

How task interactivity shapes action observation

L. M. Sacheli^{1,2*}, C. Verga¹, E. Arcangeli¹, G. Banfi^{2,3}, M. Tettamanti⁴, E. Paulesu^{1,2}

¹ Department of Psychology and Milan Center for Neuroscience (NeuroMi), University of Milano-Bicocca, 20126 Milan, Italy.

² IRCCS Istituto Ortopedico *Galeazzi*, 20161 Milan, Italy.

³ Università Vita e Salute San Raffaele, 20132 Milan, Italy.

⁴ Centro Interdipartimentale Mente/Cervello, Università degli Studi di Trento, 38068, Rovereto, Italy.

Running title: Task interactivity shapes action observation

*Corresponding author

Lucia Maria Sacheli

Piazza dell'Ateneo Nuovo 1, 20126 Milan, Italy

+ 39 02 64483776; lucia.sacheli@unimib.it

21 Abstract

22

23 Action observation triggers imitation, a powerful mechanism permitting interpersonal coordination.
24 Coordination, however, also occurs when the partners' actions are non-imitative and physically
25 incongruent. One influential theory postulates that this is achieved via top-down modulation of imitation
26 exerted by prefrontal regions. Here we rather argue that coordination depends on sharing a goal with the
27 interacting partner: this shapes action observation, overriding involuntary imitation, through the
28 predictive activity of the left ventral premotor cortex (lvPMc). During fMRI, participants played music in
29 turn with a virtual partner in interactive and non-interactive conditions requiring 50% of imitative/non-
30 imitative responses. In a full-factorial design, both perceptual features and low-level motor requirements
31 were kept constant throughout the experiment. Behaviorally, the interactive context minimized
32 visuomotor interference due to the involuntary imitation of physically incongruent movements. This was
33 paralleled by modulation of neural activity in the lvPMc, which was specifically recruited during the
34 interactive task independently of the imitative/non-imitative nature of the social exchange. This lvPMc
35 activity reflected the predictive decoding of the partner's actions, as revealed by multivariate pattern
36 analysis. This demonstrates that, during interactions, we process our partners' behavior to prospectively
37 infer their contribution to the shared goal achievement, generating motor predictions for cooperation
38 beyond low-level imitation.

39

40 Keywords

41 Joint action, fMRI, MVPA, ventral premotor cortex, motor prediction.

42

43

44 **Introduction**

45 What tells apart interactive from non-interactive actions? Are the same cognitive mechanisms responsible
46 for coding the observed movements of others, independently of whether we need to coordinate with them?
47 Despite claims on the uniqueness of interaction for human life (Di Paolo et al. 2010), the clear-cut
48 evidence is lacking on what singles out the perception of the actions of an interactive partner, especially
49 when it comes to its neural underpinnings. The ability to coordinate with a partner is one of the earliest
50 achievements of social development: describing the neurophysiological signatures of effective non-verbal
51 motor interactions might thus provide a crucial reference point for studying social cognition in health and
52 pathology with a minimalistic and controlled approach.

53 Vast neurophysiological evidence in human and non-human primates shows that each time we
54 observe an action we *simulate* it in an involuntary imitation with a fronto-parietal “mirror” system
55 (Rizzolatti and Sinigaglia 2016). However, it has been shown that motor simulation (Novembre et al.
56 2014; Hadley et al. 2015) and visuomotor integration (Sacheli et al. 2015; Era et al. 2018) play a key role
57 in online interpersonal coordination also in the case of non-imitative interactions, which fluently occur
58 even when the pursuit of a common goal imposes that the two partners perform physically different
59 actions (Sebanz et al. 2006). Prima facie, this seems to suggest that the “mirror” fronto-parietal network
60 might not be the only relevant neural system entailed in interactions, interactive actions being rather
61 coded by higher-order brain areas that gate a purely imitative neural activity (*Top-down control*
62 hypothesis). This is suggested by experimental evidence of the role of top-down control to perform (non-
63 interactive) actions that differ from an observed one (Campbell and Cunnington 2017; Cross and Iacoboni
64 2014; Brass et al. 2009): these studies showed that prefrontal areas, in their coding of task rules, might
65 inhibit “mirroring” motor simulation when needed. This comes with a performance cost, the so-called
66 *visuomotor interference effect*.

67 Nevertheless, this top-down control hypothesis does not account well for all social exchanges, as
68 the performance cost entailed in active suppression of motor simulation (Cross and Iacoboni 2014)
69 unlikely occurs during cooperative interactions, as shown by the ease whereby they take place in real life.
70 Thus, the key question is still open on which mechanisms allow observed interactive actions gaining a

71 special status, transcending simple "mirror" imitation while not paying the processing price of cognitive
72 control.

73 In a recent behavioral experiment (Sacheli et al. 2018a), we provided evidence for an alternative
74 hypothesis: we showed that the distinctive feature of interactions might be the possibility to integrate
75 motor representations regarding the partner's actions within a *Dyadic Motor Plan* (Sacheli et al. 2018a),
76 i.e., a motor plan representing each partner's contribution to the shared goal achievement. In short, our
77 results (Sacheli et al. 2018a) showed that, during an interactive context, participants do not suffer from
78 visuo-motor interference when required to execute non-imitative responses to a partner's action; instead,
79 they show decay in performance when the partner violates the expectations that the agent holds about the
80 partner's contribution to the achievement of the shared goal. We interpret these results as follows: since
81 when interacting we aim to achieve a shared goal (Butterfill 2011, e.g., playing a melody together),
82 knowledge of the shared goal generates expectations about which contribution the partner will provide
83 (e.g., which notes he/she will play in a turn-taking duet); this allows predicting which action the partner
84 will perform to achieve the shared goal. Such expectations would be represented in the Dyadic Motor
85 Plan, and motor simulation would be only recruited to anticipate the effects of the partner's actions and
86 monitor whether they meet expectations. Within a predictive coding framework (Kilner et al. 2007), this
87 would translate in the recruitment of fronto-parietal areas to anticipate the partners' goal and the effect of
88 his/her action when the shared goal or the strategy to achieve it are still unknown and thus need to be
89 inferred (e.g., when the hitter needs to guess what move the setter will do in a volleyball match); on the
90 contrary, knowledge of the shared goal would lead to signal attenuation (Kaiser and Schutz-Bosbach
91 2018) as long as the partner's action is correctly anticipated within the Dyadic Motor Plan representation.
92 Notably, this theoretical account is compatible with a premotor hypothesis of goal representations (Kilner
93 2011; Umiltà et al. 2001), independently of the imitative or non-imitative nature of the interaction.

94 In the present study, we aimed to address these dilemmas at the neurophysiological level. We
95 capitalized on our novel interactive behavioral paradigm (Sacheli et al. 2018a) to directly compare with
96 functional magnetic resonance imaging (fMRI) the *Top-down control* and the *Dyadic Motor Plan*
97 hypotheses and to characterize the neurophysiological signatures describing how a partner's action is
98 represented as a motor interaction unfolds. Participants observed the partner's action either in the

99 presence or absence of interactive shared goals and during imitative or non-imitative exchanges. In the
100 interactive, *Joint Action* condition, the participants shared with their partner the goal of playing pre-
101 learned four-note melodies by alternating playing one note each through the pointing or grasping gestures.
102 In the control, perceptually-matched, *Non-Interactive* condition, the participants' and their partner's
103 actions and tones were unrelated, and the participants were cued on which pair of notes to play in two
104 consecutive trials, independently of the notes their partner played. Both the Joint Action and the Non-
105 Interactive condition required 50% of imitative and 50% of non-imitative responses to the partner's
106 moves, allowing us to measure the emergence of visuomotor interference in non-imitative trials. Crucial
107 for our aim, this was done in a full factorial design and by keeping constant, throughout the
108 interactive/non-interactive and imitative/non-imitative social contexts, both perceptual features and low-
109 level motor requirements, differently from previous attempts.

