants’ response, thus allowing them to respond as soon as they were able to guess the prediction. Responses were provided by pressing the key (Z) or (M) on the keyboard, corresponding to the left or right location of the action alternatives. The location of the descriptors remained constant between sessions of each participant, but it was counterbalanced among participants. The participant’s response signalled the end of the trial. A schematic representation of the task is provided in **Supplementary Figure 1**. Each session consisted of two identical blocks of 288 trials each for a total of 576 trials, lasting around 35 minutes. For each block, videos were randomly presented. For the final analysis, trials from the two sessions were merged, for a total of 1152 trials (32 trials per cell).

Data of the 2AFC task was treated as in the main experiment. After using Kolmogorov–Smirnov for checking normality in the distribution of the data, the d' and c values were entered into a 2 (Object type: Sweet, Fruit) x 3 (Context: Congruent, Incongruent, Neutral) x 3 (Interruption point: Late, Middle, Early) RM ANOVA. For all analyses, estimates of the effect size were obtained using the partial η squared (η_p^2 , Cohen, 1988) for

ANOVA effects. Post-hoc pairwise comparisons were carried out using the Duncan's test to correct for multiple comparisons and analyse significant interaction effects.



Supplementary Figure 1: Schematic representation of the action prediction task

Kinematics analysis

To show the different kinematics profiles for the two types of actions performed towards big (krapfen and apple) vs. small (cream puff and strawberry) objects, we measured specific kinematic parameters from the actor's right hand by using a dedicated software for motion analysis (Kinovea® version 0.8.15, Kinovea Inc). Specifically, for each of the 96 videos, we selected four video-frames (8, 13, 18, 23) at four different time-points of the entire motor act resulting in a very early, early, middle and late actions phases. In the very early action phase (frame 8), the actor appeared still, making kinematics fully uninformative. Then, in the early action phase (frame 13), the actor was at the very beginning of the reaching movement, with kinematics being slightly informative. Later, in the middle action phase (frame 18), the pre-shaping of the grasp movement was

more visible, thus kinematics was still ambiguous but more informative than the previous one. Lastly, the late action phase (frame 23) showed the actor's hand approaching the object with kinematics cues clearly signalling the action intention. The extracted kinematic parameters for each selected frame were the following: *thumb-index distance*, considered as the segment between the fingertips of the thumb and of the index finger; *thumb-little finger distance*, considered as the segment connecting the fingertips of the thumb and of the little finger; *wrist-table distance*, considered as the segment between the radial styloid process and the table surface. We selected these measures to detect the different degree of thumb opponency movement of the index and little fingers during precision and whole-hand grips and the different height of the reaching-to-grasp movements for the eating and moving actions. The extracted kinematics measurements were entered into separate RM-ANOVA with Object Size (Big, Small) and Action (Eat, Move) as categorical factors and Frame (8, 13, 18, 23) as dependent factors. For each cell of design, we considered as the error term the variability of 24 measures, considering the videos of 2 possible actors, each performing 2 repetitions of the same action on two different object types (i.e., fruits, cakes) in 3 different contexts (congruent, incongruent, neutral). **Supplementary Figure 3** provides an example for the extracted kinematics parameters.

