Hierarchical Integration of Communicative and Spatial Perspective-Taking Demands in Sensorimotor Control of Referential Pointing

Rui(睿) Liu(刘),[a] Sara Bögels,[a] Geoffrey Bird,[b,c] W. Pieter Medendorp,[a] Ivan Toni[a]

[a]*Donders Institute for Brain, Cognition and Behaviour, Radboud University*
[b]*Department of Experimental Psychology, University of Oxford*
[c]*Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology & Neuroscience, King's College London*

Abstract

Recognized as a simple communicative behavior, referential pointing is cognitively complex because it invites a communicator to consider an addressee's knowledge. Although we know referential pointing is affected by addressees' physical location, it remains unclear whether and how communicators' inferences about addressees' mental representation of the interaction space influence sensorimotor control of referential pointing. The communicative perspective-taking task requires a communicator to point at one out of multiple referents either *to instruct* an addressee which one should be selected (communicative, COM) or *to predict* which one the addressee will select (non-communicative, NCOM), based on either which referents can be seen (Level-1 perspective-taking, PT1) or how the referents were perceived (Level-2 perspective-taking, PT2) by the addressee. Communicators took longer to initiate the movements in PT2 than PT1 trials, and they held their pointing fingers for longer at the referent in COM than NCOM trials. The novel findings of this study pertain to trajectory control of the pointing movements. Increasing both communicative and perspective-taking demands led to longer pointing trajectories, with an under-additive interaction between those two experimental factors. This finding suggests that participants generate communicative behaviors that are as informative as required rather than overly exaggerated displays, by integrating communicative and perspective-taking

---

Correspondence should be sent to Rui Liu, Donders Institute for Brain, Cognition and Behaviour, Radboud University, the Netherlands. E-mail: ruiliu.psy@gmail.com

information hierarchically during sensorimotor control. This observation has consequences for models of human communication. It implies that the format of communicative and perspective-taking knowledge needs to be commensurate with the movement dynamics controlled by the sensorimotor system.

1. Introduction

Pointing one's finger to a location in space in order to signal a referent for an addressee is one of the biomechanically simpler instances of human communicative behavior, yet referential pointing is cognitively complex: It invites communicators to consider the presumed knowledge of addressees, an instance of recipient design (Blokpoel et al., 2012; Clark & Murphy, 1982; Tomasello, Carpenter, & Liszkowski, 2007). It has been argued that referential pointing is a fundamental human communicative behavior, relevant to understand other forms of human communication, with or without accompanying speech (Cooperrider, 2016; de Ruiter, 2000; de Ruiter, Bangerter, & Dings, 2012; Enfield, Kita, & de Ruiter, 2007; Enfield & Sidnell, 2017; Liszkowski, Brown, Callaghan, Takada, & de Vos, 2012; McNeill, 1992; Tomasello, 2008; Tomasello et al., 2007). Referential pointing provides an ecologically valid testing ground for understanding the integration of sociocognitive information with sensorimotor processing (Creem-Regehr, Gagnon, Geuss, & Stefanucci, 2013; Murillo Oosterwijk et al., 2017; Pouw, Trujillo, & Dixon, 2020; Winner et al., 2019). Previous works have shown how communicators organize their pointing movements incorporating the physical location of addressees (Özyürek, 2002; Cleret de Langavant et al., 2011; Pezzulo, Iodice, Ferraina, & Kessler, 2013; Murillo Oosterwijk et al., 2017; Winner et al., 2019). However, it remains unclear whether and how communicators' inferences about addressees' mental representation of the interaction space influence planning and control of referential pointing.

Inferring others' mental representation of space is conceptualized as spatial perspective-taking, and it is differentiated in two levels according to the cognitive demands involved (Kessler & Rutherford, 2010; Michelon & Zacks, 2006; Salatas & Flavell, 1976; Surtees, Apperly, & Samson, 2013). Level-1 perspective-taking (PT1) is required to infer *whether* part of a visual scene can be seen by another observer. It relies on calculating another observer's line-of-sight in an *egocentric* frame of reference. Level-2 perspective-taking (PT2) is required to infer *how* a visual scene is perceived by another observer. It relies on transforming frames of reference between self and others, adopting an *allocentric* frame of reference. Perspective-taking is considered a prerequisite of intentional communication (Brown-Schmidt & Hanna, 2011; Clements-Stephens, Vasiljevic, Murray, & Shelton, 2013; Keysar, Barr, Balin, & Brauner, 2000). Previous studies have examined how interlocutors make use of perspective information in language comprehension and production, for example, by manipulating shared and privileged information (Director Task, Barr, 2008; Brown-Schmidt & Hanna, 2011; Keysar et al., 2000). Those studies have gained insights into how our knowledge about the other's perspective influences the generation of linguistic cues and joint attention. However, it remains unclear whether and how perspective-taking and communica-

tive demands interface with sensorimotor processes during the production of communicative gestures. Here, we differentiate between three potential mechanisms coupling a communicator's sensorimotor system with the cognitive operations handling communicative and perspective-taking demands.

We developed a communicative perspective-taking (CPT) task in which communicative and perspective-taking demands were manipulated independently, in a full-factorial design. In this task, a communicator and an addressee sit on the opposite side of a table. The communicator points to one among a set of digital stimuli presented at different locations on the table while facing the addressee. A visual barrier is interposed between the addressee and the digital stimuli, allowing the communicator a full view of the stimuli and the barrier, whereas the addressee can see only a subset of the stimuli through the barrier. The communicator points either *to instruct* the addressee which digital stimulus should be chosen, when the addressee does not know the target (communicative pointing, COM) or *to predict* which stimulus the addressee will choose, when the addressee knows the target (non-communicative pointing, NCOM). To decide which stimulus to point to, on each trial, the communicator has to spontaneously infer either which target stimulus can be seen by the addressee through the barrier (PT1), or which of the stimuli seen by the addressee looks like the target when seen from the addressee's perspective (PT2).

We recorded the kinematics of the communicator's right hand throughout the experiment and analyzed temporal and spatial features of the pointing movements. Following previous studies, we focused on kinematic features relevant to parameterizing the sensorimotor processes of movement planning and execution (Becchio, Sartori, & Castiello, 2010; Cleret de Langavant et al., 2011; Murillo Oosterwijk et al., 2017). Each pointing movement was segmented into four stages: planning, approach, holding, and returning, approximately corresponding to the "rest," "stroke," "hold," and "retraction" stages used in the analysis of co-speech gestures (Kita, Gijn, & Hulst, 1998; McNeill, 1992). Temporally, we focused on the duration of the planning stage (i.e., the period between presentation of the stimuli and movement onset) and of the holding stage (i.e., the period between arrival at the end-position and onset of the return movement away from the end-position). Spatially, we computed displacement and variability of the approach trajectories, parameters known to index movement magnitude and movement corrections, respectively (Bays & Wolpert, 2007; Gallivan, Chapman, Wolpert, & Flanagan, 2018; Körding & Wolpert, 2006). Building on existing work, we expect to validate the current experimental setup by reproducing the observations that PT2 requires longer planning time than PT1 (Kessler & Rutherford, 2010; Michelon & Zacks, 2006; Surtees, Apperly, & Samson, 2016) and that communicative demands result in longer holding times and exaggerated pointing trajectories (Becchio et al., 2010; Cleret de Langavant et al., 2011; Krishnan-Barman, Forbes, Hamilton, & de, 2017; Murillo Oosterwijk et al., 2017; Peeters, Chu, Holler, Hagoort, & Özyürek, 2015; Trujillo, Simanova, Özyürek, & Bekkering, 2020).

The kinematic parameters considered in this study have been chosen to differentiate between three hypotheses on the mechanism coupling cognitive and sensorimotor systems during referential pointing (Table 1). Hypothesis 1 ("two steps model") proposes that those systems are independent: A communicator first decides where to point, according to

Table 1
Hypotheses and predictions of perspective-taking, communication, and their interaction effects on pointing trajectories

Hypothesis	Description	Effect on Trajectory
Two-steps model	Perspective-taking only determines referent selection and has no impact on movement trajectory. Communication solely affects movement trajectory	Communication
Leaky two-steps model	Perspective-taking determines referent selection thus affects movement trajectory near the end-position. Communication affects movement trajectory independent from perspective-taking	Communication, perspective-taking
Interactive model	Perspective-taking and communicative demands jointly affect movement trajectory, indicating sensorimotor integration of the two demands	Interaction between Perspective-taking and Communication

perspective-taking demands (Kessler & Rutherford, 2010; Michelon & Zacks, 2006; Surtees et al., 2016), and then plans and controls the movement according to communicative demands. This hypothesis predicts a null effect of perspective-taking on the bulk of the pointing trajectory and limits perspective-taking effects to planning time, while higher communicative demands would lead to more emphatic pointing trajectories as shown previously (Cleret de Langavant et al., 2011; Murillo Oosterwijk et al., 2017; Winner et al., 2019). Hypothesis 2 ("leaky two steps model") proposes that perspective-taking effects, occurring during the determination of the pointing location before movement onset, might leak into later stages of movement execution when the pointing finger approaches the end-position (the "apex" in co-speech gesture literature), as communicators might only consider the accuracy of the pointing end-positions that directly determines the task performance. Under the framework of optimal feedback control (Schwartz, 2016; Scott, 2004, 2016; Todorov & Jordan, 2002, 2003), deviations away from the predetermined forward kinematics (i.e., the anticipatory internal model that specifies end-effectors and joint kinematics) will be corrected only for the task-relevant dimensions. Specifically, the principle of minimum intervention of sensorimotor control argues that only the end-position of the movement, instead of the whole trajectory, is modified by the sensorimotor system, unless external perturbation while moving hampers task performance (Nashed, Crevecoeur, & Scott, 2012; Keyser, Medendorp, & Selen, 2017; Todorov & Jordan 2003). This sensorimotor coding mechanism would predict the leakage of perspective-taking into late-stage movement execution due to the control of end-position; and it would occur in parallel to an early effect of communicative demand on the approach trajectory (Cleret de Langavant et al., 2011; Murillo Oosterwijk et al., 2017; Winner et al., 2019). Hypothesis 3 ("interactive model") proposes that communicators might bear in mind the addressee's perspective as well as communicative demands while planning and controlling the pointing movements. This hypothesis predicts an under-additive interaction effect of perspective-taking and communicative demands in the trajectory. The reason is that the more control signals are present, the more noise emerges in the

sensorimotor system (i.e., "signal-dependent noise"; Faisal, Selen, & Wolpert, 2008; Harris & Wolpert, 1998; Wolpert & Ghahramani, 2000; Wolpert, Doya, & Kawato, 2003). Hypothesis 3 implies that the two types of cognitive demands might share a common coding in the sensorimotor interface, allowing those different cognitive demands to be integrated to generate a more efficient control policy ("control policy" refers to internal rules on how anticipated consequences of a movement and real-time sensory information are used to generate motor commands; Schwartz, 2016; Scott, 2004, 2016; Todorov & Jordan, 2002, 2003).