110 We had the following expectations. According to the *Dyadic Motor Plan* hypothesis, we
111 expected to replicate our behavioral results on the lack of visuomotor interference for incongruent actions
112 in the Joint Action condition only (Sacheli et al. 2018a). We posited that this would translate in
113 dissociation between the neural correlates of the partner's action observation in the interactive vs. non-
114 interactive context and that the neural correlates of interactivity would be independent of the imitative or
115 non-imitative nature of the motor exchange. We expected the neural correlates of the Non-Interactive
116 condition to be modulated by the congruence between the partner's and the participant's movements: we
117 envisaged a stronger recruitment of brain regions responsible for executive motor control (e.g., prefrontal
118 and right parietal areas, Campbell and Cunnington 2017; Cross and Iacoboni 2014; Brass et al. 2009) in
119 incongruent (i.e., non-imitative) as compared to congruent (i.e., imitative) trials. Such effect of action
120 congruence was not expected in Joint Action, where we envisaged a stronger recruitment of the neural
121 correlates of motor predictive processes (e.g., in the left fronto-parietal network), independently of action
122 congruence. On the contrary, the *Top-down control* hypothesis would be satisfied by the emergence of
123 visuomotor interference effects in non-imitative as compared to imitative trials, independently of task
124 interactivity. At the neurophysiological level, this would translate in the stronger recruitment of the neural
125 regions responsible for executive control in the non-imitative than the imitative condition, during both the
126 Joint Action and Non-Interactive tasks.

127 **Materials and Methods**

128

129 **Participants**

130 Twenty-four participants took part in the study (13 men, 11 females, age range 19-27 years, mean $23.75 \pm$
131 2.02). All participants were right-handed as confirmed by the Edinburgh Handedness Inventory ([Oldfield](#)
132 [1971](#); mean Laterality Quotient 92.5 ± 11.5), reported normal or corrected-to-normal vision, and were
133 naive as to the purpose of the experiment.

134 The experimental protocol was approved by the Ethics Committee of the University of Milano-
135 Bicocca (Italy) and by the Ethics Committee of the IRCCS Galeazzi Orthopaedic Institute (Comitato
136 Etico dell'Ospedale San Raffaele di Milano), where data collection was run. The study was carried out
137 according to the ethical standards of the 1964 Declaration of Helsinki and later amendments. All
138 participants gave their written, informed consent to take part in the study and were debriefed as to the
139 purposes of the study at the end of the experimental procedures. Professional musicians were not
140 recruited.

141

142 **Learning phase (2-4 days before the fMRI session)**

143 Before performing the fMRI experiment (2-4 days before the fMRI session), participants underwent a
144 learning phase aimed to teach them the correct association between color-cues and melodies (Joint Action
145 condition) and pairs of notes (Non-Interactive condition) and to familiarize with the task. This learning
146 phase (about 40 min) was identical to the one described in ([Sacheli et al. 2018a](#)). By design, we planned
147 to test only participants who successfully completed the Learning phase (threshold 80% of accuracy). All
148 participants in the sample passed the threshold.

149

150 **Experimental Design**

151 Stimuli and apparatus. The MRI-compatible response device (BrainTrends Ltd) consisted of a custom-
152 made 5-cm wooden cube placed next to the right hand of the subject lying in the scanner. The cube was
153 connected to a standard device to record behavioral measures in the fMRI scanner (Resonance
154 Technology Inc., Northridge, CA, USA). Touch-times on the cube were recorded by activating buttons (1

155 cm wide), one located on the top and two on the sides of the response-cube. Before each trial, participants
156 positioned their right-hand index finger over a start-button (1 × 1 cm) located 2 cm to the right of the
157 cube. They were instructed to either press the top button on the cube with their index finger (pointing
158 action) or to press the side buttons with their thumb and index fingers (grasping action). Pressing the top
159 button generated a G note (~392 Hz) and pressing the side buttons generated a C note (~261 Hz). The two
160 sounds had the same duration (100 ms). A third, raspberry-like sound (duration 100 ms) was emitted as an
161 error signal. Auditory feedback was provided via headphones.

162 The participants responded to sequences of visual stimuli consisting of a set of pictures depicting a virtual
163 partner in different positions (Fig. 1a): (i) starting-position, (ii) implied-motion posture (depicting the
164 pointing/grasping actions at mid-flight), and (iii) end-position (depicting the end of the pointing/grasping
165 action). The end-position image included a small colored square at the center of the partner's cube that
166 gave the color-cued instructions for playing melody/pair of notes.

167

168 Interactive (Joint Action) vs. Non-Interactive tasks. There were separate sessions (fMRI runs) for the
169 Joint Action and the Non-Interactive tasks, which were presented in counterbalanced order between the
170 participants. During the two tasks, identical stimuli were presented and the participants alternated with
171 their partner in generating the notes. The conditions differed only for task instructions: in the Joint Action
172 condition the color-cue (red, orange, blue or light blue) indicated which of four four-note melodies the
173 participant had to play together with their partner, in alternating turns of playing one note each (i.e.,
174 participants played two of the four notes in turn with their partner while remembering the full melody). In
175 the Non-Interactive condition, the color-cue (yellow, green, pink or violet) was associated with one of
176 four pairs of notes that the participants had to play in two consecutive trials independent of the notes that
177 their partner was playing. For instance, the color-cue could specify: **Joint Action** condition, red melody
178 C-C-G-G, orange melody C-G-C-G, blue melody G-G-C-C, light-blue melody G-C-G-C; **Non-Interactive**
179 condition, green pair C-G; yellow pair C-C; pink pair G-C; violet pair G-G (Suppl Figure S1; the
180 association between colors and melodies/pairs of notes was counterbalanced between participants). Thus,
181 all color-cues conveyed the same amount of information the participant needed to perform the task in two
182 consecutive trials: in the example given above, both the red and the green cues indicated that the

183 participants had first to play a C note and then a G note (see [Fig. 1b](#)). Importantly, however, the color-cue
184 in the Joint Action condition also indicated the notes that the partner would play, whereas the partner's
185 notes were irrelevant in the Non-Interactive condition and therefore not specified.

186 Stimuli presentation and randomization were controlled by E-Prime2 software (Psychology Software
187 Tools Inc.). Visual stimuli were delivered using VisuaStim fiber-optic goggles (800 x 600 pixel
188 resolution).

189

190 Procedure and trial timeline. Each Joint Action/Non-Interactive fMRI run comprised of 128 trials, as it
191 included 16 repetitions of each four-note melody (Joint Action) or pair of notes (Non-Interactive), each of
192 which composed two trials. A "trial" was counted as each time a participant performed a pointing or
193 grasping action on the cube. Accordingly, the task was made of series of two turn-taking productions of
194 pointing/grasping gestures to generate, with a virtual partner, four-note musical sequences ("melodies" in
195 Joint Action and "pairs of notes" in Non-Interactive), each melody being made of two consecutive trials,
196 i.e., Trial1 and Trial2 (see below paragraph "[Types of Trial](#)"). Each run was divided into two blocks (64
197 trials each) with a 30 s break in between. Within each run, the order of melodies (Joint Action) or pairs of
198 notes (Non-Interactive) was pseudorandomized so that each melody/pair could not be consecutively
199 repeated more than twice. The instructions were set up so that participants played either a G- or a C-note
200 50% of the time, and the combination of the participant's and their partner's actions was congruent
201 ("imitative", e.g., pointing in response to the partner's pointing action) or incongruent ("non-imitative",
202 e.g., grasping in response to the partner's pointing action) in 50% of the trials. The order of
203 congruent/incongruent trials was randomized within each block.

204 The pseudo-randomization also included "null trials" showing a black screen, having the same
205 duration as an entire melody/pair of notes (randomly variable between 8 and 6.25 s, mean duration $7.12 \pm$
206 0.59 s) and occurring, on average, every five trials. Overall, each run contained 16 null trials equal to 114
207 s.

208

<Please place Figure 1 here>

209

210 Fig. 1a illustrates the trial timeline, which was identical in the Joint Action and Non-Interactive
211 conditions. For each trial, the virtual partner always took the first turn. Each trial started with the image of
212 a fixation cross (400 ms), followed by an image of the partner in the starting position (200 ms plus a
213 variable stimulus onset asynchrony ranging from 100 to 700 ms), then in the implied-motion position
214 (duration 50% times 100/200 ms), and then in the final position, which was presented synchronously with
215 the partner's note. The image showing the partner's final position also included the color-cue indicating
216 which melody (Joint Action) or pairs of notes (Non-Interactive) the participant had to play. The partner's
217 note constituted the "go-signal" for the participants to release the start button and play their note. The
218 correct note would be played if the response was correct; otherwise, an error signal would sound. The
219 participants were told to complete the task as quickly and correctly as possible. The association between
220 the partner's action (pressing the top, pointing, or the side buttons, grasping) and the ensuing note (G or C
221 note) was always identical to the participant's.

222 Between-trial jittering was ensured by the variable stimulus onset asynchrony (100-700 ms) and
223 by the variability in participants' response times, as the next trial started as soon as participant's response
224 at the previous one was recorded.

225 The participants familiarized themselves with the task and the apparatus in an 8-trial practice
226 block before starting each Joint Action/Non-Interactive fMRI run.

