Cognition 240 (2023) 105581

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/cognit

What's your point? Insights from virtual reality on the relation between intention and action in the production of pointing gestures

Renuka Raghavan [a,b], Limor Raviv [a,c], David Peeters [d,*]

[a] Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands
[b] Radboud University, Donders Institute for Brain, Cognition, and Behavior, Nijmegen, The Netherlands
[c] Centre for Social, Cognitive and Affective Neuroscience (cSCAN), University of Glasgow, United Kingdom
[d] Tilburg University, Department of Communication and Cognition, TiCC, Tilburg, The Netherlands

ARTICLE INFO

Keywords:
Pointing gesture
Intention
Action
Communication
Virtual reality

ABSTRACT

Human communication involves the process of translating intentions into communicative actions. But how exactly do our intentions surface in the visible communicative behavior we display? Here we focus on pointing gestures, a fundamental building block of everyday communication, and investigate whether and how different types of underlying intent modulate the kinematics of the pointing hand and the brain activity preceding the gestural movement. In a dynamic virtual reality environment, participants pointed at a referent to either share attention with their addressee, inform their addressee, or get their addressee to perform an action. Behaviorally, it was observed that these different underlying intentions modulated how long participants kept their arm and finger still, both prior to starting the movement and when keeping their pointing hand in apex position. In early planning stages, a neurophysiological distinction was observed between a gesture that is used to share attitudes and knowledge with another person versus a gesture that mainly uses that person as a means to perform an action. Together, these findings suggest that our intentions influence our actions from the earliest neurophysiological planning stages to the kinematic endpoint of the movement itself.

1. Introduction

One of the first ways in which we ontogenetically express our communicative intentions is by manually pointing at things in the world around us (Bates, Camaioni, & Volterra, 1975; Butterworth, 2003; Carpenter, Nagell, & Tomasello, 1998), and throughout life, pointing remains a ubiquitous, fundamental, and universal building block of our everyday social interactions (Cooperrider, 2020; Goldin-Meadow, 2007; Kita, 2003). In concert with speech and in its absence, pointing gestures are used for a variety of communicative purposes (Tomasello, 2008). We may, for instance, shift our addressee's attention to an entity by pointing at it to share our personal attitudes towards it ('What a beautiful flower!'). We may inform someone about something by pointing at it ('That's my car, right there.') or may point at something as a request or imperative for assistance ('Could you pass me the salt?'). Not surprisingly, it is therefore sometimes assumed that "the exact same pointing gesture will mean something completely different" as a function of the underlying intention of the speaker and the perceived degree of common ground between speaker and addressee (Tomasello, 2008, p. 3).

Here, we combine motion tracking, electrophysiology and immersive virtual reality to test whether different types of socio-communicative intentions indeed lead to identical pointing gestures, or, alternatively, to pointing gestures that differ in their kinematic profile. In other words, to what extent do people alter the specificities of their pointing movement when having different underlying intentions? Furthermore, we explore the open question of whether different types of socio-communicative intent can be distinguished not just kinematically, but also at the neurophysiological level, and specifically already at early stages preceding the actual execution of the gestural movement. To set the stage for the present study, we will first discuss relevant theoretical considerations with regard to the study of human communicative actions in general, and pointing gestures specifically.

1.1. Intentions and action kinematics

During the everyday interactions we have with others, our hands typically barely rest (Trujillo & Holler, 2021). Non-communicative hand actions such as scratching one's earlobe or grasping a cup to drink from

* Corresponding author at: Department of Communication and Cognition, Tilburg University, P.O. Box 90153, NL-5000 LE Tilburg, The Netherlands.
E-mail address: d.g.t.peeters@tilburguniversity.edu (D. Peeters).

https://doi.org/10.1016/j.cognition.2023.105581
Received 14 February 2022; Received in revised form 3 July 2023; Accepted 26 July 2023
Available online 11 August 2023

it are commonly intermixed with communicative hand gestures that may supplement concurrently produced speech (Kendon, 2004; Kita, 2003; McNeill, 1992). For successful communication to take place, a listener or addressee thus first has to segregate communicatively intended signals from movements that are not directly relevant to the message the speaker wishes to convey (Holler & Levinson, 2019; Kelly, Healey, Özyürek, & Holler, 2015). In addition, the social intention behind a speaker's communicative action needs to be inferred (Bara, 2010; Grice, 1975; Tomasello, 2008). Is my friend pointing at the window as a request for me to open it, or do they aim to shift my attention to a car driving by in the street (Kelly, Barr, Church, & Lynch, 1999)? In the absence of direct access to a speaker's intentions and other mental states, this is a repeatedly occurring yet non-trivial problem that people need to solve in their everyday communication (Clark, Schreuder, & Buttrick, 1983; Tomasello, 2008; Wittgenstein, 1955).

One source of information that could help to ease the challenge of inferring a speaker's intent is if their intentions would be visually reflected in their actions. Indeed, the experimental action literature suggests that this is the case, in that actors' intentions may be visibly translated into the specific kinematic features of the movements they produce (Becchio, Manera, Sartori, Cavallo, & Castiello, 2012; Becchio, Sartori, & Castiello, 2010; Cavallo, Koul, Ansuini, Capozzi, & Becchio, 2016; Krishnan-Barman, Forbes, & de Hamilton, 2017; Vesper & Richardson, 2014). For instance, the velocity of reach-to-grasp movements differs as a function of whether an object is grasped for communicative versus non-communicative reasons (Sartori, Becchio, Bara, & Castiello, 2009), and some kinematic properties of manual actions, such as the size or amplitude a hand movement takes, are typically exaggerated when intended to be more communicative or cooperative (McEllin, Sebanz, & Knoblich, 2018; Sacheli, Tidoni, Pavone, Aglioti, & Candidi, 2013; Trujillo, Simanova, Bekkering, & Özyürek, 2018; Vesper & Richardson, 2014). Observers, in turn, seem capable of discriminating between different intentions on the basis of subtle kinematic properties available in an actor's movement, such as the degree of deceleration of the actor's wrist during a reach-to-grasp movement (Ansuini et al., 2015; Becchio et al., 2012; Cavallo et al., 2016; Manera, Becchio, Cavallo, Sartori, & Castiello, 2011; Sartori, Becchio, & Castiello, 2011; Trujillo, Vaitonyte, Simanova, & Özyürek, 2019). These findings suggest that our intentions are, at least to some extent, visible and derivable from the hand movements we make.