2. Methods

2.1. Participants

Thirty-eight right-handed healthy participants, naive to the goal of the study, acted as the communicator in the CPT task. All reported normal or corrected-to-normal vision gave informed consent in accordance with protocols approved by the local ethics committee (Committee on Research Involving Human Subjects, region Arnhem-Nijmegen, The Netherlands. Study protocol registration number 3011157.01) and were offered financial compensation for their time. Seven were excluded due to task performance falling below a pre-defined boundary ($<70\%$ trials where the communicator correctly identified the targets), leaving 31 participants (23 females, mean $\pm$ $SD = 23.7 \pm 3.5$ years old) included in the final analysis. The excluded participants used inadequate strategies, such as simply flipping the stimuli upside-down or left-right instead of engaging in perspective-taking. Although the rejected participants might have been able to use PT2 if we had explicitly instructed them to do so, they did not spontaneously use PT2 to solve the task. Given the absence of published effect size estimates for perspective-taking and communicative demands, it was decided to recruit participants until moderate evidence was obtained for ($BF = 3$) or against ($BF = 0.3$) main and interaction effects of the experimental factors, with a maximum amount of 40 participants (Dienes, 2016; Lee & Wagenmakers, 2013; Rouder, 2014).

2.2. Experimental setup

The experimental setup is illustrated in Fig. 1a and b. A pair of co-players sat at opposite sides of a round wooden table (radius: 120 cm). The participant played the role of the communicator; the confederate played the role of the addressee. A set of *stimuli* (four digital items next to each other, each item covering 5×5 cm^2, 7.5-cm apart, Fig. 1b) was presented along the center of the *main screen* (73×44 cm^2; Fig. 1a-i and b) via a projector (DELL M410HD, refresh rate: 60 Hz; Fig. 1a-ii) under the table. A *barrier* (26 cm from the addressee's eyes, Fig. 1a-iii) was vertically interposed between the *main screen* and the addressee. The *barrier* consisted of 70 Liquid Crystal Display (LCD) tiles (1×36 cm^2 each) arranged along the whole extent of the glass panel. The transparency of each LCD tile could be independently controlled through a computer (transition time: 3.2 ms). By controlling transparency of each tile, we allowed or prevented the addressee from seeing some of the stimuli on the main screen (manipulation of PT1, Fig. 1b; also
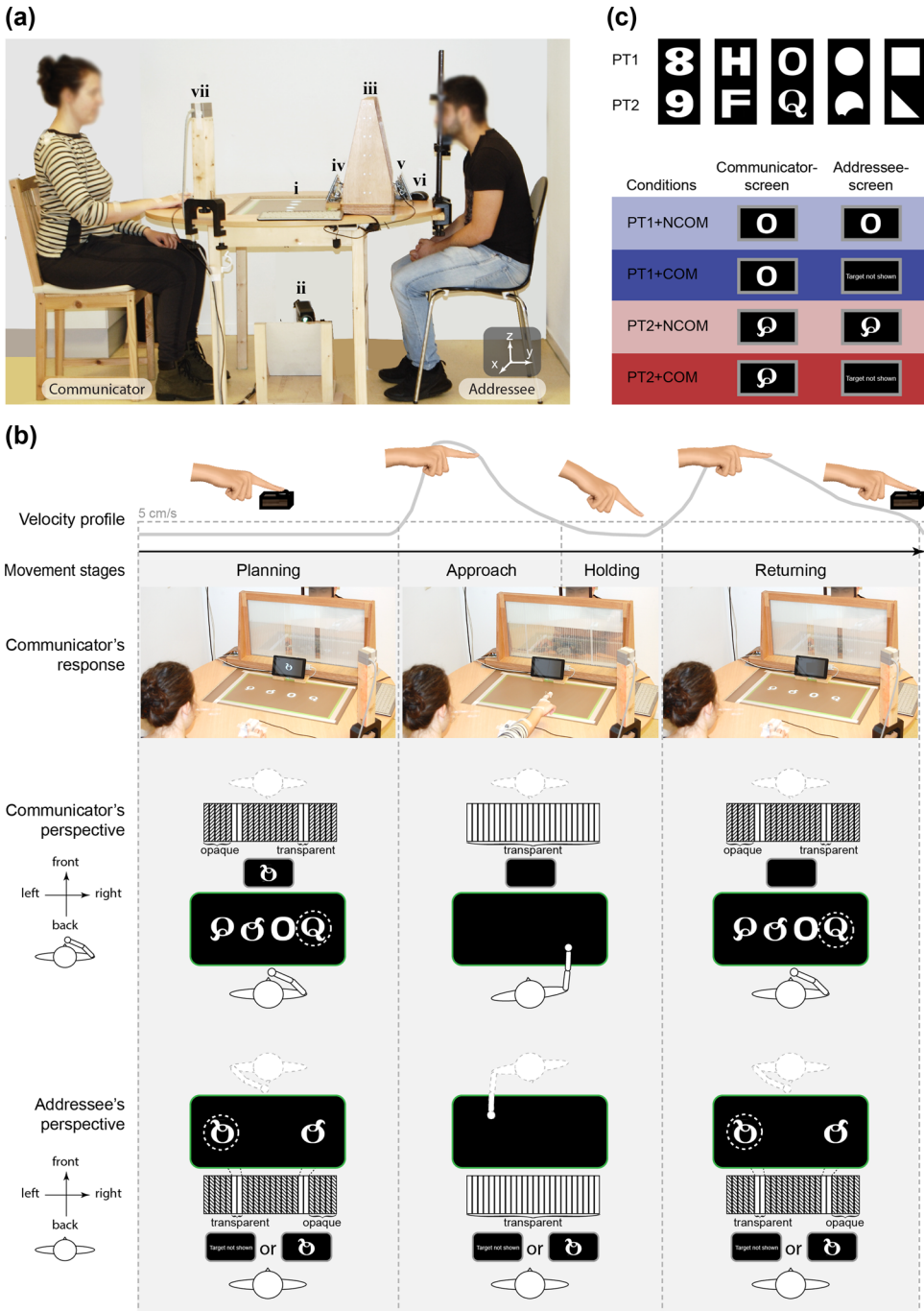
Fig 1. Experimental settings and design. (a) Two co-players sit at the opposite sides of a table. Four stimuli are presented on the main screen (i) through a projector (ii). The communicator (on the left side) needs to point at one of the stimuli, considering which target stimulus can be seen by the addressee through the barrier (iii), or

which of the visible stimuli look like the target presented on the communicator screen (iv) when seen from the addressee's perspective. Subsequently, the addressee selects a stimulus from the main screen either based on the communicator's pointing or on the target presented on the addressee screen (v), by mouse click (vi). (vii) is the source of the magnetic field used by the motion tracking system. (b). The curve indicates a schematic profile of velocity of the pointing finger. Each trial is segmented into four movement stages according to predetermined criteria. The photographs illustrate the corresponding configuration of the experimental setup and the pointing response of the communicator. The last two rows show representative Level-2 perspective-taking (PT2) trials (in communicative pointing (COM) and non- communicative pointing (NCOM) conditions) illustrating the visual scene available to the communicator (upper row) and to the addressee (bottom row). The large black rectangle represents the main screen; the small black rectangle represents the communicator screen (upper row), and the addressee screen (bottom row). The array of shadowed (opaque) and white (transparent) stripes represents the barrier; the dashed lines between the barrier and the main screen (bottom row) illustrate the addressee's line-of-sight through the transparent parts of the barrier. For illustration purpose, we use the dashed circle to represents the correct referent to point to in the example, but the circle was not presented in the real experiment. (c) The upper panel shows the visual stimuli with symmetric (Level-1 perspective-taking (PT1)) or asymmetric (PT2) features (three other mirrored versions of the stimuli were also adopted in the experiment.). The bottom panel denotes the experimental conditions, arising from using symmetric or asymmetric stimuli (PT1, PT2) and from showing or not showing the target stimulus on the addressee screen (NCOM, COM). More precisely, the communicator screen presents the target in all trials. The addressee screen presents the target in the NCOM condition, and the text "Target not shown" in the COM condition.

see Section 2.3 Experiment Task and Design). The upper part (77×14 cm^2) of the barrier was covered to prevent eye contact between the co-players. At either side of the base of the *barrier*, there were two additional small monitors (15.5×8.9 cm^2), one facing the communicator (*communicator screen*; Fig. 1a-iv) and the other the addressee (*addressee screen*, Fig. 1a-v). These small monitors allowed us to manipulate communicative demand by showing or not showing the target on the *addressee screen* (see Section 2.3 Experiment Task and Design). The experimental setup also included a *home key* for the communicator (Fig. 1a), aligned with the midline of the communicator and positioned 38 cm from the center of the *main screen*; a chin rest for the addressee, aligned with the midline of the addressee and positioned 60 cm from the center of the main screen; and a computer mouse (Fig. 1a-vi), held with the right hand of the addressee, which controlled a cursor displayed on the *main screen* after the communicator completed the pointing movement and returned to the *home key*.