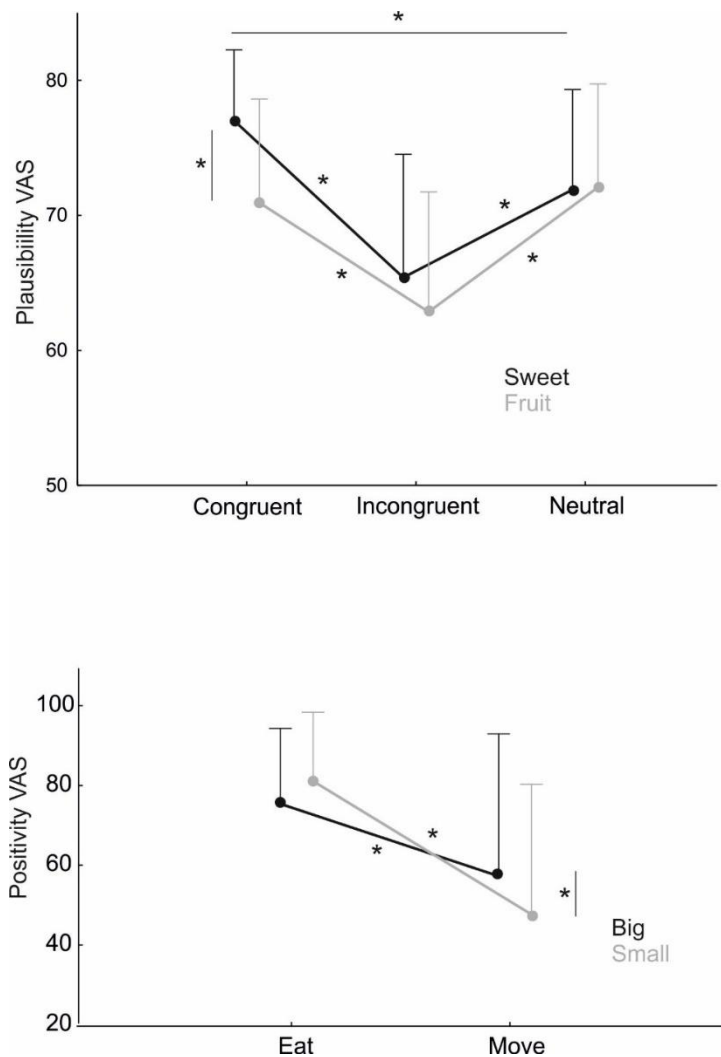
Results

Stimuli rating task results

For what concerns action plausibility, the RM-ANOVA yielded significant main effects of Context ($F_{2,38} = 5.60$, $p = 0.007$, $\eta_p^2 = 0.23$), Action ($F_{1,19} = 15.63$, $p < 0.001$, $\eta_p^2 = 0.45$), which were qualified by two-way interactions between Object type x Context ($F_{2,38} = 5.73$, $p = 0.006$, $\eta_p^2 = 0.23$), Object size x Context ($F_{2,38} = 3.50$, $p = 0.040$, $\eta_p^2 = 0.16$) and Object size x Action ($F_{1,19} = 4.75$, $p = 0.042$, $\eta_p^2 = 0.20$). Post-hoc analysis of the Object type x Context interaction showed that for the sweet objects, actions embedded in congruent contexts were more plausible than those embedded in neutral and incongruent contexts ($p < 0.001$); furthermore, actions embedded in incongruent contexts were less plausible than actions embedded in neutral contexts ($p < 0.001$). Instead, for the fruit objects, both congruent and neutral contexts were more plausible than incongruent contexts (all $ps < 0.001$) but did not differ with each other ($p = 0.419$). Furthermore, in the case of congruent contexts, plausibility was higher for sweet than fruit objects ($p < 0.001$), while no differences emerged either for incongruent ($p = 0.061$) or neutral contexts ($p = 0.897$), suggesting a stronger contextual effect for sweet than for fruit objects. Post-hoc analysis of the Object size x Context interaction showed that plausibility was higher for big than small objects in the case of congruent (big: 74.981 ± 3.056 ; small: 72.981 ± 3.326 ; $p = 0.043$) and neutral (big: 74.050 ± 3.677 ; small: 69.928 ± 3.912 ; $p < 0.001$) contexts, while no differences emerged for incongruent contexts (big: 64.5 ± 4.387 , small: 63.747 ± 4.576 , $p = 0.414$). Further, for the big objects, both congruent and neutral contexts were more plausible than incongruent contexts (all $ps < 0.001$) but did not differ between each other ($p = 0.313$). Instead, for the small objects, congruent contexts were more plausible than neutral and incongruent context (all $ps < 0.002$) and incongruent contexts were less plausible than neutral contexts ($p < 0.001$).

Last, post-hoc analysis of the Object size x Action interaction showed that, although the eating action was generally more plausible than the moving action for both the small and the big objects, the moving action was more plausible for big (65.087 ± 4.958) than small objects (56.962 ± 5.248 , $p = 0.045$), while no differences emerged for the eating action (big: 77.267 ± 3.330 , small: 80.808 ± 3.288 , $p = 0.361$).