1.2. Intentions and pointing gesture kinematics

In line with the broader action literature, experimental studies on the link between intentions and pointing gesture kinematics have shown that a person's intent also influences the movement parameters of their pointing gesture. For example, it has been observed that people may slow down their pointing gestures towards an object when their gesture is meant to convey a higher degree of novel information with respect to the location of the object they are referring to (Peeters, Chu, Holler, Hagoort, & Özyürek, 2015). Moreover, they typically hold their extended finger in apex position for longer when communicative demands increase, such as when an addressee's task is dependent on the gesture versus when it is not (Cleret de Langavant et al., 2011; Murillo Oosterwijk et al., 2017; Peeters, Chu, Holler, Özyürek, & Hagoort, 2013). Decreasing the velocity of the pointing gesture and extending the duration of its hold phase arguably allows more time for the addressee to recognize the trajectory and vector produced by the communicator's finger and subsequently derive the location and identity of the intended referent (Cooney, Brady, & McKinney, 2018; Herbort & Kunde, 2016). People thus seem to tailor the specific kinematics of their pointing gestures to the communicative needs of their addressees (Cleret de Langavant et al., 2011; Liu, Bögels, Bird, Medendorp, & Toni, 2019; Peeters et al., 2015), often in tight synchronization with concurrently produced speech (Bangerter, 2004; Chu & Hagoort, 2014; Cooperrider, Fenlon, Keane, Brentari, & Goldin-Meadow, 2021; Levelt, Richardson, & La Heij, 1985). These experimental findings are in line with the observation that naturally occurring pointing gestures in spontaneous interactions outside the experimental lab differ in the amount of space they take up as a function of whether they carry more or less foregrounded information for the addressee (Enfield, Kita, & de Ruiter, 2007; see also Cooperrider, 2017).

1.3. Types of intent preceding a pointing gesture

Pointing gestures have traditionally been descriptively classified as either *declarative* or *imperative* pointing gestures as a function of the specific socio-communicative motive driving their production (Bates et al., 1975). Declarative pointing can be broadly defined as an effort to direct an addressee's attention to some object, person, or event in the world, simply to have shared attention on that entity and/or to share information about it with the addressee (Bates et al., 1975). More recently, a further distinction has been made between "declarative as expressive" pointing, in which one aims to share with an addressee one's attitude about a common referent, versus "declarative as informative" pointing, in which an addressee is provided with assumedly relevant information (that they currently lack) about a referent (Tomasello, 2008, p. 118). Typical examples for this distinction include an infant pointing at a van driving by in the street to share attention to that interesting entity with their caregiver (declarative as expressive pointing), as opposed to pointing at your partner's car keys when they are looking for them (declarative as informative pointing; Liszkowski, Carpenter, Henning, Striano, & Tomasello, 2004; Tomasello, 2008). Imperative pointing, on the other hand, refers to "the intentional use of the listener as an agent or tool in achieving some end" (Bates et al., 1975, p. 208). Common everyday examples for imperative pointing include pointing at a window as a request for somebody to open it, or pointing at an out-of-reach salad bowl as a directive for it to be passed over during dinner.

To sum up, naturally occurring referential pointing gestures can be theoretically classified as a function of their underlying socio-communicative intent depending on whether they are used to primarily *share attention and information* with an addressee (as for the "declarative as expressive" and "declarative as informative" pointing gestures mentioned above) or not (as for imperative pointing gestures). However, they can also be categorized as a function of whether they actively and primarily *start from another person's perspective and knowledge state* (as for the "declarative as informative" pointing gestures mentioned above) or from the needs of the self (as for the "declarative as expressive" and imperative pointing gestures mentioned above). Typically, in all these triadic situations (i.e., involving a speaker, an addressee, and a certain referent), the speaker alternates gaze between the addressee and the intended referent while pointing (e.g., Bakeman & Adamson, 1984), creating a dynamic "referential triangle" between the two interlocutors and the thing they are talking about (Butterworth, 2003; Leavens, Hopkins, & Bard, 2005).

As we have seen in the previous section, perhaps surprisingly, most of the experimental work looking into the relation between adult speakers' communicative intentions and their associated pointing gesture kinematics is not shaped by the clear theoretical distinction of pointing gestures described above, in which pointing gestures are being defined and categorized as different subtypes depending on the underlying socio-communicative intent of the speaker. Instead, the experimental studies in this domain have typically looked at pointing behavior in communicative situations in which the pointing gesture was relevant to an addressee's task, and compared these to arguably less communicative situations in which the gesture was, in terms of its communicative value, largely redundant (Liu et al., 2019; Murillo Oosterwijk et al., 2017; Peeters et al., 2015; Winner et al., 2019) or to a situation in which the speaker was asked to produce the gesture for non-communicative reasons (Cleret de Langavant et al., 2011). An MEG study that did distinguish between declarative and imperative pointing gestures did

not analyze the kinematics of these movements (Brunetti et al., 2014). As such, it remains unknown whether and how different naturally occurring types of socio-communicative intentions (e.g., declarative-expressive, declarative-informative, imperative) actually shape the kinematic profile of a speaker's pointing gesture.

Indeed, to what extent do the behavioral findings (e.g., decreasing the velocity of the gesture, prolonging its stroke duration, and extending the duration of its hold stage) reported above generalize from artificially induced *more* versus *less* communicative situations in the lab to the types of pointing gestures distinguished as a function of their underlying socio-communicative intent that we observe in everyday communication? Could we consider some (e.g., declarative) pointing gestures indeed "more communicative" than other (e.g., imperative) pointing gestures? In the present study, we use an immersive virtual reality setup to fill in this gap in the literature, and directly test how different types of underlying intent shape the kinematics of referential index-finger pointing gestures. That is, we aim to disclose whether and how different types of socio-communicative intentions indeed translate into modulations of specific action kinematic features in a naturalistic communicative setting in the lab.