2.3. Experimental task and design

The CPT task required a communicator to point at one of the four stimuli presented on the main screen. A trial started with the communicator pressing and holding on the home key with the right index finger, making the barrier opaque and a fixation cross to appear on the main screen. After 1000 ms, portions of the barrier turned transparent, and the set of stimuli was displayed on the main screen. Meanwhile, a target was presented on the communicator screen (Fig. 1b, planning). The communicator decided when and where to point, then released the home key and moved the right hand toward the relevant item on the main screen, before returning to the home key. Releasing the home key triggered the stimuli and target to disappear from the main screen and from the communicator screen and the whole barrier

to become transparent (Fig. 1b, approach and holding). No constraints were imposed on the communicator's pointing movement, other than starting from and returning to the home key, and pointing without touching the main screen. Once the communicator returned to the home key, the stimuli reappeared on the main screen and the barrier became partially opaque as it was at the onset of the trial (Fig. 1b, returning). The addressee could then select an item from the stimuli array on the main screen with the mouse, triggering the appearance of a gray frame around the selected item, which was also visible to the communicator. After a jittered interval (range: 400 to 600 ms, uniform distribution), the next trial would start.

We implemented a 2×2 within-subject factorial design (Fig. 1c). First, the CPT task required the communicator to point for either a communicative (COM) or non-communicative (NCOM) purpose. The communicator was informed that the addressee was supposed to select an item that matched the target shown on the addressee screen, which was identical to the one shown on the communicator screen. In COM trials, however, the target was shown on the communicator screen but not on the addressee screen. Therefore, the addressee could complete the trial only by observing the communicator's pointing movement; in other words, the communicator was required *to instruct* the addressee which item should be chosen by pointing. In NCOM trials, the target was shown on both the communicator screen and the addressee screen. In these trials, the communicator was asked *to predict* which item of the stimuli array would be selected by the addressee. COM and NCOM trials were grouped into four testing blocks of 48 trials each, counter-balanced according to an "ABBA" scheme and the type of the first block was randomly assigned. At the onset of each block of trials, a written instruction was presented on the main screen (i.e., "Instruct" for COM blocks, "Predict" for NCOM blocks), coupled with a red or green frame at the edge of the main screen throughout the 48 trials of a block. The association of the red/green frame with COM/NCOM blocks was randomized across participants, but it was kept consistent within each participant. Second, the communicator was informed that the addressee's choice determined the success of each trial. Therefore, when deciding which item to point at, the communicator needed to adopt the addressee's perspective. Compliance with this crucial task requirement was ensured through these task features:

(a) The communicator needed to consider that the addressee could only select an item that was visible to her, and that matched the target shown on the communicator screen when seen from the addressee's view.

(b) Perspective-taking demand was manipulated through the use of the barrier and of targets with symmetric and asymmetric features. In PT1 trials, the target was symmetric along the horizontal axis (e.g., "O" in Fig. 1c) so that the target appeared the same from the perspectives of both co-players. However, due to the arrangement of the transparent and opaque elements of the barrier (Fig. 1b), some items of the stimuli array were visible to both co-players, whereas other items (including at least one duplicate of the target item) were visible only to the communicator. This manipulation was introduced so that the communicator needed to consider which items were visible from the addressee's point of view since only one of the two displayed target items is the correct answer. This task feature ensured communicators engaged in

PT1. In PT2 trials, the target was asymmetric (e.g., "Q" in Fig. 1c) so that it appeared different from the communicator's and addressee's perspectives (e.g., a "Q" from the communicator's point of view would correspond to a from the addressee's point of view). Note that the communicator could not be able to identify the correct item by simply mirroring or flipping the target (e.g., "Q" and are not reflected images), and such strategy would lower their task performance. Furthermore, the use of this strategy was discouraged by adding mirrored versions of the stimuli array on the main screen. Thus, in order to select the correct item, the communicator needed to adopt the addressee's perspective to determine how the target would look, that is, to engage in PT2.

(c) Each trial started with a new stimuli configuration such that the communicator needed to engage once more in perspective-taking.

(d) Other task details minimized sensory differences between conditions and movement predictability. To match physical features across stimuli arrays, each trial contained both symmetric and asymmetric stimuli. Targets for the four experimental conditions were counterbalanced across the four stimuli locations on the main screen. To minimize stereotyped movements, the items were pseudo-randomly arranged such that no more than three trials with targets at the same location on the main screen were presented in succession. Although the participants only played the role of communicator in the task, they also experienced and understood the role of the addressee during training blocks (see Section 2.4 Experimental Procedure).

2.4. Experimental procedure

The experiment started with training blocks in which participants familiarized themselves with the experimental setup and the roles of the CPT task (started as the role of communicator, then alternated between the roles). The training blocks consisted of at least three blocks of five COM trials and five NCOM trials. Additional blocks were performed until a block was performed with more than 90% accuracy when the participant played the role of communicator. In the training blocks, the players alternated in the roles of communicator and addressee, exchanging positions at the table and experiencing both points of view. During the training blocks, the addressee (played by the confederate) performed the NCOM trials by selecting stimuli on the main screen according to the target shown on the addressee screen (i.e., the players might select a different item based on their own judgement). Then, during the four testing blocks (separated by self-paced breaks), regardless of the trial conditions, the addressee always selected stimuli indicated by the communicator rather than following the targets presented on the addressee screen. This procedure ensured that the communicator received exactly matched feedback from the addressee's behaviour across the COM and NCOM conditions. Trial-by-trial and overall accuracy were presented to the players on the main screen in training blocks, and only overall accuracy was provided at the very end of the testing blocks. Before and after the testing blocks, participants performed 20 pointing movements (baseline trials) to each of the four locations on the main screen where the stimuli

array was displayed, without the presence of the addressee, and with the barrier fully opaque. These measurements provided a baseline reference of each participant's pointing movements in a solo setting without experimental manipulations.

2.5. Kinematic recording and analysis

We sampled the position and orientation of five sensors at 240 Hz, using an electromagnetic tracking system (LIBERTY, Polhemus). The sensors were attached to the communicator's right index finger (on the second distal phalanx, on the proximal phalanx, and the metacarpal joint), wrist (corpus radii), and to the right shoulder. Kinematic data were processed with Kinemagic Toolbox (Verhagen, Dijkerman, Medendorp, & Toni, 2012, see also https://github.com/Donders-Institute/kinemagic) using MATLAB R2016b (MathWorks). The data were filtered using a third-order bidirectional Butterworth low-pass filter at a cut-off frequency of 15 Hz. The velocity and acceleration of the index finger were calculated as the first and second derivatives of the position. Position was defined within a Euclidian reference system centered on the home key, with the x-axis along the communicator's transverse plane, the y-axis along the communicator's sagittal plane, and the z-axis along the communicator's coronal plane (the coordinate system is illustrated in the bottom-left panel of Fig. 1a). Movement onset-and-offset was defined using a criterion of index fingertip velocity in the sagittal plane (y-z plane) over-or-under 5 cm/s for at least 200 ms. Hence, each pointing movement was segmented into four stages (Fig. 1b): Planning (or "rest" in some co-speech gestures literature; Kita et al., 1998; McNeill, 1992)—the period between presentation of the stimuli and movement onset; Approach ("stroke")—the period between movement onset and arrival near the location of the selected stimulus; Holding ("hold")—the period between arrival near the selected stimulus and onset of the return movement away from the selected stimulus; and Returning ("retraction")—from the onset of the return movement until the time the home key was pressed. Only the first three stages were analyzed since the returning movement does not concern referential pointing, and it was constrained by the additional requirement to press the home key.

We examined temporal and spatial aspects of kinematic data. Temporally, we focused on the duration of two movement stages: planning and holding. The spatial features of the pointing movements were quantified as average displacement and variability of the trajectories in three-dimensional space during the approach stage as indices of movement magnitude and movement correction, respectively (see Section 1 Introduction). To compare the temporal development of spatial features of the movement across participants, each approach trajectory was first normalized into 100 time points on a participant level with cubic-spline interpolation. We then extracted time series of average displacement and variability of the trajectories relative to the corresponding values in the baseline movement (performed in the solo setting). Specifically, for each time point on the time-normalized trajectories, we estimated an ellipsoid along the eigenbases of the covariance matrix of the trajectories. The center of the ellipsoid at each time point provided the average position of the pointing finger; the volume of the ellipsoid was computed by taking the 95% confidence interval of the Mahalanobis distance according to a Chi-square distribution. Based on the

ellipsoid at each time point, displacement was calculated as the Euclidian distance from the home key to the center of the ellipsoid; and trajectory variability was calculated as the volume of the ellipsoid. Trials with incorrect responses (i.e., pointing movement toward an incorrect item) were excluded from the kinematic analysis (mean: 6.45%; range: 0.52%–17%). One participant was excluded from the trajectory analysis because of invalid baseline measuring.