227

228 Types of Trial. Since the color-cues directed participants on what to do in two consecutive trials and
229 appeared at the end of the partner's first move, each melody or pair of notes effectively contained two
230 trials of different "type". In Trial1 (corresponding to the first half of the melody or pair of notes), the
231 participants observed their partner's actions before having seen the color-cue, while in Trial2
232 (corresponding to the second half of the melody or pair of notes) the participants had already seen (in the
233 preceding Trial1) the cue, already knew what to do before observing their partner's action and were thus
234 expected to respond faster.

235 Overall, the participants performed 32 trials per each of the 8 experimental conditions: Joint
236 Action-Trial1-Congruent, Joint Action-Trial1-Incongruent, Joint Action-Trial2-Congruent, Joint Action-

237 Trial2-Incongruent, Non-Interactive-Trial1-Congruent, Non-Interactive-Trial1-Incongruent, Non-
238 Interactive-Trial2-Congruent, Non-Interactive-Trial2-Incongruent.

239

240 **Behavioral data statistical analyses**

241 We measured Accuracy (Acc), i.e., the proportion of correct responses over non-excluded trials, and
242 Response Times (RTs), i.e., the time delay between the go-signal (corresponding to the onset of the end-
243 position picture and of the partner's note) and the instant when the participant pressed a button, measured
244 in correct trials only. We also measured reaction times, i.e., the time delay between the go-signal and the
245 instant the participant released the start button, to exclude from the analysis of both Acc and RTs the trials
246 in which participants made a false start. Overall, valid trials (included in the fMRI analyses) were $30.52 \pm$
247 2.94 per condition in Joint Action (equal to $\sim 95.38\%$ of the trials) and 30.77 ± 1.7 per condition in Non-
248 Interactive (equal to $\sim 96.16\%$ of the trials).

249 For behavioral data analyses, we calculated the individual mean Acc and RTs for each condition,
250 excluding from the analysis of RTs any outlier values that fell 2.5 SDs above or below the individual
251 mean of each experimental condition. Raw Acc and RTs data are reported in Suppl Table S1. Accuracy
252 data were at ceiling (mean Acc = 0.99 ± 0.3 in Joint Action and 0.99 ± 0.2 in Non-Interactive).

253 RTs data were analyzed in the statistical programming environment R (R 3.3.3, [R Core Team 2014](#)).
254 Linear mixed effects models were used as the main statistical procedure. RTs were analyzed as a
255 continuous dependent variable using linear mixed effects models, fitted using the LMER function in
256 "lme4" R package (version 1.1-15, [Bates et al. 2015](#)). As fixed effects, Trial-type (factorial, 2 levels:
257 Trial1 vs. Trial2), Congruence of Actions (factorial, 2 levels: Congruent vs. Incongruent), Task (factorial,
258 2 levels: Joint Action vs. Non-Interactive), and their interactions were tested. The by-subject random
259 intercept was included as a random factor to account for between-subject variability. The inclusion of a
260 main effect or interaction in the final model was assessed performing a series of Likelihood Ratio Tests,
261 including a parameter if it significantly increased the model's goodness of fit. The results of this
262 procedure are summarized in Table 1.

263 The best final fitting model included all factors (Trial-type, Congruence, and Task) and their
264 interactions. We report the significant fixed effects of the final best fitting model with significance levels

265 based on Satterthwaite's degrees of freedom approximation implemented in "lmerTest" R package (version
266 3.0-1, [Kuznetsova et al. 2017](#)). Moreover, to directly contrast single levels of the significant interactions,
267 post-hoc procedures were carried out on the best fitting model with the "phia" R package (version 0.2-1,
268 [Rosario Martinez 2015](#)), applying Bonferroni correction for multiple comparisons. All tests of significance
269 were based upon an α level of 0.05.

270

271 **MRI data acquisition and analyses**

272 Data acquisition. MRI scans were acquired using a Siemens Magnetom Avanto 1.5 T scanner (Siemens
273 AG, Erlangen, Germany) equipped with gradient-echo echo-planar imaging (EPI) [repetition time (TR)
274 2150 ms, echo time (TE) 40 ms, flip angle 90°, 28 slices, slice thickness 4 mm, interleaved slice
275 acquisition, matrix 64 × 64, FOV 250 × 250 mm]. Overall, 300 scans per run were acquired. The first two
276 volumes recorded from each functional run were removed to allow for steady-state tissue magnetization.
277 MPRAGE high-resolution T1-weighted structural images were also acquired (flip angle 35°, TE 5 ms, TR
278 21 ms, FOV 256 x 192 mm, matrix 256 × 256, TI 768, for a total of 160 axial slices with 1 x 1 x 1 mm
279 voxels).

280

281 Preprocessing. After image reconstruction, raw data visualization and conversion from the DICOM to the
282 NIfTI format were performed with MRICron (www.mricron.com) software. All subsequent data analyses
283 were performed in MATLAB R2014b (MathWorks) using Statistical Parametric Mapping software
284 (SPM12, Wellcome Department of Imaging Neuroscience, London, UK). First, slice-timing correction as
285 implemented in SPM12 was applied. Then, the fMRI scans were realigned and unwarped to account for
286 any movement during the experiment and to reduce the effect of magnetic field distortions; the unwarped
287 images were co-registered with the T1-weighted structural image of each participant, which was then
288 segmented and stereotactically normalized into the SPM12 template (tmp.nii) to allow for group analyses
289 of the data. Deformation fields used for T1 segmentation were then applied to the co-registered functional
290 scans. At this stage, the data matrix was interpolated to produce voxels 2 × 2 × 2 mm in dimension. The
291 stereotactically normalized scans were smoothed using a Gaussian filter of 10 × 10 × 10 mm to improve
292 the signal-to-noise ratio.

293 Artifact Detection Tools (ART, Whitfield-Gabrieli, http://www.nitrc.org/projects/artifact_detect) was
294 used to identify outlier scans in global signal and movement for each participant. Time-points were
295 marked as outliers when scan-to-scan variations in the global signal exceeded three standard deviations
296 from the mean, and when the compounded measure of movement parameters exceeded 1 mm scan-to-scan
297 movement (on average, excluded volumes were $5.6\% \pm 4.7\%$ in Joint Action and $5.04\% \pm 4.8\%$ in Non-
298 Interactive). Outlier scans were excluded from the single-subject analysis. No participant included in the
299 sample showed more than 20% of outlier time-points in an experimental condition.

300

301 Univariate statistical analyses of fMRI data. A two-step statistical analysis, based on the general linear
302 model (GLM), was performed. The blood oxygen level-dependent (BOLD) signal associated with each
303 experimental condition was analyzed by convolution with a canonical hemodynamic response function
304 (Worsley and Friston 1995). No global normalization was performed. The time series was high-pass
305 filtered at 128 s and pre-whitened by means of an autoregressive model AR(1). This first step implied a
306 fixed-effect analysis, in which condition-specific effects were calculated. We separately analyzed the
307 Joint Action and Non-Interactive runs. In each run, we characterized the BOLD signal associated with
308 observation of the mid-flight hand-posture in each trial (event-related design, see Fig.1a). At this first
309 level of statistical analysis (single-subject level), the experiment conforms to a 2x2 design having Trial-
310 type (Trial1 vs. Trial2) and Congruence of Actions (Congruent vs. Incongruent) as within-subject factors.
311 We thus modeled four regressors of interest per each run (Trial1_Congruent, Trial1_Incongruent,
312 Trial2_Congruent, Trial2_Incongruent), by only including correct trials in which no false-start was
313 recorded. Separate regressors also modeled experimental confounds, including error trials and false-starts
314 and the realigning parameters calculated in the preprocessing step. Finally, we characterized the effects
315 associated with each experimental condition (Trial1-Congruent, Trial1-Incongruent, Trial2-Congruent,
316 Trial2-Incongruent) in both the Joint Action and Non-Interactive runs, each with a weight of +1 for the
317 regressor-of-interest and a weight of zero for all other regressors. These eight effects were then entered
318 into a second-level full-factorial ANOVA that conformed to random effect analyses. The ANOVA
319 included Task (Joint Action vs. Non-Interactive), Trial-type (Trial1 vs. Trial2) and Congruence of
320 Actions (Congruent vs. Incongruent) as within-subject factors.

321 All univariate analyses were conducted at the whole-brain level. We report the regional effects
322 meeting a family-wise error rate (FWER) correction at the cluster-level. The regional effects that also
323 survived a voxel-wise FWER correction are indicated in the tables and in the main text.