1.4. Neural correlates of socio-communicative intent

Producing a pointing gesture takes cognitive resources and planning time. At a stage preceding the onset of the gestural movement, theory-of-mind-related processes (e.g., processing the addressee's perspective on a referent, determining the degree of common ground between oneself and one's addressee), attention-related processes (allocating more or less attentional resources to inspecting properties of the referent), and movement planning mechanisms (planning the execution of the upcoming hand action) must dynamically interact to allow for the production of a contextually-appropriate gesture (Liu et al., 2019; Peeters et al., 2015).

To date, only a handful of studies have aimed at characterizing neural activity preceding the production of communicative pointing gestures during this early planning stage. These studies suggest that an increase in communicative demands when planning a pointing gesture leads to enhanced activation in the right posterior superior temporal sulcus (pSTS) and in areas often related to theory-of-mind and mentalizing, such as parts of medial prefrontal cortex (Brunetti et al., 2014; Cleret de Langavant et al., 2011; see also Enrici, Adenzato, Cappa, Bara, & Tettamanti, 2011; Willems et al., 2010). Specifically, enhanced activation of medial frontal areas has been detected prior to the production of declarative pointing gestures when compared to imperative pointing gestures (Brunetti et al., 2014). These findings are in line with an observed correlation between activation measured over frontal brain regions in 14-month-old infants and their frequency of declarative (but not imperative) pointing four months later (Henderson, Yoder, Yale, & McDuffie, 2002). Nevertheless, other results indicate enhanced involvement of an arguably critical area of the theory-of-mind network (right temporo-parietal junction; right TPJ) prior to the production of pointing gestures in general, regardless of whether the gesture is produced with a declarative or an imperative motive (Brunetti et al., 2014). Indeed, considering one type of pointing gesture "more communicative" than another may obscure the fact that both declarative and imperative pointing gestures are typically driven by a socio-communicative motive.

In sum, earlier studies have identified a variety of cortical regions and networks putatively involved in planning and executing a pointing gesture. However, we currently do not understand well at what point in time preceding the production of a pointing gesture different intentions may potentially start translating into different patterns of underlying brain activity. The current study therefore analyzes the potential role of intent in modulating electrophysiological brain activity at the earliest stages of planning the execution of a pointing gesture. As such, by combining kinematic and electrophysiological measurements in a single study, the potential influence of different types of intent on planning and producing a pointing gesture can be studied from the earliest neurophysiological planning stages to the kinematic endpoint of the gesture itself.

1.5. The present study

The experiment presented below makes use of immersive virtual reality technology to compare kinematic and electrophysiological correlates of people's intent when they plan and produce pointing gestures in three different conditions: (i) pointing gestures produced to direct an addressee's attention to a referent and share attention (henceforth called "declarative pointing"); (ii) pointing gestures produced to provide an addressee with new information related to a referent ("informative pointing"); and (iii) pointing gestures produced in order to get an addressee to perform an action on a referent ("imperative pointing"). Note that this tripartite theoretical distinction has a long tradition in the study of speech acts (e.g., Grice, 1975; Searle, 1969).

Participants were immersed in a series of virtual 3D environments while the kinematics of their dominant hand and their electroencephalogram (EEG) were continuously recorded. We opted for immersive virtual reality as a mode of dynamic and interactive stimulus display as it allows for a combination of high ecological validity and high experimental control in language research using life-size virtual interlocutors in naturalistic settings in the lab (Huizeling, Peeters, & Hagoort, 2022; Legault et al., 2019; Pan & Hamilton, 2018; Parsons, 2015; Peeters, 2019; Tromp, Peeters, Meyer, & Hagoort, 2018). These virtual agents outperform human confederates in terms of the consistency and replicability of the subtleties of all aspects of their behavior, allowing for reproducible research across participants and labs (Hömke, Holler, & Levinson, 2018; Kuhlen & Brennan, 2013; Pan & Hamilton, 2018; Peeters, 2020).

At the kinematic level, previous experimental work suggests that "an increase in communicative demands" may lead to pointing gestures that have a longer stroke duration, a lower stroke velocity, a later gesture initiation time, and a longer hold duration (Liu et al., 2019; Murillo Oosterwijk et al., 2017; Peeters et al., 2015). It is unclear, however, how such kinematic differences would map onto the different types of socio-communicative intent (i.e., declarative, informative, imperative) typically driving the production of pointing gestures in everyday life and referred to in the theoretical literature.

As we have seen above, the three types of socio-communicative intent we focus on in the current study can be contrasted along at least two theoretical axes. First, while declarative and informative pointing gestures are used to *share attention and information* with an addressee, imperative pointing gestures rather aim to use the addressee as a tool to achieve something. As such, the production of declarative and informative pointing gestures could be considered intrinsically more communicatively demanding in nature than the production of imperative pointing gestures (cf. Brunetti et al., 2014). Hence, based on the empirical literature discussed above, one would predict slower gestural movement (i.e., a longer stroke duration and lower stroke velocity) and longer gesture hold stages (i.e., a longer gesture initiation time and a longer hold duration) for declarative and informative versus imperative pointing gestures. After all, unlike declarative and informative situations, the rationale behind the production of imperative pointing gestures is primarily to use the addressee as a tool to achieve something, rather than to primarily in a communicative way share attention and information on a referent (cf. Tomasello, 2008).

Second, in contrast, if we consider actively *starting from another person's perspective and knowledge state* (rather than from the self) as the critical factor that makes the pointing gesture "more communicative" and increases its "communicative demand", one would predict slower

gestural movement (i.e., a longer stroke duration and lower stroke velocity) and longer gesture hold stages (i.e., a longer gesture initiation time and a longer hold duration) for informative compared to declarative and imperative conditions. Indeed, unlike declarative and imperative pointing gestures, informative pointing gestures are actively driven by the desire to communicatively inform the addressee about information that is assumed to be novel and/or relevant *to them*.