2.6. Statistical analysis

To quantify evidence for or against the null hypothesis, we used Bayes Factors (BFs). Stage durations were included into Bayesian repeated-measure analysis of variance (ANOVAs) in JASP 0.9.2 (JASP Team, 2019) with two factors (perspective-taking and communicative demands) with two levels each. We used a multivariate Cauchy prior with a fixed effects scale factor of $r = .50$, a random effects scale factor of $r = 1$, and a scale covariate of $r = .35$ (Rouder, Morey, Speckman, & Province, 2012). Post hoc comparisons were based on the default t test with a Cauchy $(0, r = .71)$ prior (Rouder, Speckman, Sun, Morey, & Iverson, 2009). Statistical inferences on the time courses of displacement and trajectory variability were made using non-parametric cluster-based permutation tests (Maris & Oostenveld, 2007), in order to control for multiple comparisons in the context of data with complex spatiotemporal dependencies. The cluster-forming threshold was $\alpha = .05$ (two-tailed). *P*-values corrected for multiple comparisons (p_{MC}) were computed by ranking the test statistics among sample values from 1000 Monte Carlo runs. We used post hoc exploration of significant effects to evaluate specific main and interaction effects using a Bonferroni procedure (p_{corr}) to correct for multiple comparisons in those pairwise contrasts. These tests were performed with MATLAB R2016b.

3. Results

3.1. Temporal effects of perspective-taking and communicative demands on the pointing movements

We performed a Bayesian repeated-measure ANOVAs on the duration of three-movement stages (planning, approach, and holding). Consistent with our predictions on planning and holding times (see Section 1 Introduction), the BFs suggested extreme evidence for a main effect of perspective-taking on planning time (BF = 1.264e+21, Fig. 2a) and a main effect of communication on holding time (BF = 1.960e+3, Fig. 2c). More precisely, participants took longer to initiate a movement during PT2 trials than during PT1 trials (BF = 2.864e+15; PT2: 4.420 s; PT1: 2.459 s) and held their finger on the end-position location for a longer time at the end of communicative than non-communicative movements (BF = 151.400; COM: 0.249 s; NCOM: 0.209 s). There was no strong evidence to conclude that approach time differed across conditions (Table 2).
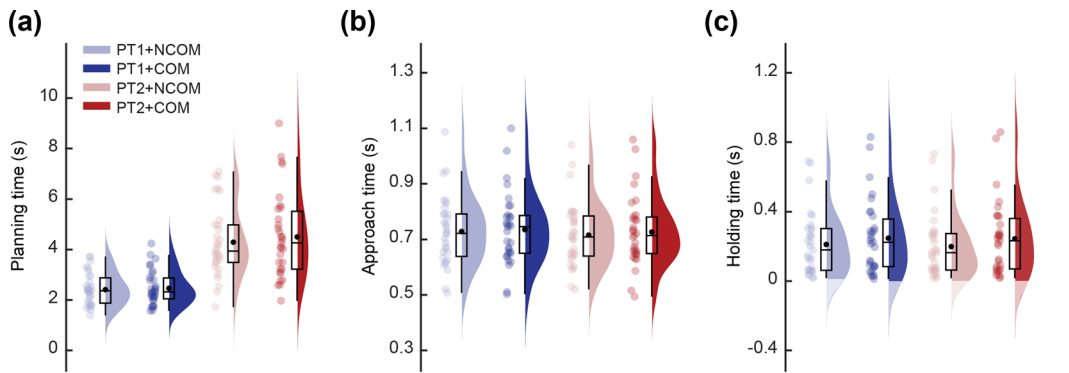
Fig 2. Temporal features of the pointing movements. Durations of (a) planning, (b) approach, and (c) holding stage. The curved areas illustrate kernel density estimates of the data distribution. The colored dots are jittered raw data and the black dots represent the group mean. The boxes indicate the interquartile range (IQR) and the whiskers are maximum and minimum with 1.5 IQR. The task conditions are coded with the color of light blue, dark blue, light red, and dark red for PT1 + NCOM, PT1 + COM, PT2 + NCOM, and PT2 + COM, respectively.

Table 2
Bayes factors (BFs) of the effects of perspective-taking and communicative demands and their interaction on the duration of each movement stage (against the null hypothesis)

Effect (Duration)	Planning	Approach	Holding
Communication effect	0.285	2.036	1.960e+3
Perspective-taking effect	1.264e+21	4.468	0.272
Interaction effect	0.235	0.287	0.297

Note. BFs > 3 or < 0.3 provide moderate evidence and > 10 or < 0.1 strong evidence in favor or against the presence of an experimental effect (Lee & Wagenmakers, 2013).

3.2. Spatial effects of perspective-taking and communicative demands on the pointing movements

To compare the spatial dynamics of the pointing movements, we normalized the raw trajectories (Fig. 3a) with respect to equal time intervals per participant. There are three main findings. First, in line with previous findings, displacement of pointing trajectories showed an early effect of communicative demand (from 9% until 100% of the approach time, $p_{MC} < 0.001$, Fig. 3b), with longer trajectories in COM than in NCOM trials ($t_{29} = 7.531$, $p_{corr} < 0.001$, Cohen's $d = 1.375$. A post hoc analysis revealed that the COM trajectories ended further away from the starting position, compared to the NCOM trajectories (BF = 1.460e+7; COM: 38.14 cm; NCOM: 38.04 cm). Second, there was also a main effect of perspective-taking (from 23% until 100% of the approach time, $p_{MC} < .001$, Fig. 3b), with longer trajectories in PT2 than in PT1 trials ($t_{29} = 8.394$, $p_{corr} < .001$, Cohen's $d = 1.533$). A post hoc analysis revealed that the PT2 trajectories ended further away from the starting position, compared to the PT1 trajectories (BF = 1.019e+13; PT2: 38.18 cm; PT1: 38.00 cm). Third, in favor of Hypothesis 3 ("Interactive model", see Section 1
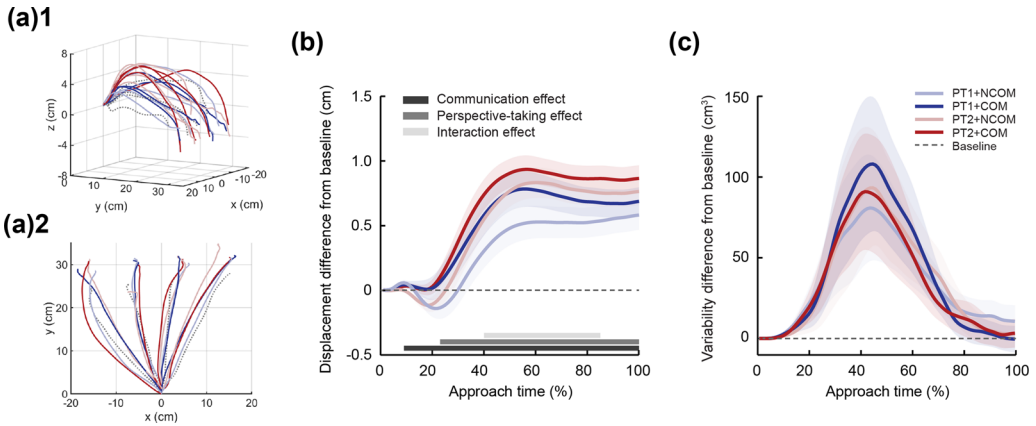
Fig 3. Spatial features of the pointing movements. (a) Illustration of raw pointing trajectories in each experimental condition, both in 3D view (Panel a1) and over the horizontal plane (Panel a2). Each curve represents a randomly chosen trial from a participant. (b) Time courses of the difference in trajectory displacement between task-related and baseline movements, taken as reference (dashed line). Each curve represents group averages; shaded areas indicate ±1 standard error of the mean (SEM) over participants. The bars along the x-axis indicate significant effects of communicative demand (black bar; between 9% and 100% of the approach time, $p_{MC} < .001$); perspective-taking (dark gray bar; between 23% and 100% of the approach time, $p_{MC} < .001$); and interaction between perspective-taking and communicative demand (light gray bar; between 40% and 85% of the approach time, $p_{MC} < .001$). (c) Time courses of the difference in trajectory variability between task-related and baseline movements.

Introduction), there was an under-additive interaction between perspective-taking and communicative demands (from 40% until 85% of the approach time, $p_{MC} < 0.001$, $t_{29} = -5.404$, $p_{corr} < .001$; Cohen's d = .987; Fig. 3b), with a smaller effect of perspective-taking demand (longer trajectories in PT2 than PT1) during COM trials ($t_{29} = 4.670$, $p_{corr} = 0.005$, Cohen's $d = 0.853$) than during NCOM trials ($t_{29} = 17.344$, $p_{corr} < 0.001$, Cohen's $d = 3.167$). A Bayesian one-sample *t*-test suggests that this interaction effect occurred after the initial acceleration phase of the pointing movement (BF = 6521), that is, after the pointing finger reached peak velocity at approximately 35% of the approach time. Moreover, these trajectory displacement effects could not be attributed to changes in trajectory variability: The trajectories remained comparable in that respect across conditions throughout the approach time ($p_{MC} > 0.05$, Fig. 3c).