For what concerns valence positivity, the RM-ANOVA yielded significant main effect of Action ($F_{1,19} = 13.59$, $p = 0.002$, $\eta_p^2 = 0.42$), qualified by an Object size x Action interaction ($F_{1,19} = 6.04$, $p = 0.023$, $\eta_p^2 = 0.24$). Duncan post-hoc comparisons showed that, similarly to the results on plausibility, the eating action was more positive than the moving action directed toward either the big (74.371 ± 3.633 vs. 56.479 ± 6.627 ; $p < 0.001$) or the small object (80.042 ± 3.357 vs. 47.300 ± 6.029 ; $p < 0.001$). Furthermore, the valence of the moving action was higher for big than small objects ($p = 0.044$), while no between-object differences emerged for the eating action ($p = 0.200$). All other effects were not significant (all $F_s < 3.03$, $p > 0.060$, $\eta_p^2 = 0.24$, $\eta_p^2 < 0.14$), suggesting that contextual manipulation led to a modulation of plausibility, but not of positivity of the stimuli.



Supplementary Figure 2: VAS scores upper panel: plausibility for actions directed towards sweet and fruit objects embedded in congruent, incongruent, and neutral contexts; lower panel: positivity for actions directed

towards big and small objects Asterisks indicate significant comparisons ($p < 0.05$). Error bars represent SEM.