In addition to collecting behavioral data (gesture initiation time, stroke duration, stroke velocity, hold duration), we explored via a data-driven exploratory analysis of event-related potentials (ERPs) whether different types of intent already lead to different neurophysiological activity at the earliest stages of planning the socio-communicative act. Theories of language production have studied in great detail how different stages of internal encoding make use of representations stored in long term memory prior to articulating a communicative signal, but the time-course of the early cognitive differentiation between different *types of intent* at the message conceptualization stage has received less attention (e.g., Levelt, 1989). Because of the substantial increase in the complexity of the visual display that was used in the current study compared to earlier studies, and the overall methodological novelty of our approach, we formed no predictions for specific ERP component amplitude modulations prior to the onset of gesture. Rather, we aimed to explore at what point in time the process of having the intention to *share information and attention* with one's addressee would start to differ from the process of *taking their perspective and current knowledge state* into account, as a neurocognitive basis for potential differences in subsequent gesture kinematics. Purely to facilitate interpretation of the directionality of potential ERP amplitude differences across conditions, we included a non-communicative condition to the experiment in which participants produced pointing gestures in the absence of any salient communicative motivation.

2. Method

2.1. Participants

Thirty-six native speakers of Dutch (mean age = 21.9, age range = 18–30, all female) participated in the experiment. They were all right-handed (Oldfield, 1971) and Dutch was their single native language. They had normal or corrected-to-normal vision and had no history of neuropsychological disorder, dyslexia, or speech problems. They gave written informed consent and received monetary compensation for their participation. Data from 12 additional participants was recorded, but excluded prior to analysis, due to technical problems encountered during the experiment ($N = 9$) or excessive noise in the data ($N = 3$).

Sample size ($N = 36$) was a priori determined by a cumulative frequency distribution-based power analysis in R (R Core Team, 2018) using the effect sizes (η^2) of three relevant measures (mean velocity of the pointing gesture, hold duration of the pointing gesture, and the stimulus-locked P3a ERP effect) reported in the most similar previous

Fig. 1. Visual display of different parts of the experiment in the CAVE. During the experiment, participants first met primary virtual agent Sandra (top left). Next, in Sandra's absence, virtual agent Emily entered the room. She briefly introduced herself to the participant and communicated her preference for healthy food, posh clothing, and antique home lifestyle items (top right). After a subsequent brief chat between Sandra and Emily in which these preferences were not mentioned (bottom left), the main experiment started in a virtual room (bottom right) in which the participant pointed at one of three pictures per trial in different conditions (non-communicative, declarative, informative, imperative). The participant (not depicted here, see Fig. 2) was seated in the middle of the CAVE and wore shutter glasses such that the projected virtual worlds were perceived in 3D.

study in the field (i.e., Peeters et al., 2015). At a significance level of 0.05, power estimates of 0.83, 0.97, and 0.81 were found for a projected sample size of 36 participants. We were therefore confident that this sample size would allow for statistically detecting any potential kinematic and/or electrophysiological effect if reliably present in the data.

2.2. Apparatus

The experiment was carried out in a Cave Automatic Virtual Environment (CAVE) located at the Max Planck Institute for Psycholinguistics in Nijmegen, the Netherlands. This CAVE setup has been described in detail before (Eichert, Peeters, & Hagoort, 2018). It consists of three screens (255 × 330 cm, VISCON GmbH, NeukirchenVluyn, Germany) arranged in right angles (see Fig. 1). During the experiment, each of the three screens was illuminated by two projectors (F50, Barco N.V., Kortrijk, Belgium) via two mirrors placed behind each screen. Two vertically displaced images were projected onto each set of two mirrors, which in turn reflected the projection onto the screen such that the two images overlapped in the middle of the screen.

The CAVE system further made use of 10 infrared motion capture cameras (Bonita 10, Vicon Motion Systems Ltd., UK) for optical tracking. Six of the ten cameras were fixed on the upper edges of the screens, oriented downwards, and four were fixed at the bottom, directed upwards. The orientation of all cameras was towards the center of the CAVE. The motion capture system made use of Tracker 3 software (Vicon Motion Systems Ltd., UK). The cameras tracked the positions of the retroreflective spherical markers (fixed in a pattern on a motion-tracking glove that the participants wore), by optical-passive motion tracking. Auditory input was presented through two speakers (Logitech, US) positioned at the bottom edge of the screen that was facing the participant (henceforth: "the middle screen", see Fig. 1).

Participants sat in the center of the CAVE in a chair, located at 166 cm facing the middle screen, with one additional screen on either side of them. The critical picture stimulus materials (see below) were presented on the middle screen extending across approximately 60° of participants' horizontal visual field. The primary virtual agent Sandra (see below) was rendered on the screen to the left of the participant, (directly adjacent to the middle screen) such that she was seated at an angle of 45° relative to the plane between the participant and the middle screen (see Fig. 1, bottom right panel).

During the experiment, participants wore 3D-glasses (SMI Eye-Tracking Glasses 2 Wireless, SensoMotoric Instruments GmbH, Teltow, Germany) that immersed them into the presented virtual environments. Glasses were equipped with a 60-Hz binocular camera with automatic parallax compensation. Both shutter device and recording interface were placed on a table behind the participants during the experiment. These glasses have been described in detail before (Eichert et al., 2018).

The experiment was monitored from a control room that was situated behind the room with the CAVE system, such that a large window in the wall behind the participant allowed the experimenter visual access to display and participant's performance in the experiment. Once the participant was seated in the CAVE, the control room was not within their field of vision.

Programming of the experiment was done using the 3D application software Vizard (Vizard, Floating Client 5.4, WorldViz LLC, Santa Barbara, CA), built on Python. This software was also used to run the experiment and to record the behavioral data.