We also implemented a post hoc analysis to quantify differences between solo pointing and CPT pointing. We contrasted the trajectories of the four task conditions to those of baseline trials (solo pointing without an addressee). Task-related movements had larger displacements starting around 30% of the approaching time and afterward (39%, 32%, 32%, and 26% for NCOM+PT1, NCOM+PT2, COM+PT1, and COM+PT2, respectively). The variability of task-related trajectories was above baseline from 38%, 40%, 29%, and 25% of the approaching time and afterward in NCOM+PT1, NCOM+PT2, COM+PT1, and COM+PT2 condition, respectively; Monte Carlo *p*-values were all smaller than .001. These results revealed that the mere presence of an addressee had a substantial influence on the organization of the point-

ing movement, encouraging communicators to move further away from the starting position. We present the group-averaged differential length and differential variability (relative to the baseline per condition) of movement trajectories in Fig. 3b and c. Fig. 3a provides an example of raw single-trial trajectories to the four stimuli locations in each of the four conditions and at baseline.

4. Discussion

This study tests whether and how communicative and perspective-taking demands are integrated at the sensorimotor level during referential pointing. Communicators changed temporal and spatial features of their pointing movements according to the communicative demand of the task and the spatial perspective of the addressee. The CPT task successfully reproduced well-documented effects in the existing literature on perspective-taking and communicative pointing. Communicators took longer to initiate their movement when inferring how the addressee perceived the referents (PT2) than when discriminating what could be seen by the addressee (PT1; Galati & Avraamides, 2013; Galati, Dale, & Duran, 2019; Kessler & Rutherford, 2010; Michelon & Zacks, 2006; Samson et al., 2010; Surtees et al., 2016). Furthermore, communicators held their finger near the selected stimulus for longer when their pointing movement was required to communicate the target identity to the addressee (Becchio et al., 2010; Cleret de Langavant et al., 2011; Krishnan-Barman et al., 2017; Murillo Oosterwijk et al., 2017; Peeters et al., 2015). The novel finding of this study is that the sensorimotor system controls the trajectory of referential pointing by integrating the communicative demands of the task and the spatial perspective of the addressee ("interactive model"). Both communicative demand and PT2 processes influenced motor responses in a similar manner, that is, lengthened trajectories from an early stage of the pointing movement. Crucially, the effect of PT2 on pointing trajectories decreased when communicators conveyed instructive information to addressees (communicative pointing) as compared to when communicators predicted addressees' response (non-communicative pointing). This evidence for the "interactive model" argues against the alternative possibility of a functional independence between (i) identification of a relevant stimulus according to perspective-taking demands, and (ii) movement specification according to communicative demand. This possibility would have resulted in an effect of perspective-taking limited to the planning stage ("two-steps model") or independent effects of perspective-taking and communicative demands during movement execution ("leaky two-steps model").

4.1. Effects of communicative demand

The CPT task effectively modulated the communicative demands experienced by the communicators; they held their pointing finger near the end-position for longer when they thought the addressee had to identify the correct referent by means of their pointing movements. This observation is consistent with findings from other referential pointing studies (Cleret de Langavant et al., 2011; Murillo Oosterwijk et al., 2017; Peeters et al., 2015; Winner et al., 2019),

as well as naturalistic communicative interactions (Clark & Murphy, 1982; Hilbrink, Gattis, & Levinson, 2015; Levinson, 2016; Richardson, Dale, & Kirkham, 2007). A longer holding time allows the addressee to accumulate more sensory evidence about the end location of the pointing finger (Bangerter, 2004; Sacheli, Tidoni, Pavone, Aglioti, & Candidi, 2013; Sartori, Becchio, Bara, & Castiello, 2009; Vesper, Schmitz, & Knoblich, 2017). Furthermore, this study also confirms that increasing communicative demand led participants to generate longer movement trajectories, starting shortly after movement onset (Murillo Oosterwijk et al., 2017; Winner et al., 2019). A longer holding time and exaggerated trajectory provide a cue that, by virtue of it being longer than necessary and thus instrumentally dysfunctional, ostensively marks the action as communicative (Cleret de Langavant et al., 2011; Cooney, Brady, & McKinney, 2018; Gredebäck & Melinder, 2010; Murillo Oosterwijk et al., 2017; Peeters et al., 2015; Winner et al., 2019).

Details of the experimental design and empirical observations exclude a number of alternative interpretations. For instance, the effects of communicative demand were not a by-product of increased trial-by-trial variability in the movement trajectories or systematic variations in the end location of the pointing finger. Neither could the effects be attributed to feedback from the addressee, as the addressee's responses were matched across conditions, that is, the confederate chose the item indicated by the communicators regardless of the task condition (see Section 2.4. Experimental Procedure). Moreover, the addressee's face was occluded by the barrier so that the communicators were unable to make use of social cues, like gaze or facial expressions (see Section 2.2. Experimental Setup and Fig. 1b).

It is important to emphasize that the effect of communicative demand on movement trajectories occurred over and above the systematically longer trajectories and larger variance observed in the joint than in the solo-task settings, across experimental conditions. Taken together, these observations confirm that communicative demand leads to kinematic re-organization of pointing movements (Chu & Hagoort, 2014; Kita & Özyürek, 2003), generating a spatially exaggerated movement (Becchio et al., 2010; Cleret de Langavant et al., 2011; Krishnan-Barman et al., 2017; Murillo Oosterwijk et al., 2017; Trujillo et al., 2020). These findings raise the possibility that communicative demand leads individuals to change their motor control policy. More precisely, studies of spatially guided movements have repeatedly shown that participants tested in isolation (as in the solo condition of this study) control the end-position of the pointing finger (Nashed et al., 2012; Keyser et al., 2017; Todorov et al., 2003) and/or minimize spatial variability of the finger position (Harris & Wolpert, 1998; Wolpert & Ghahramani, 2000; Wolpert, Ghahramani, & Jordan, 1994). In contrast, during referential pointing, communicators might focus on trajectory control, leading to a tight integration between communicative demands and sensorimotor trajectory control (Chu & Hagoort, 2014; Murillo Oosterwijk et al., 2017; Winner et al., 2019).

4.2. Effects of perspective-taking

When solving the CPT task, participants took longer to start their movement during trials that required them to infer how the addressee perceived the referent. This observation is in line with the established notion that PT2 requires additional cognitive operations as compared to

PT1, for example, inhibition of one's own perspective and transformation from an egocentric to allocentric frames of reference (Galati & Avraamides, 2013; Galati et al., 2019; Kessler & Rutherford, 2010; Michelon & Zacks, 2006; Samson et al., 2010; Surtees et al., 2016). The current study goes further in providing novel evidence of how visuospatial perspective-taking affects spatial features of movement performance. Trials requiring PT2 evoked longer trajectories than trials requiring PT1. The effect started during the initial acceleration phase of the movement, an indication that perspective-taking influences the forward kinematics of a movement (Gallivan et al., 2018; Körding & Wolpert, 2004; Todorov et al., 2003).

The effect of PT2 on pointing trajectory cannot be explained by trajectory adjustments driven by visual feedback of the selected stimulus and its relation to the point of view of the addressee. During the performance of the pointing movements, both target and stimuli disappeared from the communicator screen and the main screen; and the whole barrier turned transparent. The results also revealed that variability of the pointing trajectories, indexing movement correction, remained comparable across levels of visuospatial perspective-taking. These observations suggest that the difference in trajectory displacement is due to perspective-taking effects on the initial forward kinematics of the pointing movements, computed during the planning phase before movement onset.

It could be argued the effect of PT2 on pointing trajectory has a communicative goal, for example, participants ostensively mark their pointing movements for the addressee. However, this possibility does not explain why perspective-taking demand would lead participants to selectively deploy an ostensive marker on the movement trajectory, but then remove it from the holding time, where there are no differences between the two perspective-taking conditions. This inconsistency in PT2 effects over the approach and holding phases of the movement does not match with the effects of communicative demands, where participants ostensively tag both phases of the movement.

Studies on the impact of different frames of reference in coding spatial targets of a movement offer an alternative interpretation of the effect of PT2 on pointing trajectory. Specifically, it is possible that the communicators' coding of the location of the selected stimulus was influenced by the use of different frames of reference in perspective-taking conditions. More precisely, an allocentric frame of reference in PT2 trials. An allocentric frame is required for communicators estimating how the stimuli were perceived by the addressee. In contrast, communicators could estimate whether the objects could be seen by the addressee (PT1 trials) by using an egocentric frame of reference (Crawford, Henriques, & Medendorp, 2011; Flavell, Everett, Croft, & Flavell, 1981; Hart & Moore, 1973; Kessler & Rutherford, 2010; Michelon & Zacks, 2006; Salatas & Flavell, 1976; Surtees et al., 2013). The sensorimotor system uses those frames of reference when planning and executing a movement (Carrozzo, Stratta, McIntyre, & Lacquaniti, 2002; Chen, Byrne, & Crawford, 2011, 2014; Fiehler, Wolf, Klinghammer, & Blohm, 2014; Proulx, Todorov, Taylor Aiken, & Sousa, 2016). In this study, participants aimed their movements at systematically different end-positions in PT1 and PT2, with comparable levels of movement correction (no difference in movement variability), an indication of different initial forward kinematics in the two perspective-taking conditions (Gallivan et al., 2018; Körding & Wolpert, 2004; Todorov et al., 2003). Therefore, it could be suggested that the effect of PT2 on pointing trajectory might arise from differences in the estimation

of the stimulus location due to the different frames of reference used to solve PT1 and PT2 problems.