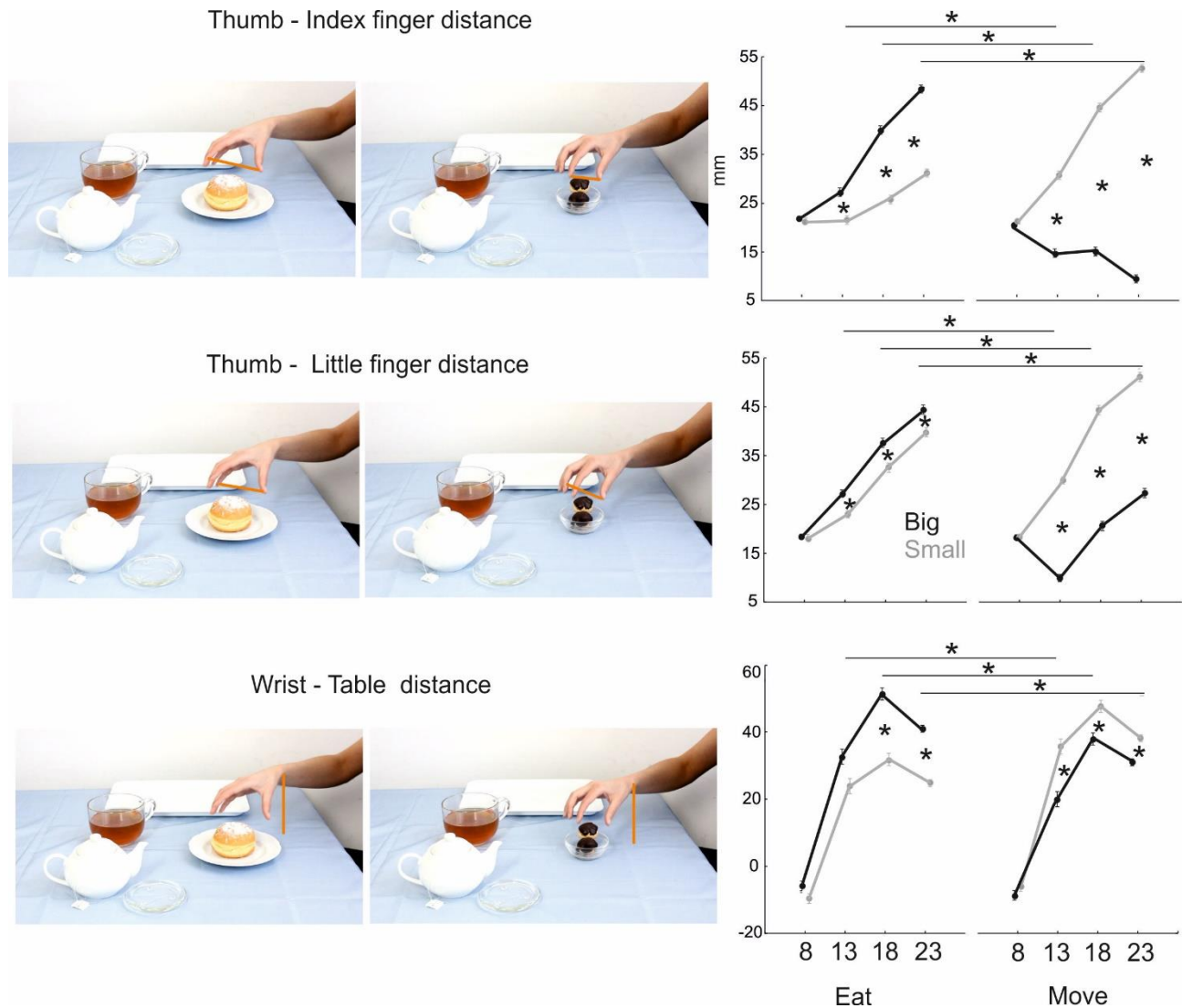
Kinematics analysis results

The RM-ANOVA on the *thumb-index finger distance* (**Supplementary Figure 3**) showed significant effects of Object size ($F_{1,92} = 70.38$, $p < 0.001$, $\eta_p^2 = 0.43$), Action ($F_{1,92} = 13.85$, $p < 0.001$, $\eta_p^2 = 0.14$), Object size x Action interaction ($F_{1,92} = 591.09$, $p < 0.001$, $\eta_p^2 = 0.86$), Frame ($F_{3,276} = 637.09$, $p < 0.001$, $\eta_p^2 = 0.87$), Frame x Object size ($F_{3,276} = 99.92$, $p < 0.001$, $\eta_p^2 = 0.52$) and Frame x Action ($F_{3,276} = 46.13$, $p < 0.001$, $\eta_p^2 = 0.33$) interactions, all qualified by a three-way Frame x Object size x Action interaction ($F_{3,276} = 586.05$, $p < 0.001$, $\eta_p^2 = 0.86$). Duncan post-hoc analysis of this interaction showed that, at frame 8, the thumb-index finger distance was not distinguishable between the two object sizes for either action (all $ps > 0.291$). At later action phases, for the eating action, it was progressively smaller (indicating greater thumb opponency) for the small than the big object (all $ps < 0.001$), while for the moving action it was progressively greater for the small than the big objects (all $ps < 0.001$). Further, focusing on the difference between actions, at frames 13, 18 and 23 the thumb-index finger distance was smaller for the moving than the eating actions directed to the big object (all $p < 0.001$), and smaller for the eating than the moving actions directed to the small objects (all $p < 0.001$). No differences emerged at frame 8 ($ps > 0.509$).

The RM-ANOVA on the *thumb-little finger distance* (**Supplementary Figure 3**) showed significant effects of Object size ($F_{1,92} = 114.19$, $p < 0.001$, $\eta_p^2 = 0.55$), Action ($F_{1,92} = 18.93$, $p < 0.001$, $\eta_p^2 = 0.17$), Object size x Action interaction ($F_{1,92} = 241.07$, $p < 0.001$, $\eta_p^2 = 0.72$), Frame ($F_{3,276} = 1288.49$, $p < 0.001$, $\eta_p^2 = 0.93$), Frame x Object size interaction ($F_{3,276} = 66.93$, $p = 0.001$, $\eta_p^2 = 0.42$), Frame x Action interaction ($F_{3,276} = 14.78$, $p = 0.001$, $\eta_p^2 = 0.14$), all qualified by a significant three-way Object size x Action x Frame interaction ($F_{3,276} = 132.98$, $p < 0.001$, $\eta_p^2 = 0.59$). Post-hoc on this interaction showed that, for what concerns the eating action, the thumb-little finger distance was higher for big than small objects at frame 13, 18 and 23 (all $ps < 0.001$). No differences emerged at frame 8 ($p = 0.930$). For what concerns the moving action, the thumb-little finger distance was higher for small than big objects at frames 13, 18 and 23 (all $ps < 0.001$). No differences emerged at frame 8 ($p = 0.990$). Further, focusing on the difference between actions, at frames 13, 18 and 23, for big objects, the thumb-little finger distance was higher for the eating than the moving actions directed to big objects (all $ps < 0.001$), while it was higher for the moving than the eating actions directed toward small objects (all $ps < 0.001$). No differences emerged at frame 8 ($p > 0.959$).

The RM-ANOVA on the *wrist-table distance* (**Supplementary Figure 3**) showed significant effects of Object size x Action interaction ($F_{1,92} = 34.05$, $p < 0.001$, $\eta_p^2 = 0.27$), Frame ($F_{3,276} = 1382.75$, $p < 0.001$, $\eta_p^2 = 0.94$), Frame x Object size interaction ($F_{3,276} = 13.67$, $p < 0.001$, $\eta_p^2 = 0.13$), all qualified by a three-way Frame x Object size x Action interaction ($F_{3,276} = 17.59$, $p < 0.001$, $\eta_p^2 = 0.16$). Post-hoc on this interaction showed that for what concerns the eating action, the wrist-table distance was higher for big than small objects at frames 18 and 23 (all $ps < 0.001$). No differences emerged at frames 8 ($p = 0.557$) and 13 ($p = 0.144$). For what concerns the moving action, the wrist-table distance was smaller for big than small objects at frames 13, 18 and 23 (all

ps ≤ 0.011). No differences emerged at frame 8 (p = 0.532). Further, focusing on the difference between actions, at frames 13, 18 and 23, the wrist-table distance was higher for the eating than the moving actions directed to big objects (all ps < 0.001), while it was higher for the moving than the eating actions directed to small objects (all ps < 0.001). No differences emerged at frame 8 (p > 0.211).