2.3. Virtual environments

The virtual environments, all the objects (chair, table, etc.) used in them, and the virtual agents were adapted or created in-house using Autodesk Maya. A total of three virtual scenes were used. The first scene contained a brick wall across all screens, creating the feel of an alleyway, with one entrance to the left and one to the right (see Fig. 1, top panels). This was used as the first environment and was visible on the screens as the participants entered the CAVE. The introduction of the experiment with the two virtual agents (see below) took place in this environment. A second environment was used briefly as a calibration space. It contained one blue sphere, one white sphere, and a yellow sphere in different positions on the middle screen in the CAVE. The spheres were used as targets for hand-tracking calibration before the main experiment began. The experimenter could toggle between the calibration environment and the experimental environment using buttons on the keyboard in the control room computer. The third scene was the main experimental environment, consisting of a carpeted room, with some wooden shelves against the walls (see Fig. 1, bottom right panel). A wooden table was positioned in the center on the middle screen, and three small tablet-like screens were placed on the table, equally spaced. Picture stimuli appeared as triplets on these screens. The primary virtual agent sat on a chair on the left screen. All objects were scaled to realistic sizes and were designed to be neutral in color and appearance.

2.4. Virtual agents

Two female virtual agents, both adapted from stock avatars produced by WorldViz, were used in the experiment. Both agents appeared Caucasians in their late twenties. One agent (based on stock avatar "casual03_f_highpoly") played the role of the primary agent and was given the name Sandra. She was blonde haired and was dressed in casual brown trousers, a dark blouse and a blue cardigan, with casual dull colored shoes (see Fig. 1). As the primary agent, Sandra gave the participants instructions and interacted with them throughout the experiment. The second agent (based on stock avatar "business03_f_25_spec"), who introduced herself as 'Emily', was brown haired and was dressed in a sleek cut blue pantsuit with black heeled shoes. Movements of both agents during the experiment were fully pre-programmed. Sandra's resting facial expression was neutral with a slight smile, while Emily displayed a more open smile during her presence in the virtual space.

All speech of both agents was recorded prior to the experiment from two female native speakers of Dutch who resembled the two agents in age and ethnicity. Recordings were made in a soundproof both, sampled at 44.1 kHz (stereo, 16-bin sampling resolution), and spoken at normal speech rate with natural intonation. The agents' mouth movements were tuned to the amplitude of the sound signal ("lip-sync"), and head and body movements were coordinated to render the behavior of the agents as natural as possible.

2.5. Picture stimuli

Picture stimuli for the main experiment were selected based on the results of a pre-test, reported in Appendix A. The pre-test assured that the triplets of pictures used on the different trials in the experiment were matched within each triplet in terms of visual complexity, overall salience, and degree of familiarity of the depicted objects to the participants.

2.6. Experimental design and procedure

As a cover story that served to hide the goal of the study, participants were instructed that they would take part in an exploratory virtual reality experiment that was aimed at selecting content to create virtual spaces customized for different people with varying preferences. It was portrayed as a series of interactive games to help shed light on this issue. Participants were further instructed prior to the start of the experiment that they would be engaging with two virtual agents named Sandra and Emily. Important in the light of the *informative* condition, Emily was described as a person who liked posh clothing, healthy food, and antique home lifestyle items. During EEG montage, participants familiarized themselves with the picture triplets that were used in the experiment. After EEG montage, a short calibration session for the motion tracking equipment in the CAVE preceded the experiment.

The experiment then started with virtual agent Sandra entering the first virtual environment and introducing herself to the participant (Fig. 1, top left panel). She then briefly left the room under the pretext of completing some last minute checks next door. In Sandra's absence, virtual agent Emily entered the room through a (virtual) door to the right of the participant and looked around the space, while also introducing herself to the participant (Fig. 1, top right panel, and Fig. 2). In a short narrative, she casually mentioned her preference for elegant/stylish clothing, healthy food, and antique home lifestyle items. Once Emily's monolog to the participant had finished, Sandra returned and met Emily (Fig. 1, bottom left panel). They briefly spoke, after which Emily left, and Sandra and the participant would start the actual, main experiment. The rationale behind this procedure was to provide the participant with information about Emily that was not known to Sandra. As such, this manipulation allowed for including an *informative* condition in the current experiment (see below).

The experiment consisted of four main blocks, corresponding to three main experimental conditions and a non-communicative block. It was designed to have a single response mode, i.e., the participant was instructed by Sandra to communicate solely using pointing gestures for the entire experiment. Each block consisted of 45 trials during which the participant manually pointed at one of three pictures that were presented in the virtual environment on three virtual tablet-like screens (see Fig. 1, bottom right panel). Appendix B reports a control experiment that explores the extent to which the behavioral findings reported below were dependent on the response mode.

The first block in the experiment was always the familiarization block, which informally served as the non-communicative point of visual reference for interpretation of the ERP data (see below). Sandra introduced the first block to the participant as a preliminary training phase meant for the virtual reality system to adjust itself to the participant's interaction. She asked the participant to point, for each triplet of pictures, at one of the three pictures at random. Once she had conveyed the instructions, Sandra engaged in reading a book and made no head or eye movement in the direction of the participant during this entire block. As such, this block was considered to represent a non-communicative condition, included purely to aid in visual interpretation of any potential ERP differences in the three communicative blocks.

Because the familiarization block was always the first block in the experiment and designed to be faster than the test blocks to elicit a random response, we will not statistically compare the elicited kinematic data from this block to the kinematic data elicited in the three following communicative blocks. However, we considered the electrophysiological data recorded on each trial prior to gesture onset as an appropriate, informal, non-communicative point of reference for visual comparison to the three communicative conditions.

At the start of the declarative block, Sandra asked the participant to point at the one picture in each triplet of pictures that was the food item (for 15 trials), clothing item (for 15 trials), or home lifestyle item (for 15 items) that the participant herself would prefer. As such, in this block, participants conveyed their personal preference to Sandra. We designed this condition to resemble the expression of declarative intentions in which one shares information and attitudes for the mere sake of sharing. At the onset of each trial, Sandra looked at the participant. She then followed the participant's pointing gesture to one of the three pictures. She subsequently looked back at the participant and nodded to acknowledge that she perceived their choice, after which the next trial started.