324

325 Multivariate pattern analysis (MVPA) of fMRI data. MVPA was performed using the PyMVPA 2.6.5
326 toolbox (www.pymvpa.org) running under Python 2.7.15. We run separate analyses for the Joint Action
327 and Non-Interactive data. The analysis was performed on spmT map images ([Misaki et al. 2010](#)), obtained
328 by re-estimating the SPM12 univariate first-level analyses on preprocessed but spatially unsmoothed
329 fMRI data. In this analysis, events were labeled depending on what the participants observed (a Pointing
330 vs. Grasping action) independently of which motor response followed, i.e., independently of Congruence
331 of Actions. By doing so, this control analysis was independent of the one applied in univariate statistical
332 analysis. The factors included in this first-level analysis were thus Trial-type (Trial1 vs. Trial2) and
333 Observed Action (Pointing vs. Grasping). We then characterized the Trial1 > Trial2 contrast for each
334 observed action (i.e., both Grasping and Pointing). No Z-scoring or averaging were applied within the
335 PyMVPA toolbox. The analysis was restricted to the 222 voxels included in the left vPMc cluster showing
336 a significant Task x Trial-type interaction as assessed by the univariate statistical analyses (see Results).

337 We trained a linear support vector machine classifier to distinguish the Pointing from the
338 Grasping spmT maps, based on a leave-one-subject-out cross-validation procedure, analogous to a
339 second-level, random-effects analysis. More specifically, we performed a searchlight analysis with 4-mm-
340 radius spheres, and for each sphere we calculated the mean classification accuracy across leave-one-
341 subject-out folds, together with the confusion matrix of predicted against actual classes.

342 The significance of the classification accuracies was determined through a Monte Carlo procedure, by
343 randomly permuting the Pointing and Grasping spmT map labels in each sphere 10000 times ([Stelzer et
344 al. 2013](#)), and by then comparing the actual classification accuracy against the random permutation
345 distribution with a Bonferroni corrected threshold ($\alpha < .05$ divided by 222 searchlights becomes α_{corr}
346 $< .0002$). We report the classification accuracies and the corresponding confusion matrices for the
347 significant searchlight spheres.

348

349 **Results**

350 To ease the interpretation of the results, each section begins with a summary of what results would be
351 expected according to the Dyadic Motor Plan or the Top-down control hypotheses.

352

353 **Behavioral data analysis**

354 According to the *Top-down control* hypothesis, a main effect of Congruence would be expected, showing
355 that Incongruent trials are more difficult than Congruent ones in both the Joint Action and the Non-
356 Interactive task. The *Dyadic Motor Plan* hypothesis would rather predict that such a main effect of
357 Congruence is modulated by the Task factor: a decay in performance in Incongruent as compared to
358 Congruent trials (indexing the emergence of visuo-motor interference effects) is expected in the Non-
359 Interactive but not in the Joint Action condition.

360 The whole pattern of behavioral results is reported in [Suppl. Table S1](#) and in [Fig. 2a](#). Results
361 showed a significant main effect of Task ($F(1,23) = 8.07, p = .004$) showing that RTs were shorter in the
362 Joint Action than the Non-Interactive task, indicating a performance advantage of the Joint Action
363 condition. Results also showed a significant main effect of Trial-type ($F(1,23) = 1768.52, p < .001$) and
364 Congruence ($F(1,23) = 26.77, p < .001$) indicating that, overall, RT were shorter in Trial2 than in Trial1
365 and in Congruent than Incongruent trials. These main effects were further specified by two significant
366 interactions.

367 The Trial-type x Congruence ($F(1,23) = 8.59, p = .003$) interaction showed that, although RTs in
368 Trial1 (both in the Congruent and Incongruent conditions) were significantly slower than RTs in Trial2
369 (all $ps < .001$), RTs in Incongruent-Trial1 trials were also slower than Congruent-Trial1 trials (adj means
370 975.00 ms vs. 932.68 ms, $\chi^2 = 32.85, p < .001$); on the contrary, RTs in Incongruent-Trial2 trials were not
371 slower than Congruent-Trial2 trials (adj means 740.09 ms vs. 728.37 ms, $\chi^2 = 2.52, p = .68$). This
372 indicates that visuomotor interference (indexed by higher RTs in Incongruent as compared to Congruent
373 trials) was maximal in Trial1 and reduced in Trial2.

374 More importantly, the Task x Congruence significant interaction ($F(1,23) = 6.03, p = .014$, [Fig.](#)
375 [2b](#)) showed that RTs in the Incongruent condition were slower than the Congruent one only in the Non-
376 Interactive task (adj means 871.38 ms vs. 831.54 ms, $\chi^2 = 29.26, p < .001$); this difference was not

377 significant in the Joint Action task (adj means 843.71 ms vs. 829.52 ms, $\chi^2 = 3.68$, $p = .33$). Importantly,
378 RTs in the Non-Interactive-Incongruent condition were also slower than those in the Joint Action-
379 Incongruent one (adj means 871.38 ms vs. 843.71 ms, $\chi^2 = 14.06$, $p = .001$). This pattern of results is in
380 line with the Dyadic Motor Plan hypothesis.

381

382 <Please insert Figure 2 here>

383

384 **Whole-brain univariate statistical analysis of the fMRI data**

385 According to the *Top-down control* hypothesis, a main effect of Congruence would be expected at the
386 neurophysiological level as well: it would imply a stronger recruitment of the neural regions responsible
387 for executive motor control (e.g., prefrontal and perhaps right parietal areas) in Incongruent as compared
388 to Congruent trials, independently of the Task performed (i.e., both in the Joint Action and in the Non-
389 Interactive task). The *Dyadic Motor Plan* hypothesis would rather predict that such a main effect of
390 Congruence is modulated by the Task factor and that a difference between the neural correlates of
391 Congruent and Incongruent trials is present in the Non-Interactive condition only. In the Joint Action
392 condition, the Dyadic Motor Plan hypothesis would be satisfied by a stronger recruitment of the neural
393 correlates of motor predictive processes (e.g., in the left fronto-parietal network), which should be
394 independent of action Congruence.

395 With regard to the main effects, the ANOVA only showed a significant effect of Trial-type,
396 indicating that a wide fronto-parietal and occipito-temporal network was more active in the Trial1 than
397 the Trial2 condition, possibly reflecting motor preparation and attentional processes (see [Suppl Table](#)
398 [S2a](#)). The main effect of Trial-type also showed that the paracentral lobule and the angular gyrus were
399 more active in Trial2 than in Trial1 (see [Suppl Table S2b](#)). No significant main effect of Congruence and
400 main effect of Task were found.

401 Crucial for the aim of the present study, however, the ANOVA also revealed a highly significant
402 Task x Trial-type interaction selective for a cluster ($k = 222$, $p_{\text{FWER-corr}} = .037$) in the pars opercularis of the
403 inferior frontal gyrus, in the left ventral premotor cortex (lvPMc), at the border between Brodmann's areas
404 6 and 44 (local maxima $MNI -58, 10, 10$, $Z\text{-score} = 4.96$, $p_{\text{FWER-corr}} = .013$, [Fig. 3a](#)). Specifically for Joint

405 Action, this area was more strongly activated for Trial1 than Trial2 (Simple effect of Trial-type in Joint
406 Action, Z-score = 4.55, [Fig. 3b,c](#)). On the contrary, there was no such an effect in the Non-Interactive
407 condition ([Fig. 3b,c](#); Simple effect of Trial-type in Non-Interactive, Z-score = 1.38 in the neighbor MNI
408 coordinate -50, 8,10). The Task x Trial-type interaction did not show any significant effect in the opposite
409 direction, i.e., no brain region was more active in Non-Interactive than in Joint Action in Trial1 as
410 compared to Trial2. A control conjunction analysis also revealed that the result that emerged in the lvPMc
411 was independent of action Congruence, as also indicated by the lack of a significant Task x Trial-type x
412 Congruence triple interaction. Indeed, the conjunction of the Task x Trial-type interactions separately
413 calculated for Congruent and Incongruent trials revealed a significant cluster in the lvPMc that completely
414 overlapped with the one reported in Figure 3 (see [Suppl Figure S2](#)).

415

416 <Please insert Figure 3 here>

417

418 Finally, the ANOVA showed a significant Task x Congruence interaction for a cluster in the right
419 supramarginal gyrus, at the border between Brodmann's area 2 and area 40 ($k = 208$, $p_{\text{FWER-corr}} = .047$; local
420 maxima *MNI* 62, -28, 36, Z-score = 3.88; Figure 4). This area showed less activation in Congruent as
421 compared to Incongruent trials selectively for the Non-Interactive condition, in line with previous studies
422 showing a crucial role of right parietal areas in imitation control (only required in the Incongruent
423 condition) during non-interactive tasks ([Sowden and Catmur 2015](#); [Brass et al. 2009](#)). Crucially, the
424 activation of this area showed no dissociation between Congruent and Incongruent trials during the Joint
425 Action condition. More generally, no region showed an effect of Congruence in Joint Action, as predicted
426 by the Dyadic Motor Plan hypothesis.