Supplementary Figure 3 Left panel: Example of kinematic parameters extracted at frame 23 for the thumb-index finger distance, the thumb-little finger distance, and the wrist-table distance. Right panel: deployment of the different kinematics profiles at the selected frames (8, 13, 18, 23) for the big and small objects for eating and moving actions. Asterisks indicate significant comparisons ($p < 0.05$). Error bars represent SEM.

Behavioural results

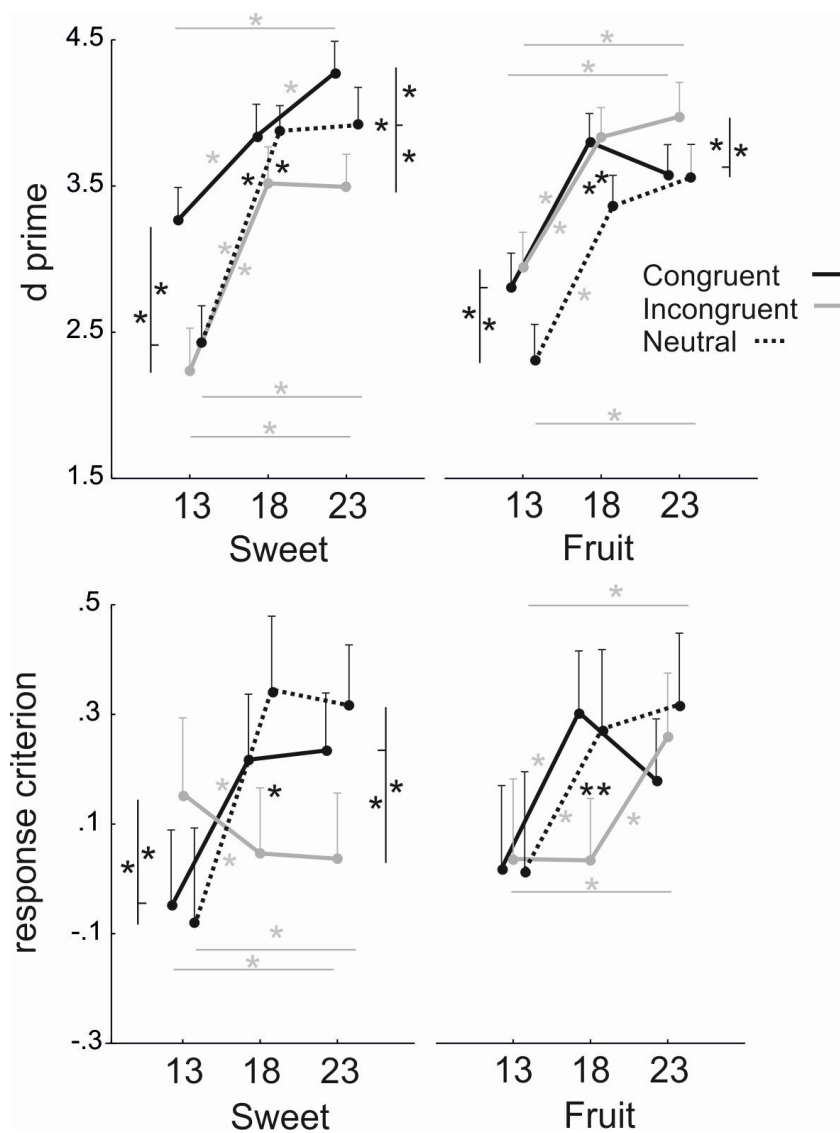
The RM-ANOVA on the d' values (**Supplementary Figure 4**) yielded significant main effects of Context ($F_{2,38} = 11.62$, $p < 0.001$, $\eta_p^2 = 0.38$), Interruption point ($F_{2,38} = 42.61$, $p < 0.001$, $\eta_p^2 = 0.69$), significant two-way interactions between Object type and Context ($F_{2,38} = 21.77$, $p < 0.001$, $\eta_p^2 = 0.53$), Context and Interruption point ($F_{4,76} = 4.17$, $p = 0.004$, $\eta_p^2 = 0.18$), all qualified by a significant three-way Object type x Context x Interruption point interaction ($F_{4,76} = 3.57$, $p = 0.010$, $\eta_p^2 = 0.16$).