In the informative block, Sandra asked participants to point at one image in each triplet of images that was the food item (for 15 trials), clothing item (for 15 trials), or home lifestyle item (for 15 items) that Emily preferred. Note that the participant, unlike virtual agent Sandra, was aware of the (healthy) food, (elegant/stylish) clothing, and (antique) home lifestyle preferences of Emily. This manipulation thus captured two important characteristics of informative intentions – the reliance on theory-of-mind abilities in order to think from another person's perspective, and the lower personal, subjective value of the provided information relative to the expression of a declarative intention. As in the declarative block, Sandra looked at the participant at trial onset, then followed their pointing gesture towards the picture that was pointed at, after which she looked back at the participant and nodded to acknowledge that she registered their choice.

In the imperative block, Sandra asked participants to point at one picture in each triplet of pictures that was the food item (for 15 trials), clothing item (for 15 trials), or lifestyle item (for 15 items) that the participant wanted to take a closer look at. Pressing a button on a small

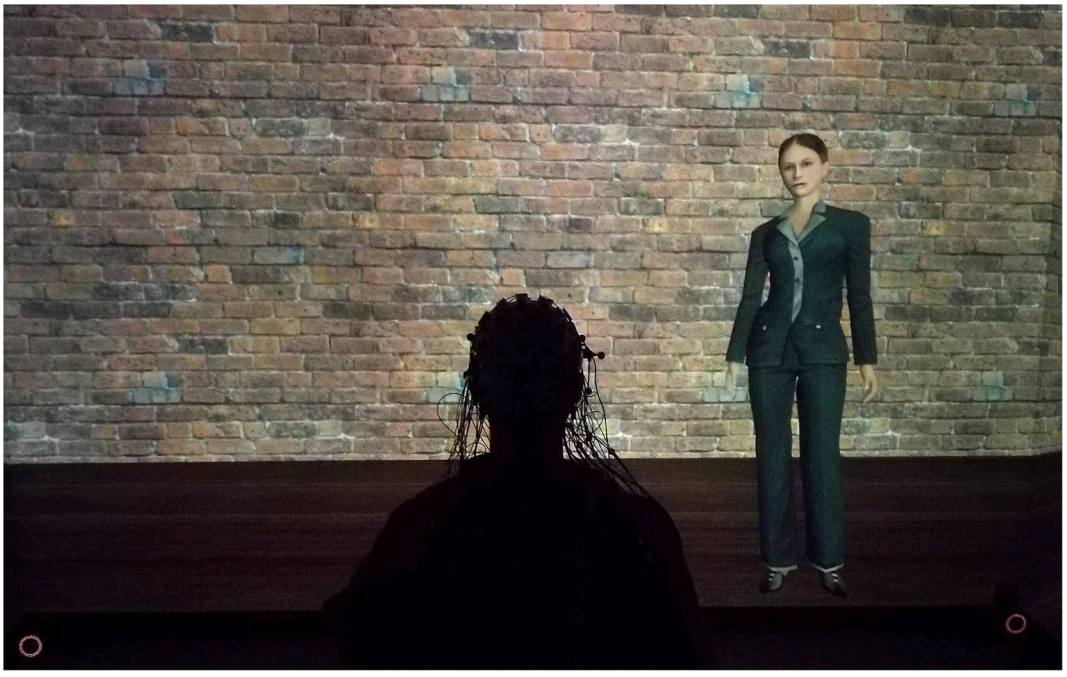

Fig. 2. A participant in the experiment, wearing a 64-channel EEG cap, listening to virtual agent Emily in the CAVE environment. By wearing 3D shutter glasses, participants experienced the projected immersive virtual environments in 3D.

white button box that was located on her left thigh, Sandra would move forward the selected picture towards the participant, after which it would revert to its original position. As such, participants would use Sandra "as a tool" to perform an action at their request. As in the two other communicative blocks, Sandra looked at the participant at trial onset and followed their pointing gesture towards the picture that was pointed at. She then pressed the button, which made the selected picture move forward, after which she looked back at the participant, nodded, and the picture moved back.

In all three communicative blocks, participants' selection of one of three pictures was recorded online by the experimenter. In a Wizard-of-Oz procedure, and unknown to the participant, the experimenter pressed one of three keys on a keyboard to indicate which of the three pictures the participant pointed at. Upon registration of the button press, Sandra's gaze then automatically followed the gesture towards the correct picture. The button press also allowed, in the imperative block, for the correct picture to move forward. Appendix C shows that the latency of the experimenter's button press was stable across conditions and cannot have led to any of the kinematic differences in the participant's behavior reported in the Results section below.

The three communicative blocks were presented in fully counterbalanced order across participants (six participants per order, six orders). The 15 trials in each of the three picture categories (food, clothing, home lifestyle) were always presented in a blocked manner in the same order (i.e., food, clothing, home lifestyle). Before presentation of each set of 15 trials, Sandra repeated the instructions for that block, specified for the type of upcoming picture category (food, clothing, or home lifestyle). The presentation order of triplets within each set of 15 trials was always fully randomized. Also the appearance of the three pictures in each trial was fully randomized across the three tablet-like screens. Participants could take a self-timed break after each block of 45 trials. At the end of the experiment, Sandra thanked the participant for taking part. The experiment lasted on average 35–40 min.

2.7. Kinematic recording and analysis

Kinematic data was recorded continuously throughout the experiment using Vizard (Vizard, Floating Client 5.4, WorldViz LLC, Santa Barbara, CA) and the optical motion-tracking system (see above). The motion-tracking software continuously captured the coordinate positions (3D) of retroreflective spherical markers that were fixed on a motion-tracking glove, which participants wore on their dominant right hand. Raw behavioral data sampled at 62.5 Hz consisted of hand-tracker coordinates relative to the center of the CAVE (0,0,0), the trial and block number, stimulus names, and relative positions of the stimuli.