427 No other main effect or interaction was significant.

428

429 <Please insert Figure 4 here>

430

431 **Multivariate classification of fMRI activation patterns in the left vPMc**

432 We also run a searchlight MVPA on the same fMRI data to further explore the functional role of the
433 lvPMc activation in the Joint Action condition, which was revealed by the univariate statistical analysis.
434 We reasoned that if the lvPMc activation reflected a deep processing of the observed actions in Joint
435 Action, a supporting vector machine classifier would be able to tell observed pointing from observed
436 grasping activation maps for data recorded during the Joint Action condition. Accordingly, we found that
437 the MVPA classifier distinguished Pointing from Grasping data significantly above chance in the Joint
438 Action condition (classification accuracy of the significant searchlight = 72.92%, $p_{\text{corr}} = .02$). The
439 significant searchlight was within the left vPMc, centered on the *MNI* coordinate -57, 7, 2, in Brodmann's
440 area 44, with a significant confusion matrix ($\chi^2 = 10.83$, $p = .01$). See [Fig. 3d](#).

441 The same results were not expected in the Non-Interactive condition, because the univariate
442 statistical analysis revealed lower activation of lvPMc in this condition. As a matter of fact, searchlight
443 permutation testing indicated that no sphere significantly discriminated between observed pointing and
444 grasping actions in Non-Interactive data (all $p_{\text{s,corr}} > .3$).

445

446

447 Discussion

448 There is a broad theoretical consensus that human behavior is guided by the anticipation of
449 desired effects during one's action planning and execution and during action observation (e.g., [Wolpert et](#)
450 [al. 2003](#); [Hommel 2009](#); [Kilner 2011](#)). Also, it is widely believed that such anticipations become even
451 more prominent during joint actions ([Candidi et al. 2015](#)), when agents form motor representations that
452 specify the interaction outcomes ([Loher et al. 2013](#); [Loher and Vesper 2016](#); [Keller et al. 2016](#)), which in
453 turn guide and influence individual motor planning ([Pfister et al. 2014](#); [Yamaguchi et al. 2017](#)). The
454 results of the present study show that these joint representations, which we propose form a Dyadic Motor
455 Plan, are strictly motoric in nature, as they are paralleled by a modulation of lvPMc activity, and
456 consistent with a predictive coding hypothesis ([Kilner et al. 2007](#)).

457 We reach such conclusions having tested the hypothesis that an interactive context shapes how the
458 brain processes the observed actions of an interacting partner. We evaluated whether perceptually
459 identical visual stimuli (i.e., the partner's movements) might be differently coded depending on whether

460 they are relevant to achieve an interactive shared goal, e.g., to play a specific melody in turn-taking. The
461 study was motivated by both every-day life and experimental evidence of the ease whereby non-imitative
462 interactions take place. We investigated why visuomotor interference disappears in these instances
463 (Sacheli et al. 2018a), while it is well present in non-interactive situations (Campbell and Cunnington
464 2017; Cross and Jacoboni 2014; Brass et al. 2009). We hypothesized that different neural computations
465 take place during action observation depending on task interactivity. Crucially, within this line, the
466 computations characterizing "interactive" action observation should be identical during imitative and non-
467 imitative social exchanges: this would explain why imitative and non-imitative interactions occur with
468 similar ease when they are guided by a shared goal.

469

470 In our experiment, we compared two alternative theoretical standpoints: the Dyadic Motor Plan
471 and the Top-down control hypotheses. According to our *Dyadic Motor Plan* hypothesis (Sacheli et al.
472 2018a), interactive agents represent the partner's actions within a unitary motor representation describing
473 both their own and the partners' required contribution to the achievement of a shared goal. At the
474 neurophysiological level, this would translate in the recruitment of fronto-parietal areas to anticipate the
475 partners' goal and effect when the shared goal is still unknown and thus needs to be inferred (e.g., in
476 Trial1 in the present experiment); on the contrary, knowledge of the shared goal would lead to signal
477 attenuation (Kaiser and Schutz-Bosbach 2018) as long as the partner's action is correctly anticipated
478 within the Dyadic Motor Plan representation (e.g., in Trial2 in the present experiment). These fronto-
479 parietal areas responsible for predictive motor processes would be equally recruited during imitative and
480 non-imitative exchanges providing that they occur within a shared-goal setting. On the contrary, a *Top-*
481 *down control* hypothesis would have been satisfied by the specific additional activations, beyond
482 premotor regions, of brain areas associated with executive control - e.g., anterior prefrontal areas - to
483 inhibit automatic imitation during incongruent trials independently of the task interactivity. Behaviorally,
484 visuomotor interference should then have been observed in both non-interactive and interactive contexts.

485 Our results are firmly in favor of the Dyadic Motor Plan hypothesis. We show that no behavioral
486 visuomotor interference emerged during the Joint Action condition for the physically incongruent trials,
487 the subjects being faster in this condition than in the Non-interactive one (Figure 2). Importantly, the fact

488 that the reaction times were overall faster in the Joint Action than in the Non-Interactive condition (as
489 shown by the main effect of Task) ensures that, thanks to the preceding training phase, the Joint Action
490 and Non-Interactive tasks were matched for difficulty at the moment of the fMRI session: this also rules
491 out that our results can be accounted for by a stronger attentional demand or working-memory load in the
492 Joint Action condition.

493 At the neurophysiological level, we found two main results. First, the main task-specific
494 difference was in premotor rather than prefrontal cortex, in the lvPMc, pars opercularis of the inferior
495 frontal gyrus. These results were obtained with an unconstrained whole-brain analysis, corrected for
496 multiple comparisons (Figure 3). The MVPA analysis, which was focused on the volume of interest
497 showing the Task x Trial-type significant effect in the univariate analysis, was aimed to further explore
498 the functional role of the lvPMC activation in the Joint Action condition: the analysis revealed that it was
499 possible to decode from the lvPMC activation maps collected during the Joint Action task what specific
500 action (grasping vs. pointing) the partner was performing. Such decoding was not possible from the
501 activation maps recorded during the Non-Interactive task, which generally showed lower recruitment of
502 the lvPMC as revealed by the univariate statistical analysis. Although a formal comparison between
503 decoding accuracy between the two experimental conditions was not performed, the overall pattern of
504 results suggests that different types (or at least degrees) of motor simulation of the partner's action might
505 be at play depending on task interactivity.

506 The second result concerns a right posterior parietal cluster, which showed a Task x Congruence
507 effect that was due to the stronger recruitment of this brain regions in Incongruent as compared to
508 Congruent trials selectively for the Non-Interactive condition (Figure 4), in line with previous studies on
509 imitation control ([Sowden and Catmur 2015](#); [Brass et al. 2009](#)). Accordingly, this right parietal activation
510 seems to be specific for imitation control when no shared goal is present. This result is in line with
511 evidence for an important role played by the posterior parietal cortex in dealing with conflict information
512 ([Liston et al., 2006](#)), especially when a stimulus activates conflicting action plans ([Coulthard et al., 2008](#)):
513 in our Non-Interactive Incongruent trials, the stimulus activates both an imitative response (due to the
514 motor information conveyed by the partner's action) and a non-imitative one (specified by the color cue).
515 This might be the reason for the lower activation of the right posterior parietal cortex in Congruent as

516 compared to Incongruent trials, because in the Congruent condition the two sources of information
517 activate a coherent response. Crucially, the posterior parietal cortex was equally active in Congruent and
518 Incongruent trials in the Joint Action condition, when no region showed an effect of Congruence, as
519 predicted by the Dyadic Motor Plan hypothesis.