Duncan Post-hoc comparisons showed that, for what concerns the sweet objects, d' scores at the early interruption point were higher for congruent ($3.269 \pm .222$) than incongruent ($2.236 \pm .292$, $p < 0.001$) and neutral contexts ($2.427 \pm .254$, $p < 0.001$), which did not differ from each other ($p = .212$); at the middle interruption point, d' scores were lower for incongruent ($3.518 \pm .250$) than congruent ($3.841 \pm .219$, $p = 0.049$) and neutral contexts ($3.877 \pm .173$, $p = 0.030$), which did not differ from each other ($p = 0.114$); at the late interruption point, d' scores were higher for congruent ($4.274 \pm .216$) than neutral ($3.917 \pm .258$, $p = 0.019$) and incongruent contexts ($3.494 \pm .223$, $p < 0.001$), and were lower for incongruent than neutral contexts ($p < 0.001$). Further, d' scores were always higher for middle and late than early interruption points independently from context (all $ps < 0.009$). Conversely, they were higher for late than middle interruption points for congruent contexts only ($p = 0.006$), while no differences emerged for incongruent and neutral contexts (all $ps > 0.777$).

For what concerns the fruit objects, d' scores at the early interruption point were lower for neutral (2.293 ± 0.247) than congruent ($2.809 \pm .232$, $p < 0.001$) and incongruent contexts ($2.937 \pm .247$, $p < 0.001$), which did not differ from each other ($p = 0.372$); at the middle interruption point, they were lower for neutral ($3.363 \pm .210$) than congruent ($3.802 \pm .195$, $p = 0.007$) and incongruent contexts ($3.834 \pm .202$, $p = 0.004$), which did not differ from each other ($p = 0.820$); at the late interruption point, they were higher for incongruent ($3.975 \pm .234$) than congruent ($3.582 \pm .201$, $p = 0.017$) and neutral contexts ($3.559 \pm .226$, $p = 0.012$), which did not differ from each other ($p = 0.870$). Further, d' scores were always higher for middle and late than early interruption points independently from context (all $ps < 0.001$), while no differences emerged between middle and late interruption points for either context (all $ps > 0.128$).

The RM-ANOVA on response criterion (**Supplementary Figure 4**) yielded a significant main effects of Context ($F_{2,38} = 4.02$, $p = 0.026$, $\eta_p^2 = 0.17$) and Interruption point ($F_{2,38} = 6.66$, $p = 0.003$, $\eta_p^2 = 0.26$) as well as significant two-way interaction between Context and Interruption point ($F_{4,76} = 8.50$, $p < 0.001$, $\eta_p^2 = 0.31$), qualified by a significant three-way interaction between Object type, Context and Interruption point ($F_{4,76} = 2.74$, $p = 0.034$, $\eta_p^2 = 0.13$). Duncan Post-hoc comparisons showed that, for what concerns the sweet objects, the response criterion at the early interruption point was higher for incongruent (0.155 ± 0.138) than congruent (-0.051 ± 0.140 , $p = 0.044$) and neutral contexts (-0.086 ± 0.179 , $p = 0.019$), which did not differ from each other ($p = 0.687$); at the middle interruption point, it was lower for incongruent (0.046 ± 0.120) than neutral contexts (0.345 ± 0.134 , $p = 0.004$), while no differences emerged with congruent contexts (0.217 ± 0.119 , all $ps > 0.076$); at the late interruption point, it was lower for incongruent (0.036 ± 0.129) than congruent (0.234 ± 0.105 , $p = 0.048$) and neutral contexts (0.316 ± 0.110 , $p = 0.006$), which did not differ from each other ($p = 0.409$). Further, for congruent contexts response criterion was lower for early than middle ($p = 0.009$) and late ($p = 0.005$) interruption points, which did not differ from each other ($p = 0.847$); for neutral contexts it was lower for early than middle ($p < 0.001$) and late ($p < 0.001$) interruption points, which did not differ from each other ($p = 0.761$); for incongruent contexts, no differences emerged among interruption points (all $ps > 0.202$).