Prior to kinematic data analysis, we plotted participants' hand movements in space over time for each trial. All trials were manually inspected, and any trial containing two (rather than one) pointing gestures (3,25% of all trials), as well as trials on which there was a technical issue in terms of recording hand position (2,10% of all trials) were removed. One participant was fully excluded from kinematic data analysis due to a technical problem during kinematic recording. In total, we were left with a dataset of 4465 trials equally distributed over the three communicative conditions (35 participants × 45 trials × 3 conditions). As a next step, we calculated the mean value per dependent variable (see next paragraph) per participant and removed outlier trials that were 2.5SD away from the participant's mean on that variable. This resulted in a remaining dataset of 4194 trials for the kinematic analysis (declarative: 1419 trials; informative: 1391 trials; imperative: 1384 trials). Raw data, hand movement plots for all individual trials, preprocessing and data analysis scripts are openly available on OSF (https://osf.io/rqpxf/?view_only=6e19f11f6e8746e3950aa408eadbc5af).

For every participant and every trial, the responses on four kinematic dependent variables were calculated. First, *Gesture Initiation Time* was defined as the interval (in milliseconds) between the visible onset time of the picture triplet on the screens in the virtual environment and the onset of the pointing gesture response within each trial. As such, it corresponds to the time it took participants to start pointing. For each trial and each sample within a trial, we first determined whether the participant's hand moved, and if so, whether it moved forward or backward. To avoid an influence of minor hesitations or inconsistencies in the recording system, we only included sequences of 10 or more samples in which the participant's hand was moving in the same direction (i.e., either forward or backward), with samples corresponding to 16 milliseconds intervals. The temporal interval between the onset of the picture triplet and the onset of the longest monotonic movement forward was then considered as the Gesture Initiation Time on each trial. We opted for this conservative and objective monotonic approach as it does not require one to make a relatively arbitrary judgment of how much a participant's hand is required to move forward for it to be taken as the onset of the gesture.

Second, *Stroke Duration* was defined as the duration of the participant's hand moving forward (in milliseconds) until it reached gesture apex, i.e., the point in time where the longest monotonic movement forward stopped. This measure hence corresponded to the duration of participants' hand moving forward in the direction of the referent.

Third, the mean *Stroke Velocity* was defined as the ratio between the distance the hand travelled between gesture initiation and gesture apex, and the stroke duration (i.e., the time interval between gesture initiation and gesture apex). As such, it corresponds to the average speed of the participant's hand moving forward in the direction of the referent.

Fourth, *Hold Duration* was defined as the interval between gesture apex and the onset of the longest monotonic movement of the participant's hand moving backward in space over time, i.e., the onset of the retraction phase of the gesture. As such, it corresponds to the duration of the participant keeping their hand in apex position.

The definition of these four dependent variables is in line with the definition and use of the same constructs in previous work in this domain (e.g., Chu & Hagoort, 2014; Peeters et al., 2015).

Kinematic data were then further analyzed in R (R Core Team, 2021). Specifically, we used separate mixed effects regression models generated by the lme4 and lmetest packages in R (Bates, Mächler, Bolker, & Walker, 2015; Zeileis & Hothorn, 2002) to compare the three communicative, experimental conditions on the four dependent variables described above (i.e., Gesture Initiation Time, Stroke Duration, Stroke Velocity, and Hold Duration). All regression models had an identical fixed and random effect structure, which included the maximal random effect structure justified by the data and our hypotheses. This included a fixed effect for Experimental Condition (a 3-level categorical variable defined using forward difference coding to make the following contrasts: Declarative vs. Informative, and Informative vs. Imperative), random intercepts for participants, and random by-participant slopes with respect to the effect of experimental condition. The corresponding lme4 model syntax was *Dependent Variable* ~ *Experimental Condition* + *(1+ Experimental Condition | Participant)*.

2.8. Electrophysiological recording and analysis

Each participant's EEG was recorded continuously from 59 active electrodes (Brain Products, Munich, Germany) held in place on the scalp by an elastic cap (Neuroscan, Singen, Germany). Three additional, external electrodes were attached to record participants electrooculogram (EOG) – one below the left eye (to monitor for vertical eye movement/blinks), and two on the lateral canthi next to the left and right eye (to monitor for horizontal eye movements). Finally, one electrode was placed over the left mastoid bone and one over the right mastoid bone. The continuous EEG was recorded with a sampling rate of 500 Hz, a low cut-off filter of 0.01 Hz, and a high cut-off filter of 200 Hz. All electrode sites were referenced online to the electrode placed over the left mastoid and re-referenced offline to the average of the right and left mastoids. Fig. 3 presents the equidistant placement of electrodes over the scalp.

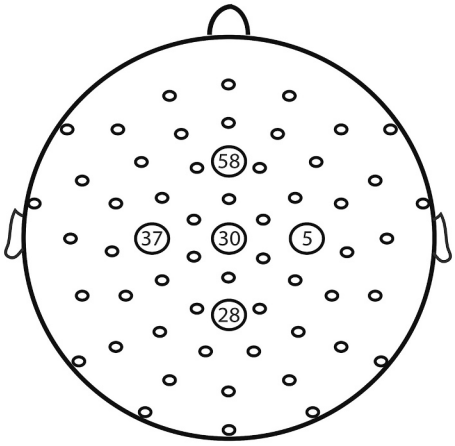

Fig. 3. Equidistant placement of electrodes across the scalp. The channels here marked with a number refer to the channels used for plotting the event-related potential outcomes in Fig. 5.

Fieldtrip (Oostenveld, Fries, Maris, & Schoffelen, 2011) was used for pre-processing and analysis of EEG data. In preparation for a stimulus-locked ERP analysis, continuous EEG was low-pass filtered at 40 Hz. Independent component analysis (ICA) was used on continuous (low-pass filtered) data, and ICA components that corresponded with horizontal eye movements and eye-blink artifacts were removed from the data. A limited number of bad channels were reconstructed using an average of 7 neighboring channels. Finally, trials were filtered using visual artifact rejection, leading to removal of 6.67% of the trials across the entire dataset prior to analysis.

For the ERP analysis, an epoch from 100 ms preceding stimulus onset to 400 ms after stimulus onset was selected. Event-related potentials were truncated at 400 ms post-stimulus onset due to the fastest pointing gesture responses starting to contaminate the EEG data with motor artifacts after this point. The pre-stimulus period of 100 ms was used as a baseline.