520

521 The lvPMc is a classical "mirror" area: its activity has been associated with several processes
522 linked to action observation, such as action categorization, understanding, and interpretation ([Rizzolatti
523 and Sinigaglia 2016](#)). However, evidence in human ([Avenanti et al. 2018](#)) and non-human primates
524 ([Umiltà et al. 2001](#); [Kohler et al. 2002](#)) show that the action observation fronto-parietal network, and in
525 particular specific neuronal populations within the lvPMc, can represent the goals and effects of produced
526 and observed actions independently of the movements needed to achieve them. Moreover, the lvPMc is a
527 critical hub for auditory-motor integration ([Kaplan and Iacoboni 2007](#); [Lahav et al. 2007](#)). We suggest that
528 the process underlying the recruitment of the lvPMc during Joint Action might be the prediction of the
529 partner's action goal and effect, i.e., of the combination between the grasping/pointing action and the
530 ensuing musical output. Although we acknowledge that the lvPMc is also responsible for the processing
531 of non-motor information like those conveyed by abstract hierarchical sequences ([Fiebach and Schubotz
532 2006](#); [Schubotz and Cramon 2004](#)), we interpret its stronger recruitment in the Trial1 of the Joint Action
533 condition as indicative of the activation of auditory-motor associations because this is fully consistent
534 with our previous behavioral evidence ([Sacheli et al. 2018a](#)).

535 Indeed, in the study by Sacheli and colleagues ([2018a](#)) we demonstrated that, only for the Joint
536 Action task, participants, after initial training with the task and action/sound associations, are vulnerable
537 to a subsequent reversal of such associations between the (grasping vs. pointing) actions and the (C vs. G)
538 notes produced by the partner (see Fig. 4 in [Sacheli et al. 2018a](#)). We interpreted these findings by
539 suggesting that, during the Joint Action task, participants try to predict their partners' action effects from
540 observation: when such predictions are violated, via a reversed association of actions and (auditory)
541 effects, performance decays. The studies by Ticini and colleagues ([Ticini et al., 2012, 2017](#)) support the
542 hypothesis that such action-sound associations are coded in the motor system. Crucially, these
543 modulations took place even though in our paradigm both the Non-Interactive and the Joint Action

544 conditions entailed the use of visually-cued instructions, so that the participants could, in principle, plan
545 their response on the basis of the color-cue alone, or by using an abstract representation of the musical
546 sequence, while completely ignoring the partner's movements. Our present and previous results ([Sacheli
547 et al. 2018a](#)) show that this was not the case, as no visuo-motor interference (in the Non-Interactive
548 condition) or effect of reversed action-note association (in Joint Action, [Sacheli et al. 2018a](#)) would
549 emerge if the participants did not process the partner's movement features. To us, these behavioral results
550 are strong evidence that, as a whole, the task is represented in a motor rather than an abstract, rule-based
551 code.

552 In the present experiment, the best correlate of such predictive computations in Joint Action is
553 represented by the lvPMc activity, for at least two reasons. First, the MVPA analysis revealed that the
554 lvPMc activity encodes the pointing or grasping movements of the partner at mid-flight, when the
555 guesswork of the observer about the partner's goal is maximized; second, the activation of the lvPMc in
556 Joint Action was maximal in Trial1 and much reduced in Trial2, when the shared goal is already known
557 and the sounds generated by the partner are no longer a surprise. It is, therefore, reasonable to interpret as
558 "predictive" the overall pattern of activity seen in lvPMc during the Joint Action condition. Accordingly,
559 the reduction of lvPMc activity in Trial2 for the Joint Action condition is interpretable in terms of
560 predictive coding ([Kilner et al. 2007](#)). Knowledge of the melody allows the agent to represent in advance
561 the expected partner's action and combined musical effect within a Dyadic Motor Plan: it thus works as a
562 "prior", to use a Bayesian jargon, biasing action perception and resulting in signal attenuation (the
563 reduction of lvPMc activity in Trial2), as long as expectations are not violated and prediction errors do
564 not occur. Crucially, these modulations were specific for Joint Action.

565 Evidence that the Joint Action condition was not associated with longer reaction times further
566 testifies that our results cannot be accounted for by potential between-task differences in the complexity
567 of motor planning. Between-task differences could be only due to differences in the participants' social
568 (interactive vs. non-interactive) mind-sets, which in Joint Action is a "we-mode"([Gallotti and Frith 2013](#)).

569

570 How do our results compare with previous findings? It is worth mentioning that a modulation of
571 the recruitment of the lvPMc depending on task interactivity is variably reported in the literature on motor

572 interaction. A prominent role of parietal -only in some cases coupled with premotor- activations was
573 shown by studies applying interactive paradigms requiring on-line interpersonal adaptations to perform
574 synchronous goal-directed actions or to generate specific joint hand configurations ([Newman-Norlund et al. 2008](#);
575 [Kokal et al. 2009](#); [Sacheli et al. 2015, 2018b](#); [Era et al. 2018](#)). On the contrary, other studies
576 showed that premotor areas are crucial to support interpersonal coordination while playing music in a duet
577 ([Hadley et al. 2015](#)) and are recruited when participants observe contingent interactions ([Georgescu et al.](#)
578 [2009](#), but see [Eskenazi et al. 2015](#)) or imagine performing joint as compared to single-agent actions
579 ([Wriessnegger et al. 2016](#)).

580 At first glance, these inconsistencies may seem to suggest that the neural correlates of interactive
581 behaviors might rather be task-dependent, against easy generalizations. Nevertheless, these findings are
582 not difficult to reconcile. Altogether, they even more strongly suggest that the whole point about
583 interactions is that they entail predictions on motor acts and outcomes ([Vesper et al. 2013](#); [Kourtis et al.](#)
584 [2013](#)) directly derived from the presence of shared goals ([Candidi et al. 2015](#), see also [Pesquita et al.](#)
585 [2018](#), [Kourtis et al. 2019](#)): the latter guide individual motor planning and modulate the activity in fronto-
586 parietal areas allowing for the anticipation of the partner's contribution, in terms of outcomes in the
587 environment. The specific area coding such outcomes depends on the nature of the outcome itself, e.g.,
588 hand-configuration or target-object in parietal areas ([Newman-Norlund et al. 2008](#); [Kokal et al. 2009](#);
589 [Sacheli et al. 2015, 2018b](#); [Era et al. 2018](#), see [Zapparoli et al. 2018](#)) vs. action effects in premotor ones
590 ([Hadley et al. 2015](#)). However, what can be generalized is the suggestion that task interactivity shapes
591 sensorimotor simulation in the fronto-parietal network and promotes the prediction of the partner's
592 contribution toward the shared goal achievement, without requiring top-down executive control as it
593 rather happens in non-interactive contexts ([Campbell and Cunnington 2017](#); [Cross and Iacoboni 2014](#);
594 [Brass et al. 2009](#)). In our view, this strictly depends on the ability to integrate both the partner's and one's
595 action within a unitary motor representation that we define a Dyadic Motor Plan.

596

597 To conclude, we demonstrate that task interactivity shapes how we code someone else's actions
598 even in perceptually-matched conditions, where task interactivity is only induced by the presence (in the
599 Joint Action condition) or absence (in the Non-Interactive condition) of a shared goal requiring a

600 contribution from both partners. Notably, we did so by keeping constant, between social contexts, both
601 perceptual features and instructions, as the latter always refer to generating specific notes depending on
602 color-cues. As a final neurobiological consideration, our results highlight that the very same lvPMc,
603 belonging to the "mirror" motor system (Rizzolatti and Sinigaglia 2016), contributes to the deciphering of
604 motor goals beyond mere low-level movement imitation. A recent neurophysiological study on non-
605 human primates (Ferrari-Toniolo et al., 2019) showed that there is a population of left premotor neurons
606 that preferentially discharge when a monkey cooperates with a partner to achieve a motor goal,
607 independently of the specific visual input that the monkey receives at a given moment (Ferrari-Toniolo et
608 al., 2019): this is in line with our suggestion that that the activity of premotor regions might be shaped by
609 social motor experience and acquire the possibility to code higher-order "joint" motor representations.

610 Coming back to our initial question, we suggest that, during an interaction, our brain does not
611 represent online what the partner is doing right now, rather his/her prospective contribution to the shared
612 goal achievement: this is what singles out observation of a partner's action from passive action
613 observation in non-interactive contexts, and what might become crucial to identify computational and
614 neurobiological models of effective interactive exchanges, paving the way to the study of social cognition
615 in health and pathology (see for instance Candidi et al., 2017; Curioni et al., 2017) with a minimalistic
616 and controlled approach.

617

618

619 **List of Supplementary Materials**

620 Supplementary Figure S1 and S2; Supplementary Tables S1 and S2.

621

622

623 Fundings

624 This work was supported by the Milan Center for Neuroscience (NeuroMi) and the Italian Ministry of Health
625 (Ricerca Corrente: protocol L3025).