For what concerns the fruit objects, the response criterion at the early interruption point did not differ between contexts (congruent: 0.019 ± 0.151 , incongruent 0.036 ± 0.146 , $p = 0.417$, neutral: 0.016 ± 0.180 , all $p_s > 0.184$); at the middle interruption point, it was lower for incongruent (0.033 ± 0.120) than congruent (0.304 ± 0.112 , $p = 0.009$) and neutral contexts (0.273 ± 0.145 , $p = 0.021$), which did not differ from each other ($p = 0.728$); at the late interruption point, the response criterion did not differ between contexts (congruent: 0.180 ± 0.112 , incongruent 0.259 ± 0.116 , $p = 0.853$, neutral: 0.318 ± 0.130 , all $p_s > 0.833$). Further, for congruent contexts, response criterion was lower at early rather than middle interruption points ($p = 0.006$), while the other differences were not significant (all $p_s > 0.115$); for incongruent contexts it was higher for late than early ($p = 0.029$) and middle ($p = 0.029$) interruption points, which did not differ from each other ($p = 0.976$); for neutral contexts it was lower at early rather than middle ($p = 0.014$) and late ($p = 0.004$) interruption points, which did not differ from each other ($p = 0.646$).



Supplementary Figure 4: Behavioral results. Upper panel: participants' sensitivity in predicting the action towards sweet or fruit objects for the three contexts (congruent: black line, incongruent: grey line, neutral: dashed line) at the three interruption points (early: frame 13, middle: frame 18, late: frame 23). Lower panel:

participant's response criterion for the same conditions. Data points represent group averages. Asterisks indicate significant comparisons ($p < 0.05$) between interruption points (grey) and contexts (black). Error bars represent SEM.

Discussion

The present pilot study was aimed at validating and selecting the experimental stimuli used in the main experiment by behavioral evaluations and kinematic analysis of videos. First, action plausibility results showed that, in congruent contexts, the observed actions were more plausible for the sweet than the fruit objects. Importantly, the modulation of action plausibility according to contextual congruency was not associated to a parallel modulation of the positivity evaluation of the stimuli, suggesting that contextual effects could not be ascribed to emotional valence. The action positivity results revealed that, independently from context, the action to eat was considered more plausible and positive than the action to move, probably because the eating action has more relevance for survival purposes. In a similar vein, the moving action was considered more plausible and positive when directed to the big than the small objects, while no difference emerged according to object size for the eating action.

Results of the action prediction task suggested that, while increasing with later interruption time across conditions, only for sweet objects and congruent context performance was the highest for the late interruption point, while middle and late interruption points did not differ in any other condition. More importantly, and in line with previous studies ([Amoruso and Urgesi, 2016](#); [Amoruso et al., 2016](#); [Amoruso et al., 2019](#); [Amoruso et al., 2020](#)) contextual cues modulated action prediction, at least for the sweet objects, facilitating, since the early interruption point, the recognition of actions embedded in congruent contexts and worsening, only at later interruptions points, the recognition of those embedded in incongruent contexts. This result is in keeping with previous findings of the temporal deployment of contextual modulation effects on action prediction ([Amoruso et al., 2016](#)), with a facilitation of performance exerted by congruent contexts at earlier phases of action processing, followed by a later inhibition exerted by incongruent contexts. The congruence of contextual information, however, influenced action prediction only for the sweet objects. For the fruit objects, no differences emerged between congruent and incongruent contexts at early and middle interruption points, with performance being even higher for incongruent contexts at the late interruption point. This means that, at least for what concerns the objects used in the present study (i.e., krapfen, cream puff, apple, strawberry), the association between the eating and the moving action with a set- or cleared-table contextual scenario, respectively, was confirmed for sweets, but not for fruits.

Differences between the two objects also emerged in terms of response criterion, which was generally low, revealing no response bias, particularly at the early interruption time (i.e., when non kinematics cues were available). Response criterion, however, tended to become more positive, revealing a tendency to report a moving than eating action, at later interruption time depending on context and object type. Indeed, a greater bias toward a 'to move' response was noted when actions were embedded in congruent or neutral contexts than in incongruent ones, but this occurred at both middle and late interruption times for sweets and only at

the middle interruption time for fruits. At the late interruption time of fruit actions, when participants were better for incongruent contexts, a comparable bias to report a moving action was observed in all contexts.