4.3. Interaction between perspective-taking and communicative demands

The main finding of this study is the under-additive interaction between perspective-taking and communicative demands, indicating that the sensorimotor system makes use of the two types of information in an integrated manner. Communicators lengthened the pointing trajectories after implementing PT2 relative to PT1 but proportionally less when they engaged in intentional communication, that is, when knowing that the addressee's response depended on their pointing movements. The interaction effect occurred after the initial acceleration phase of the movement, an indication that the interaction is likely to reflect control processes occurring during movement execution (Gallivan et al., 2018; Körding & Wolpert, 2004; Todorov et al., 2003). The under-additive nature of the interaction suggests a hierarchical order in how perspective-taking and communicative information are integrated during sensorimotor control. In communicative pointing, the motor controller could privilege the trajectory constraints determined by the distal goal of the movement, that is, to produce a communicative signal that needs to be correctly interpreted by an addressee. Over-lengthened trajectories, determined by a linear summation of the factors shaping the main effects of perspective-taking and communicative demands, reduce the legibility of the movement (Winner et al., 2019) and increase the noise level in the sensorimotor system (Faisal et al., 2008; Harris & Wolpert, 1998; Schwartz, 2016; Scott, 2004, 2016; Wolpert et al., 2003), thus increasing the ambiguity experienced by the addressee when interpreting the observed movement. A motor controller privileging the communicative demands of the movement would imply a reduction of the weight allocated to visuospatial parameters estimated during PT2, that is, over-estimation of target distance observed in the main effect of perspective-taking. This hypothetical change of control parameters (not only the end-position but also the entire trajectory) would provide communicators with the means to finely adjust the pointing movement for the addressee and to follow the Gricean maxim of quantity (Grice, 1975), that is, to generate behaviors as informative as required rather than overly exaggerated displays. These findings indicate that the sensorimotor system can actively integrate different types of conceptual knowledge. In summary, the effect of perspective-taking on trajectory length is present in non-communicative situations, but less so in communicative ones, because then it is overruled by the communicative demands of the situation.

Other empirical observations rule out the possibility that the interaction is a generic effect of task difficulty, either from perspective-taking demand before target selection or from communicative demand at the end of pointing movements (holding stage); the interaction effect was limited to a spatial feature of the pointing movements (trajectory displacement), in the context of statistical evidence against similar interactions on temporal features of those movements, throughout planning, approach, and holding phases of the movements. It is also unlikely that the interaction is due to a ceiling effect caused by biomechanical constraints (e.g., the participants could not adequately reach the intended stimulus given their arm length), which should increase as the movement approaches the intended stimu-

lus; in that case, we should have observed stronger interaction effects in the later section of the movement. However, the interaction effect occurred in the middle section of the movement.

4.4. Interpretational issues

The experimental design relies on a previously established operationalization of visuospatial perspective-taking (Surtees, Butterfill, & Apperly, 2012, 2013, 2016) and did not include additional conditions to demonstrate that participants considered the distance between an observer and the visual stimulus (PT1) or the angular disparity between a participant and an observer's viewpoints (PT2). The findings of the perspective-taking manipulation are consistent with previous studies that have differentiated PT2 from PT1 but could be tested further by manipulating mental rotation demands in perspective-taking. However, we would like to emphasize that the experimental setup ensures communicators used PT1 and PT2 in the corresponding trials. Namely, due to the varying arrangement of transparent and opaque elements of the barrier, during PT1 trials, the communicator needed to consider which of multiple potential stimuli were visible from the addressee's point of view. Simply pointing to one of the stimuli visible to the communicator was not a viable option. Similarly, during PT2 trials, we used targets that appeared different from the communicator's and addressee's perspectives, and that were intermixed with left-right-mirrored versions. Accordingly, during PT2 trials, the communicator needed to adopt the addressee's perspective to determine how the target would look. Simply pointing to the mirror copy of the target was not a viable option (see *2.1.* Participants and 2.3. Experimental Task and Design).

It can be seen that participants produced systematically longer trajectories in the joint than in the solo-task settings, across experimental conditions. This observation extends the notion that joint task settings enhance perspective-taking beyond the selection of the correct referent (Freundlieb, Kovács, & Sebanz, 2016, 2017), showing that joint-task settings change how participants spatially organize their movements (Ciardo, Lugli, Nicoletti, Rubichi, & Iani, 2016; Coello, Quesque, Gigliotti, Ott, & Bruyelle, 2018; Creem-Regehr et al., 2013; Ondobaka, Kilner, & Friston, 2017; Proffitt & Linkenauger, 2013; Proulx et al., 2016). More generally, the current findings raise the possibility that movements with different distal goals use different control policies, providing a structural ground to descriptive differentiations between instrumental, coordinative, and communicative functions of body movements (Vesper & Sevdalis, 2020).

It could be argued that because the task involved several repetitions of the same movements, the participants might have produced stereotypical responses that are qualitatively different from the occasional referential pointing movements occurring during daily communicative interactions. However, participants consistently produced movements that were sensitive to the communicative demands. Specifically, participants produced more ostensive movements during communicative than non-communicative trials throughout the experiment. To understand the mechanistic implementation of human communication, it is fundamental to use experimental settings that allow pairs of individuals to coordinate on a trial-by-trial

basis (Redcay & Schilbach, 2019; Wheatley, Boncz, Toni, & Stolk, 2019) while retaining the ability to quantify their behaviour under controlled and reproducible conditions. However, it remains to be seen whether the current findings, obtained in the context of highly controlled experimental conditions, can be generalized to settings involving the same fundamental phenomenon, but more communicative options, as during natural dialogue with co-speech gestures.

It could also be argued that the use of a confederate might introduce biases in the findings, similar to those reported during face-to-face dialogue (Kuhlen & Brennan, 2013). However, the current settings involve non-verbal stereotyped responses with considerably fewer degrees of freedom than multimodal utterances in spoken dialogue. Furthermore, there was no demand on the confederates to deceive the participants, nor possibilities for the confederates to provide their response in a different manner. Accordingly, the minimal risks of using a confederate in the current circumstances should be weighed against the interpretational advantage of excluding differences in the behaviour of the addressee as a source of experimental effects.

5. Conclusion

This study provides evidence that the sensorimotor system makes use of communicative and perspective-taking information in planning and controlling pointing trajectories, suggesting that during movement execution, communicative constraints are prioritized over the spatial biases evoked by perspective-taking. The finding contributes to understanding the interface between sensorimotor and conceptual abilities in humans, showing that communicative and perspective-taking knowledge penetrates the sensorimotor system. Put differently, sensorimotor processes are not a simple production mechanism, downstream to and encapsulated from cognitive decisions. This observation has consequences for models of human communication. It implies that the format of communicative and perspective-taking knowledge needs to be commensurate with the movement dynamics controlled by the sensorimotor system. The current study highlights the value of bridging largely separated research traditions in sensorimotor control and human communicative behaviour, including co-speech gestures and other referential actions, and it opens the way for studying neural mechanisms integrating other agents' knowledge into communicators' behaviour under controlled yet ecologically relevant experimental conditions.

Open Research Badges

This article has earned Open Data and Open Materials badges. Data are available at https://osf.io/h8yad and materials are available at https://osf.io/z2pq4.

References

Bangerter, A. (2004). Using pointing and describing to achieve joint focus of attention in dialogue. *Psychological Science*, *15*(6), 415–419. https://doi.org/10.1111/j.0956-7976.2004.00694.x

Barr, D. J. (2008). Pragmatic expectations and linguistic evidence: Listeners anticipate but do not integrate common ground. *Cognition*, *109*(1), 18–40. https://doi.org/10.1016/j.cognition.2008.07.005

Bays, P. M., & Wolpert, D. M. (2007). Computational principles of sensorimotor control that minimize uncertainty and variability. *The Journal of Physiology*, *578*(2), 387–396. https://doi.org/10.1113/jphysiol.2006.120121

Becchio, C., Sartori, L., & Castiello, U. (2010). Toward you: The social side of actions. *Current Directions in Psychological Science*, *19*(3), 183–188. https://doi.org/10.1177/0963721410370131

Blokpoel, M., van Kesteren, M., Stolk, A., Haselager, P., Toni, I., & Van Rooij, I. (2012). Recipient design in human communication: Simple heuristics or perspective taking? *Frontiers in Human Neuroscience*, *6*, 253. https://doi.org/10.3389/fnhum.2012.00253

Brown-Schmidt, S., & Hanna, J. (2011). Talking in another persons shoes: Incremental perspective-taking in language processing. *Dialogue & Discourse*, *2*(1), 11–33. https://doi.org/10.5087/dad.2011.102

Carrozzo, M., Stratta, F., McIntyre, J., & Lacquaniti, F. (2002). Cognitive allocentric representations of visual space shape pointing errors. *Experimental Brain Research*, *147*(4), 426–436. https://doi.org/10.1007/s00221-002-1232-4

Chen, Y., Byrne, P., & Crawford, J. D. (2011). Time course of allocentric decay, egocentric decay, and allocentric-to-egocentric conversion in memory-guided reach. *Neuropsychologia*, *49*(1), 49–60. https://doi.org/10.1016/j.neuropsychologia.2010.10.031

Chen, Y., Monaco, S., Byrne, P., Yan, X., Henriques, D. Y. P., & Crawford, J. D. (2014). Allocentric versus egocentric representation of remembered reach targets in human cortex. *The Journal of Neuroscience*, *34*(37), 12515–12526. https://doi.org/10.1523/JNEUROSCI.1445-14.2014

Chu, M., & Hagoort, P. (2014). Synchronization of speech and gesture: Evidence for interaction in action. *Journal of Experimental Psychology: General*, *143*(4), 1726–1741. https://doi.org/10.1037/a0036281

Ciardo, F., Lugli, L., Nicoletti, R., Rubichi, S., & Iani, C. (2016). Action-space coding in social contexts. *Scientific Reports*, *6*, 22673. https://doi.org/10.1038/srep22673

Clark, H. H., & Murphy, G. L. (1982). Audience design in meaning and reference. In J. -F. Le Ny, & W. Kintsch (Eds.), *Advances in psychology* (Vol. *9*, pp. 287–299). Amsterdam: North-Holland. https://doi.org/10.1016/S0166-4115(09)60059-5