626

627

628 Acknowledgments

629 We thank Carlo Toneatto for technical assistance, and we would like to dedicate this paper to the fond
630 memory of Dr. Enea Francesco Pavone, founder of BrainTrends, for his essential assistance in the
631 development of the experimental set-up. BrainTrends has no financial or intellectual conflict of interest in
632 connection with the manuscript. **Author Contributions.** L.M.S. designed and performed research, analyzed
633 the data, and wrote the manuscript. E.A. and C.V. performed research, analyzed the data, and edited the
634 manuscript. M.T. analyzed the data and contributed to the writing of the manuscript. E.P. and G.B.
635 supervised the project design and contributed to the writing of the manuscript. **Competing Interests.** The
636 authors declare no competing interests. **Data and materials availability.** All scripts used to run the
637 experiment and all collected data are made available by authors upon request.

638 **References**

- 639 Avenanti A, Paracampo R, Annella L, Tidoni E, Aglioti SM. 2018. Boosting and decreasing action
640 prediction abilities through excitatory and inhibitory tDCS of inferior frontal cortex. *Cereb Cortex*. 28(4),
641 1282-1296.
- 642 Bates D, Maechler M, Bolker B, Walker S. 2015. Fitting Linear Mixed-Effects Models Using lme4. *Journal*
643 *of Statistical Software* 67(1), 1-48.
- 644 Brass M, Ruby P, Spengler S. 2009. Inhibition of imitative behaviour and social cognition. *Philos. Trans.*
645 *R. Soc. Lond. B Biol. Sci.* 364, 2359–2367.
- 646 Butterfill S. 2011. Joint action and development. *The Philosophical Quarterly*, 62(246), 23-47.
- 647 Campbell ME, Cunnington R. 2017. More than an imitation game: Top-down modulation of the human
648 mirror system. *Neurosci Biobehav Rev.* 75, 195-202.
- 649 Candidi M, Sacheli LM, Era V, Canzano L, Tieri G, Aglioti SM. 2017. Come together: human-avatar on-
650 line interactions boost joint-action performance in apraxic patients. *Soc Cogn Affect Neurosci.* 12(11),
651 1793-1802.
- 652 Candidi M, Sacheli LM, Aglioti SM. 2015. From muscles synergies and individual goals to interpersonal
653 synergies and shared goals: mirror neurons and interpersonal action hierarchies: comment on “Grasping
654 synergies: a motor-control approach to the mirror neuron mechanism” by D’Ausilio et al. *Phys. Life Rev.*
655 12, 126-128.
- 656 Coulthard EJ, Nachev P, Husain M. 2008. Control over conflict during movement preparation: role of
657 posterior parietal cortex. *Neuron.* 58(1), 144-57.
- 658 Cross KA, Iacoboni M. 2014. To imitate or not: Avoiding imitation involves preparatory inhibition of motor
659 resonance. *Neuroimage.* 91, 228-36.
- 660 Curioni A, Minio-Paluello I, Sacheli LM, Candidi M, Aglioti SM. 2017. Autistic traits affect
661 interpersonal motor coordination by modulating strategic use of role-based behavior. *Mol Autism.* 8, 23.
- 662 Di Paolo EA, De Jaegher H, Gallagher S. 2010. Can social interaction constitute social cognition? *Trends*
663 *Cogn Sci.* 14(10), 441-447.

- 664 Era V, Candidi M, Gandolfo M, Sacheli LM, Aglioti SM. 2018. Inhibition of left anterior intraparietal
665 sulcus shows that mutual adjustment marks dyadic joint-actions in humans. *Soc Cogn Affect Neurosci*.
666 13(5), 492-500.
- 667 Eskenazi T, Rueschemeyer SA, de Lange FP, Knoblich G, Sebanz N. 2015. Neural correlates of observing
668 joint actions with shared intentions. *Cortex*. 70, 90-100.
- 669 Ferrari-Toniolo S, Visco-Comandini F, Battaglia-Mayer A. 2019. Two Brains in Action: Joint-Action
670 Coding in the Primate Frontal Cortex. *J Neurosci*. 39(18), 3514.
- 671 Fiebach CJ, Schubotz RI. 2006. Dynamic anticipatory processing of hierarchical sequential events: a
672 common role for Broca's area and ventral premotor cortex across domains? *Cortex*. 42(4), 499-502.
- 673 Gallotti M, Frith CD. 2013. Social cognition in the we-mode. *Trends Cogn Sci*. 17(4), 160-165.
- 674 Georgescu AL, et al. 2009. Perceiving nonverbal behavior: neural correlates of processing movement
675 fluency and contingency in dyadic interactions. *Hum Brain Mapp*. 35(4), 1362-78.
- 676 Hadley LV, Novembre G, Keller PE, Pickering, MJ. 2015. Causal role of motor simulation in turn-taking
677 behavior. *J Neurosci*. 35(50), 16516-16520.
- 678 Hommel B. 2009. Action control according to TEC (theory of event coding). *Psychol Res*. 73(4), 512-526.
- 679 Kaiser J, Schütz-Bosbach S. 2018. Sensory attenuation of self-produced signals does not rely on self-
680 specific motor predictions. *Eur J Neurosci*. 47(11), 1303-1310.
- 681 Kaplan JT, Iacoboni M. 2007. Multimodal action representation in human left ventral premotor cortex. *Cogn*
682 *Process*. 8(2), 103-13.
- 683 Keller PE, Novembre G, Loehr J. 2016. Musical ensemble performance: representing self, other, and joint
684 action outcomes. In Obhi, S.S., Cross, E.S. (Eds.), *Shared Representations: Sensorimotor Foundations of*
685 *Social Life* (pp. 280-310). Cambridge, UK: Cambridge University Press.
- 686 Kilner JM, Friston KJ, Frith CD. 2007. Predictive coding: an account of the mirror neuron system. *Cogn*
687 *Process*. 8(3), 159-166.
- 688 Kilner JM. 2011. More than one pathway to action understanding. *Trends Cogn Sci*. 15(8), 352-357.

- 689 Kohler E, et al. 2002. Hearing sounds, understanding actions: action representation in mirror neurons.
690 *Science*. 297(5582), 846-848.
- 691 Kokal I, Gazzola V, Keysers C. 2009. Acting together in and beyond the mirror neuron system.
692 *Neuroimage*. 47(4), 2046-2056.
- 693 Kourtis D, Woźniak M, Sebanz N, Knoblich G. 2019. Evidence for we-representations during joint action
694 planning. *Neuropsychologia* 131, 73-83.
- 695 Kourtis D, Sebanz N, Knoblich G. 2013. Predictive representation of other people's actions in joint action
696 planning: An EEG study. *Soc Neurosci*. 8(1), 31-42.
- 697 Kuznetsova A, Brockhoff PB, Christensen RHB. 2017. “lmerTest Package: Tests in Linear Mixed Effects
698 Models.” *Journal of Statistical Software* 82(13), 1-26.
- 699 Lahav A, Saltzman E, Schlaug G. 2007. Action representation of sound: audiomotor recognition network
700 while listening to newly acquired actions. *J Neurosci*. 27(2), 308-314.
- 701 Liston C, Matalon S, Hare TA, Davidson MC, Casey BJ. 2006. Anterior cingulate and posterior parietal
702 cortices are sensitive to dissociable forms of conflict in a task-switching paradigm. *Neuron*. 50(4), 643-53.
- 703 Loehr JD, Kourtis D, Vesper C, Sebanz N, Knoblich G. 2013. Monitoring individual and joint action
704 outcomes in duet music performance. *J. Cog. Neurosci*. 25(7), 1049–1061.
- 705 Loehr JD, Vesper C. 2016. The sound of you and me: Novices represent shared goals in joint action. *Q. J.*
706 *Exp. Psychol. (Hove)* 69, 535–547.
- 707 Misaki M, Kim Y, Bandettini PA, Kriegeskorte N. 2010. Comparison of multivariate classifiers and response
708 normalizations for pattern-information fMRI. *Neuroimage*. 53, 103–118.
- 709 Newman-Norlund RD, Bosga J, Meulenbroek RG, Bekkering H. 2008. Anatomical substrates of
710 cooperative joint-action in a continuous motor task: virtual lifting and balancing. *Neuroimage*. 41(1), 169-
711 77.
- 712 Novembre G, Ticini LF, Schütz-Bosbach S, Keller PE. 2014. Motor simulation and the coordination of self
713 and other in real-time joint action. *Soc Cogn Affect Neurosci*. 9(8), 1062-1068.
- 714 Oldfield RC. 1971. The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia*.