Thus, the results confirmed that, when observing actions directed to sweet objects, a set table anticipated an eating action and a cleared table a moving action. Conversely, the two scenarios were comparably associated to either eating or moving the fruit objects and facilitated responses compared to neutral contexts, suggesting that their service was considered appropriate either before or after breakfast. In fact, the finding that performance for fruit objects at the late interruption point was even higher for incongruent than congruent contexts might suggest the expectation of having fruit after (and not before) breakfast, in contrast to our manipulation intent.

Kinematics analysis revealed that for all the experimental interruption points the considered parameters clearly differentiated between the two actions as a function of the object size. Although these were progressively enhanced with action unfolding, at the early stage (frame 13), kinematics analysis already revealed differences between conditions. A crucial finding concerned the specific differences between the two object sizes as a function of action. Indeed, the distances between the thumb and the index fingertips, between the thumb and the little fingertips and between the wrist and the table were increasingly higher for big objects than for small objects for the eating action. Instead, these effects were inverted for the moving action. Therefore, the analysis of kinematics parameters for the videos showed a double dissociation between the kinematics associated to the grasping action (whole-hand or precision grip) and the size of the object (big or small). Indeed, higher finger aperture and distance of the arm from the table were observed for eating the big than the small objects and for moving the small compared to the big objects.

According to the Pilot study, for the main experiment we focused on those objects that were more affected by the association with congruent vs. incongruent contexts, namely the sweets, thus removing fruits. Then, to simplify the experimental design for the neurophysiological measurements we removed the neutral contexts, and we focused on the comparisons of motor resonance for actions embedded in congruent and incongruent contexts. Last, considering the evidence that the contextual modulation was similar for the middle and late interruption point, the former was removed.

References

- Amoruso, L., & Urgesi, C. (2016). Contextual modulation of motor resonance during the observation of everyday actions. *NeuroImage*, 134, 74-84.
- Amoruso, L., Finisguerra, A., & Urgesi, C. (2016). Tracking the time course of top-down contextual effects on motor responses during action comprehension. *Journal of neuroscience*, 36(46), 11590-11600.
- 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.

- Amoruso, L., & Finisguerra, A. (2019). Low or high-level motor coding? The role of stimulus complexity. *Frontiers in human neuroscience*, 332.
- Amoruso, L., Narzisi, A., Pinzino, M., Finisguerra, A., Billeci, L., Calderoni, S., ... & Urgesi, C. (2019). Contextual priors do not modulate action prediction in children with autism. *Proceedings of the Royal Society B*, 286(1908), 20191319.
- Amoruso, L., Finisguerra, A., & Urgesi, C. (2020). Spatial frequency tuning of motor responses reveals differential contribution of dorsal and ventral systems to action comprehension. *Proceedings of the National Academy of Sciences*, 117(23), 13151-13161.
- Ansuini, C., Cavallo, A., Bertone, C., & Becchio, C. (2014). The visible face of intention: why kinematics matters. *Frontiers in psychology*, 5, 815.
- 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.
- Becchio, C., Koul, A., Ansuini, C., Bertone, C., & Cavallo, A. (2018). Seeing mental states: An experimental strategy for measuring the observability of other minds. *Physics of life reviews*, 24, 67-80.
- Betti, S., Finisguerra, A., Amoruso, L., & Urgesi, C. (2022). Contextual priors guide perception and motor responses to observed actions. *Cerebral Cortex*, 32(3), 608-625.
- Bianco, V., Finisguerra, A., Betti, S., D'argenio, G., & Urgesi, C. (2020). Autistic traits differently account for context-based predictions of physical and social events. *Brain Sciences*, 10(7), 418.
- Bianco, V., Finisguerra, A., & Urgesi, C. (2022). Updating implicit contextual priors with explicit learning for the prediction of social and physical events. *Brain and Cognition*, 160, 105876.
- Briggs, G. G., & Nebes, R. D. (1975). Patterns of hand preference in a student population. *Cortex*, 11(3), 230-238.
- Cavallo, A., Koul, A., Ansuini, C., Capozzi, F., & Becchio, C. (2016). Decoding intentions from movement kinematics. *Scientific reports*, 6(1), 1-8.
- Cohen, J. (2013). *Statistical power analysis for the behavioral sciences*. Routledge.
- Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G* Power 3.1: Tests for correlation and regression analyses. *Behavior research methods*, 41(4), 1149-1160.