Clements-Stephens, A. M., Vasiljevic, K., Murray, A. J., & Shelton, A. L. (2013). The role of potential agents in making spatial perspective taking social. *Frontiers in Human Neuroscience*, *7*, 497. https://doi.org/10.3389/fnhum.2013.00497

Cleret de Langavant, L., Remy, P., Trinkler, I., McIntyre, J., Dupoux, E., Berthoz, A., & Bachoud-Lévi, A.-C. (2011). Behavioral and neural correlates of communication via pointing. *PloS One*, *6*(3), e17719. https://doi.org/10.1371/journal.pone.0017719

Coello, Y., Quesque, F., Gigliotti, M.-F., Ott, L., & Bruyelle, J.-L. (2018). Idiosyncratic representation of peripersonal space depends on the success of one's own motor actions, but also the successful actions of others! *PLoS One*, *13*(5), e0196874. https://doi.org/10.1371/journal.pone.0196874

Cooney, S. M., Brady, N., & McKinney, A. (2018). Pointing perception is precise. *Cognition*, *177*, 226–233. https://doi.org/10.1016/j.cognition.2018.04.021

Cooperrider, K. (2016). The Co-organization of demonstratives and pointing gestures. *Discourse Processes*, *53*(8), 632–656. https://doi.org/10.1080/0163853X.2015.1094280

Crawford, J. D., Henriques, D. Y. P., & Medendorp, W. P. (2011). Three-dimensional transformations for goal-directed action. *Annual Review of Neuroscience*, *34*, 309–331. https://doi.org/10.1146/annurev-neuro-061010-113749

Creem-Regehr, S. H., Gagnon, K. T., Geuss, M. N., & Stefanucci, J. K. (2013). Relating spatial perspective taking to the perception of other's affordances: Providing a foundation for predicting the future behavior of others. *Frontiers in Human Neuroscience*, *7*, https://doi.org/10.3389/fnhum.2013.00596

de Ruiter, J. P. (2000). The production of gesture and speech. In D. McNeill (Ed.), *Language and gesture* (pp. 284–311). Cambridge, MA: Cambridge University Press. https://doi.org/10.1017/CBO9780511620850.018

de Ruiter, J. P., Bangerter, A., & Dings, P. (2012). The interplay between gesture and speech in the production of referring expressions: Investigating the tradeoff hypothesis. *Topics in Cognitive Science*, *4*(2), 232–248. https://doi.org/10.1111/j.1756-8765.2012.01183.x

Dienes, Z. (2016). How Bayes factors change scientific practice. *Journal of Mathematical Psychology*, *72*, 78–89. https://doi.org/10.1016/j.jmp.2015.10.003

Enfield, N. J., Kita, S., & de Ruiter, J. P. (2007). Primary and secondary pragmatic functions of pointing gestures. *Journal of Pragmatics*, *39*(10), 1722–1741. https://doi.org/10.1016/j.pragma.2007.03.001

Enfield, N. J., & Sidnell, J. (2017). *The concept of action*. Cambridge, MA: Cambridge University Press. https://doi.org/10.1017/9781139025928

Faisal, A. A., Selen, L. P. J., & Wolpert, D. M. (2008). Noise in the nervous system. *Nature Reviews. Neuroscience*, *9*(4), 292–303. https://doi.org/10.1038/nrn2258

Fiehler, K., Wolf, C., Klinghammer, M., & Blohm, G. (2014). Integration of egocentric and allocentric information during memory-guided reaching to images of a natural environment. *Frontiers in Human Neuroscience*, *8*, 636. https://doi.org/10.3389/fnhum.2014.00636

Flavell, J. H., Everett, B. A., Croft, K., & Flavell, E. R. (1981). Young children's knowledge about visual perception: Further evidence for the Level 1–Level 2 distinction. *Developmental Psychology*, *17*(1), 99–103. https://doi.org/10.1037/0012-1649.17.1.99

Freundlieb, M., Kovács, Á. M., & Sebanz, N. (2016). When do humans spontaneously adopt another's visuospatial perspective? *Journal of Experimental Psychology. Human Perception and Performance*, *42*(3), 401–412. https://doi.org/10.1037/xhp0000153

Freundlieb, M., Sebanz, N., & Kovács, Á. M. (2017). Out of your sight, out of my mind: Knowledge about another person's visual access modulates spontaneous visuospatial perspective-taking. *Journal of Experimental Psychology. Human Perception and Performance*, *43*(6), 1065–1072. https://doi.org/10.1037/xhp0000379

Galati, A., & Avraamides, M. N. (2013). Flexible spatial perspective-taking: Conversational partners weigh multiple cues in collaborative tasks. *Frontiers in Human Neuroscience*, *7*, 618. https://doi.org/10.3389/fnhum.2013.00618

Galati, A., Dale, R., & Duran, N. D. (2019). Social and configural effects on the cognitive dynamics of perspective-taking. *Journal of Memory and Language*, *104*, 1–24. https://doi.org/10.1016/j.jml.2018.08.007

Gallivan, J. P., Chapman, C. S., Wolpert, D. M., & Flanagan, J. R. (2018). Decision-making in sensorimotor control. *Nature Reviews Neuroscience*, *19*(9), 519–534. https://doi.org/10.1038/s41583-018-0045-9

Gredebäck, G., & Melinder, A. (2010). The development and neural basis of pointing comprehension. *Social Neuroscience*, *5*(5–6), 441–450. https://doi.org/10.1080/17470910903523327

Grice, P. (1975). Logic and conversation. In J. Morgan, & P. Cole (Eds.), *Syntax and semantics* (Vol. *3*, pp. 41–58). New York, NY: Academic Press. http://lefft.xyz/psycholingAU16/readings/grice1975-logic-and-conversation.pdf

Harris, C. M., & Wolpert, D. M. (1998). Signal-dependent noise determines motor planning. *Nature*, *394*(6695), 780–784. https://doi.org/10.1038/29528

Hart, R. A., & Moore, G. T. (1973). The development of spatial cognition: A review. In R. M. Downs, & D. Stea (Eds.), *Image & environment: Cognitive mapping and spatial behavior* (pp. 246–288). Chicago, IL: Aldine-Transaction.

Hilbrink, E. E., Gattis, M., & Levinson, S. C. (2015). Early developmental changes in the timing of turn-taking: A longitudinal study of mother–infant interaction. *Frontiers in Psychology*, *6*, 1492. https://doi.org/10.3389/fpsyg.2015.01492

JASP Team (2019). JASP(Version 0.9.2) [Computer software]. https://static.jasp-stats.org/JASP-0.9.2.0-64-bit.msi

Kessler, K., & Rutherford, H. (2010). The two forms of visuo-spatial perspective taking are differently embodied and subserve different spatial prepositions. *Frontiers in Psychology*, *1*, 213. https://doi.org/10.3389/fpsyg.2010.00213

Keysar, B., Barr, D. J., Balin, J. A., & Brauner, J. S. (2000). Taking perspective in conversation: The role of mutual knowledge in comprehension. *Psychological Science*, *11*(1), 32–38. https://doi.org/10.1111/1467-9280.00211

Keyser, J., Medendorp, W. P., & Selen, L. P. J. (2017). Task-dependent vestibular feedback responses in reaching. *Journal of Neurophysiology*, *118*(1), 84–92. https://doi.org/10.1152/jn.00112.2017

Kita, S., Gijn, I. van, & Hulst, H. van der (1998). Movement phases in signs and co-speech gestures, and their transcription by human coders. In I. Wachsmuth, & M. Fröhlich (Eds.), *Gesture and sign language in human-computer interaction: International gesture workshop* (pp. 23–35). Bielefeld, Germany: Springer. http://doi.org/10.1007/BFb0052986

Kita, S., & Özyürek, A. (2003). What does cross-linguistic variation in semantic coordination of speech and gesture reveal?: Evidence for an interface representation of spatial thinking and speaking. *Journal of Memory and Language*, *48*(1), 16–32. https://doi.org/10.1016/S0749-596X(02)00505-3

Körding, K. P., & Wolpert, D. M. (2004). Bayesian integration in sensorimotor learning. *Nature*, *427*(6971), 244–247. https://doi.org/10.1038/nature02169

Körding, K. P., & Wolpert, D. M. (2006). Bayesian decision theory in sensorimotor control. *Trends in Cognitive Sciences*, *10*(7), 319–326. https://doi.org/10.1016/j.tics.2006.05.003

Krishnan-Barman, S., Forbes, P. A. G., & Hamilton, A. F. D. C. (2017). How can the study of action kinematics inform our understanding of human social interaction? *Neuropsychologia*, *105*, 101–110. https://doi.org/10.1016/j.neuropsychologia.2017.01.018

Kuhlen, A. K., & Brennan, S. E. (2013). Language in dialogue: When confederates might be hazardous to your data. *Psychonomic Bulletin & Review*, *20*(1), 54–72. https://doi.org/10.3758/s13423-012-0341-8

Lee, M. D., & Wagenmakers, E. -J. (2013). *Bayesian cognitive modeling: A practical course*. Cambridge, MA: Cambridge University Press. https://doi.org/10.1017/CBO9781139087759

Levinson, S. C. (2016). Turn-taking in human communication—Origins and implications for language processing. *Trends in Cognitive Sciences*, *20*(1), 6–14. https://doi.org/10.1016/j.tics.2015.10.010

Liszkowski, U., Brown, P., Callaghan, T., Takada, A., & de Vos, C. (2012). A prelinguistic gestural universal of human communication. *Cognitive Science*, *36*(4), 698–713. https://doi.org/10.1111/j.1551-6709.2011.01228.x

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, *164*(1), 177–190. https://doi.org/10.1016/j.jneumeth.2007.03.024

McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. Chicago, IL: University of Chicago Press.