- 715 9(1), 97–113.
- 716 Pesquita A, Whitwell RL, Enns JT. 2018. Predictive joint-action model: A hierarchical predictive approach to
717 human cooperation. *Psychon Bull Rev* 25(5), 1751-1769.
- 718 Pfister R, Dolk T, Prinz W, Kunde W. 2014. Joint response-effect compatibility. *Psychon Bull Rev.* 21(3),
719 817.
- 720 R Core Team. 2014. R: A language and environment for statistical computing. R Foundation for Statistical
721 Computing, Vienna, Austria. URL <http://www.R-project.org/>.
- 722 Rizzolatti G, Sinigaglia C. 2016. The mirror mechanism: a basic principle of brain function. *Nat Rev*
723 *Neurosci.* 17(12), 757-765.
- 724 Rosario Martinez H. 2015. phia: Post-Hoc Interaction Analysis. R package version 0.2-1. [http://CRAN.R-](http://CRAN.R-project.org/package=phia)
725 [project.org/package=phia](http://CRAN.R-project.org/package=phia).
- 726 Sacheli LM, Arcangeli E, Paulesu E. 2018a. Evidence for a dyadic motor plan in joint action. *Sci Rep.* 8(1),
727 5027.
- 728 Sacheli LM, Tieri G, Aglioti SM, Candidi M. 2018b. Transitory inhibition of the left anterior intraparietal
729 sulcus impairs joint actions: a continuous theta-burst stimulation study. *J Cogn Neurosci*, 30(5), 737-751.
- 730 Sacheli LM, Candidi M, Era V, Aglioti SM. 2015. Causative role of left aIPS in coding shared goals during
731 human–avatar complementary joint actions. *Nat Commun.* 6, 7544.
- 732 Schubotz RI, von Cramon DY. 2004. Sequences of abstract nonbiological stimuli share ventral premotor
733 cortex with action observation and imagery. *J Neurosci.* 24(24), 5467-74.
- 734 Sebanz N, Bekkering H, Knoblich G. 2006. Joint action: bodies and minds moving together. *Trends Cogn*
735 *Sci.* 10(2), 70-76.
- 736 Sowden S, Catmur C. 2015. The role of the right temporoparietal junction in the control of imitation.
737 *Cereb Cortex.* 25(4), 1107-13.
- 738 Stelzer J, Chen Y, Turner R. 2013. Statistical inference and multiple testing correction in classification-based
739 multi-voxel pattern analysis (MVPA): random permutations and cluster size control. *Neuroimage.* 65, 69–82.
- 740 Ticini LF, Schütz-Bosbach S, Waszak F. 2017. Mirror and (absence of) counter-mirror responses to action

- 741 sounds measured with TMS. *Soc Cogn Affect Neurosci.* 12(11), 1748-1757.
- 742 Ticini LF, Schütz-Bosbach S, Weiss C, Casile A, Waszak F. 2012. When sounds become actions: higher-
743 order representation of newly learned action sounds in the human motor system. *J Cogn Neurosci.* 24(2),
744 464-74.
- 745 Umiltà MA, et al. 2001. I know what you are doing: A neurophysiological study. *Neuron.* 31(1), 155-165.
- 746 Vesper C, van der Wel RP, Knoblich G, Sebanz N. 2013. Are you ready to jump? Predictive mechanisms in
747 interpersonal coordination. *J Exp Psychol Hum Percept Perform.* 39(1), 48.
- 748 Wolpert DM, Doya K, Kawato M. 2003. A unifying computational framework for motor control and
749 social interaction. *Philos Trans R Soc Lond B Biol Sci.* 358(1431), 593-602.
- 750 Worsley K, Friston KJ. 1995. Analysis of fMRI time-series revisited-again. *Neuroimage* 2, 173–181.
- 751 Wriessnegger SC, Steyrl D, Koschutnig K, Müller-Putz GR. 2016. Cooperation in mind: Motor imagery
752 of joint and single actions is represented in different brain areas. *Brain Cogn.* 109, 19-25.
- 753 Yamaguchi M, Wall HJ, Hommel B. 2017. Action-effect sharing induces task-set sharing in joint task
754 switching. *Cognition* 165, 113–120.
- 755 Zapparoli L, Gandola M, Banfi G, Paulesu E. 2019. A breakdown of imagined visuomotor
756 transformations and its neural correlates in young elderly subjects. *Cereb Cortex.* 29(4), 1682-1696.
- 757

758 **Tables**759 **Table 1.** Model selection for the analysis of RTs.

Model	df	AIC	Chi2	<i>p</i>
RT ~ (1 Subject)	3	78631		
RT ~ (1 Subject) + Trial-Type	4	77101	1532.09	< .001
RT ~ (1 Subject) + Trial-Type + Congruence	5	77077	26.70	< .001
RT ~ (1 Subject) + Trial-Type + Congruence + Task	6	77070	8.145	.004
RT ~ (1 Subject) + Trial-Type * Congruence * Task	10	77064	14.76	.005

760

761

762 **Figure captions**

763 **Figure 1. Trial time-line and experimental conditions.** (a) Trial timeline, which was identical in the Joint Action and
 764 the Non-Interactive tasks. At each trial, participants first saw the partner's movement, and then they heard the ensuing
 765 note concomitant with the instant when the color-cue was shown on the screen at the center of the partner's cube-shaped
 766 response box (third picture after the fixation cross in the figure). After that, participants performed their response
 767 (bottom picture in the figure). The combination between the partner' and participant's actions creates congruent or
 768 incongruent trials (here, a congruent trial is shown). The yellow frame indicates when the onset of fMRI analyses was
 769 set, namely during observation of the partner's mid-flight posture. Given the temporal resolution of the fMRI analysis, it
 770 includes all the following events in the trial (i.e., the presentation of the color-cue and the participant's response). (b)
 771 Schematic representation of the difference between the Joint Action and the Non-Interactive tasks. The color-cue (e.g.,
 772 red or green) conveys the same amount of information regarding the action that the participant has to perform in two
 773 consecutive trials (e.g., play a C note and then a G note, in **bold**). Yet, in the Joint Action condition the color-cue also
 774 specifies what notes the partner will play, while in the Non-Interactive condition the partner's notes are irrelevant and so
 775 not specified (as represented by the question mark). Perceptually, the tasks are identical. Each musical sequence is
 776 composed of two trials, a Trial1 (first half of the musical sequence) and a Trial2 (second half of the musical sequence).

777
 778 **Figure 2. Raw behavioral data (N = 24).** (a) The figure reports the whole pattern of behavioral results for illustrative
 779 purposes, and shows the group mean reaction times in each condition. (b) The figure illustrates the Task x Congruence
 780 significant interaction (individual data averaged across Trial-type). The grey lines indicate single-subject data and the
 781 thick black lines indicate the group means. This interaction effect shows that a visuomotor interference effect, indexed
 782 by the Incongruent (Incongr) > Congruent (Congr) difference, was present in the Non-Interactive condition only. (***) p
 783 < .001.

784
 785 **Figure 3. Task x Trial-type interaction (N = 24).** (a) The results of the Task x Trial-type interaction, showing at the
 786 whole-brain level, only one, highly significant cluster in the left ventral premotor cortex (lvPMc). (b) Simple effects of
 787 Trial-type (Trial1 > Trial2) for the Joint Action (JA) and Non-Interactive (Non-Int) conditions. The activation maps are
 788 visualized at $p_{\text{uncorr}} < .001$ at the voxel-level and $p_{\text{FWE-corr}} < .05$ at the cluster-level. (c) The parameter estimate (beta
 789 values) of the effect of Trial-type (Trial1 > Trial2) as separately tested in the Joint Action (JA) and Non-Interactive
 790 (Non-Int) conditions (error bars indicate 90% of confidence interval). These plots are not independent of the effects
 791 illustrated at point (a) and (b) and are only reported for illustrative purposes. (d) The results of the multivariate pattern
 792 analysis of the fMRI data collected during the Joint Action condition. On the left, the confusion matrix reporting the
 793 classified Pointing/Grasping spmT-maps in the significant searchlight, expressed as percentage values. On the right, the
 794 plot of the actual classification accuracy (%) for the significant searchlight (red line) against the accuracy distribution of
 795 10000 random permutations (blue histogram); the black dashed line indicates the chance classification level.

796
 797 **Figure 4. Task x Congruence of actions interaction (N = 24).** The results indicate that the activity of a cluster in the
 798 right supramarginal gyrus showed dissociation between Congruent and Incongruent trials selectively for the Non-
 799 Interactive condition, while this difference was not present in the Joint Action condition. The Joint Action condition
 800 showed no significant effect of Congruence of actions in any direction and brain region. The plots of the beta values (on
 801 the right) are reported for illustrative purposes. Error bars indicate 90% of confidence interval.