Michelon, P., & Zacks, J. M. (2006). Two kinds of visual perspective taking. *Attention, Perception, & Psychophysics*, *68*(2), 327–337.

Murillo Oosterwijk, A., Boer, M., de, Stolk, A., Hartmann, F., Toni, I., & Verhagen, L. (2017). Communicative knowledge pervasively influences sensorimotor computations. *Scientific Reports*, *7*(1), 4268. https://doi.org/10.1038/s41598-017-04442-w

Nashed, J. Y., Crevecoeur, F., & Scott, S. H. (2012). Influence of the behavioral goal and environmental obstacles on rapid feedback responses. *Journal of Neurophysiology*, *108*(4), 999–1009. https://doi.org/10.1152/jn.01089.2011

Ondobaka, S., Kilner, J., & Friston, K. (2017). The role of interoceptive inference in theory of mind. *Brain and Cognition*, *112*, 64–68. https://doi.org/10.1016/j.bandc.2015.08.002

Özyürek, A. (2002). Do speakers design their cospeech gestures for their addressees? The effects of addressee location on representational gestures. *Journal of Memory and Language*, *46*(4), 688–704. https://doi.org/10.1006/jmla.2001.2826

Peeters, D., Chu, M., Holler, J., Hagoort, P., & Özyürek, A. (2015). Electrophysiological and kinematic correlates of communicative intent in the planning and production of pointing gestures and speech. *Journal of Cognitive Neuroscience*, *27*(12), 2352–2368. https://doi.org/10.1162/jocn_a_00865

Pezzulo, G., Iodice, P., Ferraina, S., & Kessler, K. (2013). Shared action spaces: A basis function framework for social re-calibration of sensorimotor representations supporting joint action. *Frontiers in Human Neuroscience*, *7*, 800. https://doi.org/10.3389/fnhum.2013.00800

Pouw, W., Trujillo, J. P., & Dixon, J. A. (2020). The quantification of gesture-speech synchrony: A tutorial and validation of multimodal data acquisition using device-based and video-based motion tracking. *Behavior Research Methods*, *52*(2), 723–740. https://doi.org/10.3758/s13428-019-01271-9

Proffitt, D. R., & Linkenauger, S. A. (2013). Perception viewed as a phenotypic expression. In W. Prinz, M. Beisert, & A. Herwig (Eds.), *Action science: Foundations of an emerging discipline* (pp. 171–197). Cambridge, MA: The MIT Press.

Proulx, M. J., Todorov, O. S., Taylor Aiken, A., & de Sousa, A. A. (2016). Where am I? Who am I? The relation between spatial cognition, social cognition and individual differences in the built environment. *Frontiers in Psychology*, *7*, 64. https://doi.org/10.3389/fpsyg.2016.00064

Redcay, E., & Schilbach, L. (2019). Using second-person neuroscience to elucidate the mechanisms of social interaction. *Nature Reviews Neuroscience*, *20*(8), 495–505. https://doi.org/10.1038/s41583-019-0179-4

Richardson, D. C., Dale, R., & Kirkham, N. Z. (2007). The art of conversation is coordination: Common ground and the coupling of eye movements during dialogue. *Psychological Science*, *18*(5), 407–413. https://doi.org/10.1111/j.1467-9280.2007.01914.x

Rouder, J. N. (2014). Optional stopping: No problem for Bayesians. *Psychonomic Bulletin & Review*, *21*(2), 301–308. https://doi.org/10.3758/s13423-014-0595-4

Rouder, J. N., Morey, R. D., Speckman, P. L., & Province, J. M. (2012). Default Bayes factors for ANOVA designs. *Journal of Mathematical Psychology*, *56*(5), 356–374. https://doi.org/10.1016/j.jmp.2012.08.001

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian *t* tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, *16*(2), 225–237. https://doi.org/10.3758/PBR.16.2.225

Sacheli, L. M., Tidoni, E., Pavone, E. F., Aglioti, S. M., & Candidi, M. (2013). Kinematics fingerprints of leader and follower role-taking during cooperative joint actions. *Experimental Brain Research*, *226*(4), 473–486. https://doi.org/10.1007/s00221-013-3459-7

Salatas, H., & Flavell, J. H. (1976). Perspective taking: The development of two components of knowledge. *Child Development*, *47*(1), 103–109. https://doi.org/10.2307/1128288

Samson, D., Apperly, I. A., Braithwaite, J. J., Andrews, B. J., Bodley Scott, S. E. (2010). Seeing it their way: Evidence for rapid and involuntary computation of what other people see. *Journal of Experimental Psychology. Human Perception and Performance*, *36*(5), 1255–1266. https://doi.org/10.1037/a0018729

Sartori, L., Becchio, C., Bara, B. G., & Castiello, U. (2009). Does the intention to communicate affect action kinematics? *Consciousness and Cognition*, *18*(3), 766–772. https://doi.org/10.1016/j.concog.2009.06.004

Schwartz, A. B. (2016). Movement: How the brain communicates with the world. *Cell*, *164*(6), 1122–1135. https://doi.org/10.1016/j.cell.2016.02.038

Scott, S. H. (2004). Optimal feedback control and the neural basis of volitional motor control. *Nature Reviews Neuroscience*, *5*(7), 532–545. https://doi.org/10.1038/nrn1427

Scott, S. H. (2016). A functional taxonomy of bottom-up sensory feedback processing for motor actions. *Trends in Neurosciences*, *39*(8), 512–526. https://doi.org/10.1016/j.tins.2016.06.001

Surtees, A., Apperly, I., & Samson, D. (2013). Similarities and differences in visual and spatial perspective-taking processes. *Cognition*, *129*(2), 426–438. https://doi.org/10.1016/j.cognition.2013.06.008

Surtees, A., Apperly, I., & Samson, D. (2016). I've got your number: Spontaneous perspective-taking in an interactive task. *Cognition*, *150*, 43–52. https://doi.org/10.1016/j.cognition.2016.01.014

Surtees, A., Butterfill, S., & Apperly, I. (2012). Direct and indirect measures of Level-2 perspective-taking in children and adults. *The British Journal of Developmental Psychology*, *30*(1), 75–86. https://doi.org/10.1111/j.2044-835X.2011.02063.x

Todorov, E., & Jordan, M. I. (2002). Optimal feedback control as a theory of motor coordination. *Nature Neuroscience*, *5*(11), 1226–1235. https://doi.org/10.1038/nn963

Todorov, E., & Jordan, M. I., (2003). A minimal intervention principle for coordinated movement. In S. Becker, S. Thrun, & K. Obermayer (Eds.), *Advances in neural information processing systems* (Vol. *15*, pp. 27–34). Cambridge, MA: MIT Press. http://papers.nips.cc/paper/2195-a-minimal-intervention-principle-for-coordinated-movement.pdf

Tomasello, M. (2008). *Origins of human communication*. Cambridge, MA: MIT Press.

Tomasello, M., Carpenter, M., & Liszkowski, U. (2007). A new look at infant pointing. *Child Development*, *78*(3), 705–722. https://doi.org/10.1111/j.1467-8624.2007.01025.x

Trujillo, J. P., Simanova, I., Özyürek, A., & Bekkering, H. (2020). Seeing the Unexpected: How Brains Read Communicative Intent through Kinematics. *Cerebral Cortex (New York, N.Y.: 1991)*, *30*(3), 1056–1067. https://doi.org/10.1093/cercor/bhz148

Verhagen, L., Dijkerman, H. C., Medendorp, W. P., & Toni, I. (2012). Cortical dynamics of sensorimotor integration during grasp planning. *Journal of Neuroscience*, *32*(13), 4508–4519. https://doi.org/10.1523/JNEUROSCI.5451-11.2012

Vesper, C., Schmitz, L., & Knoblich, G. (2017). Modulating action duration to establish nonconventional communication. *Journal of Experimental Psychology. General*, *146*(12), 1722–1737. https://doi.org/10.1037/xge0000379

Vesper, C., & Sevdalis, V. (2020). Informing, Coordinating, and performing: A perspective on functions of sensorimotor communication. *Frontiers in Human Neuroscience*, *14*, 168. https://doi.org/10.3389/fnhum.2020.00168

Wheatley, T., Boncz, A., Toni, I., & Stolk, A. (2019). Beyond the isolated brain: The promise and challenge of interacting minds. *Neuron*, *103*(2), 186–188. https://doi.org/10.1016/j.neuron.2019.05.009

Winner, T., Selen, L., Murillo Oosterwijk, A., Verhagen, L., Medendorp, W. P., van Rooij, I., & Toni, I. (2019). Recipient design in communicative pointing. *Cognitive Science*, *43*(5), e12733. https://doi.org/10.1111/cogs.12733

Wolpert, D. M., Doya, K., & Kawato, M. (2003). A unifying computational framework for motor control and social interaction. *Philosophical Transactions of the Royal Society B-Biological Sciences*, *358*(1431), 593–602. https://doi.org/10.1098/rstb.2002.1238

Wolpert, D. M., & Ghahramani, Z. (2000). Computational principles of movement neuroscience. *Nature Neuroscience*, *3*(11), 1212–1217. https://doi.org/10.1038/81497

Wolpert, D. M., Ghahramani, Z., & Jordan, M. I. (1994). Perceptual distortion contributes to the curvature of human reaching movements. *Experimental Brain Research*, *98*(1), 153–156.