The event-related potential data were analyzed using cluster-based permutation tests (Maris & Oostenveld, 2007) on the full epoch (0–400 ms). The cluster-based permutation statistic is a non-parametric, data-driven approach that controls for the family-wise error rate that is bound to arise when an effect is tested at a multitude of temporal and spatial points (Maris & Oostenveld, 2007). It was chosen as it dealt effectively with the multiple comparisons problem arising during the analysis of electrophysiological data (Maris, 2012). The cluster-based permutation tests relied on a dependent samples *t*-test comparing two conditions at every data point – i.e. the signal from each electrode per time point. Clusters were adjacent data points that were grouped together if they exceeded an alpha level of 0.05. The sum of the *t*-statistic in each positive and negative cluster was used in the cluster level statistic. A Monte Carlo method (2000 randomizations, calculating the largest cluster-level statistic for each randomization) was used on a calculated null distribution – an assumption of no difference between pairs of conditions – to compare the clusters against the null distribution. The clusters that crossed the $p < 0.05$ significance threshold were considered significant.

Raw EEG data and analysis scripts are available via the OSF entry for this project (see above for OSF-link).

3. Results

3.1. Behavioral results

Table 1 presents the average Gesture Initiation Time, Stroke Duration, Stroke Velocity, and Hold Duration per Experimental Condition, and Table 2 summarizes the results of the mixed effects regression models used to predict each of the four dependent variables. Fig. 4 shows the coefficients of these models with respect to the main effect of Experimental Condition.

The model showed a significant effect of Experimental Condition on Gesture Initiation Time (Declarative vs. Informative: $\beta = 0.225$, $SE = 0.67$, $t = 3.375$, $p = 0.00186$), indicating that the onset of pointing was significantly later in the Declarative condition compared to the Informative condition. This result means that participants started pointing significantly later in the Declarative condition compared to the Informative condition and, given our coding scheme, by extension also compared to the Imperative condition (see Fig. 4A).

We observed no significant effect of Experimental Condition on Stroke Duration, and no significant effect of Experimental Condition on Stroke Velocity (see Fig. 4B and C)

A significant effect of Experimental Condition was observed on Hold Duration (Informative vs. Imperative: $\beta = 0.289$, $SE = 0.08$, $t = 3.784$, $p = 0.0006$), such that hold durations were significantly longer in the Informative condition compared to the Imperative condition. This result indicates that people kept their index-finger in apex position for a shorter interval in the Imperative condition compared to the Informative condition (and given our coding scheme, by extension also compared to the Declarative condition) (see Fig. 4D).

3.2. Electrophysiological data

Fig. 5 presents the event-related potentials for the Declarative, Informative, and Imperative conditions, time-locked to the onset of the critical picture stimuli, for the earliest stages of gestural planning. These three communicative conditions were compared with one another, while the non-communicative condition was only plotted to visually aid in interpretating the directionality of the effects. First, cluster-based permutation tests on the full epoch (0–400 ms after stimulus onset) showed significantly ($p = 0.023$) enhanced positive amplitude for the Declarative compared to the Imperative condition, which was most pronounced between 190 ms and 348 ms after stimulus onset, and widespread over the scalp (46/59 channels contributed to the effect). Second, cluster-based permutation tests on the full epoch (0–400 ms after stimulus onset) also revealed significantly ($p = 0.031$) enhanced positive amplitude for the Informative to the Imperative condition, which was most pronounced between 220 ms and 334 ms after stimulus onset and relatively wide-spread over the scalp (41/59 channels contributed to the effect). Finally, no statistical differences were observed between the Informative and the Declarative condition (p's $>$ 0.39). Appendix D presents the results of an additional analysis

Table 1
Average gesture initiation time, stroke duration, stroke velocity, and hold duration per condition in the experiment. Values between parentheses represent standard deviations.

Condition	Gesture initiation time	Stroke duration	Stroke velocity	Hold duration
Declarative	$M = 1244$ ms ($SD = 93$)	$M = 1272$ ms ($SD = 63$)	$M = 0.134$ m/s ($SD = 0.07$)	$M = 1122$ ms ($SD = 108$)
Informative	$M = 1030$ ms ($SD = 82$)	$M = 1211$ ms ($SD = 56$)	$M = 0.143$ m/s ($SD = 0.09$)	$M = 1054$ ms ($SD = 103$)
Imperative	$M = 1037$ ms ($SD = 77$)	$M = 1213$ ms ($SD = 61$)	$M = 0.141$ m/s ($SD = 0.08$)	$M = 790$ ms ($SD = 107$)

Table 2
Outcome of regression models comparing the declarative, imperative, and informative conditions across the four kinematic measures. Significant differences are marked in bold.

Dependent Variable		Estimate	Std.Error	df	*t* value	*p* value
Gesture Initiation Time	(Intercept)	1.11268	0.07021	33.92485	15.84883	0.00000
	Declarative vs. Informative	0.22488	0.06663	33.98790	3.37524	**0.00186**
	Informative vs. Imperative	−0.02434	0.08199	34.18737	−0.29684	0.76838
Stroke Duration	(Intercept)	1.22338	0.05380	33.99108	22.73763	0.00000
	Declarative vs. Informative	0.05208	0.03983	34.29994	1.30763	0.19970
	Informative vs. Imperative	0.00832	0.04234	33.23499	0.19653	0.84539
Stroke Velocity	(Intercept)	0.13713	0.00886	33.99632	15.48484	0.00000
	Declarative vs. Informative	−0.00771	0.00663	34.14856	−1.16311	0.25285
	Informative vs. Imperative	0.00258	0.00782	34.21791	0.33053	0.74302
Hold Duration	(Intercept)	0.98470	0.15710	34.00642	6.26787	0.00000
	Declarative vs. Informative	0.05520	0.06906	34.00164	0.79933	0.42965
	Informative vs. Imperative	0.28932	0.07645	34.07298	3.78450	**0.00060**

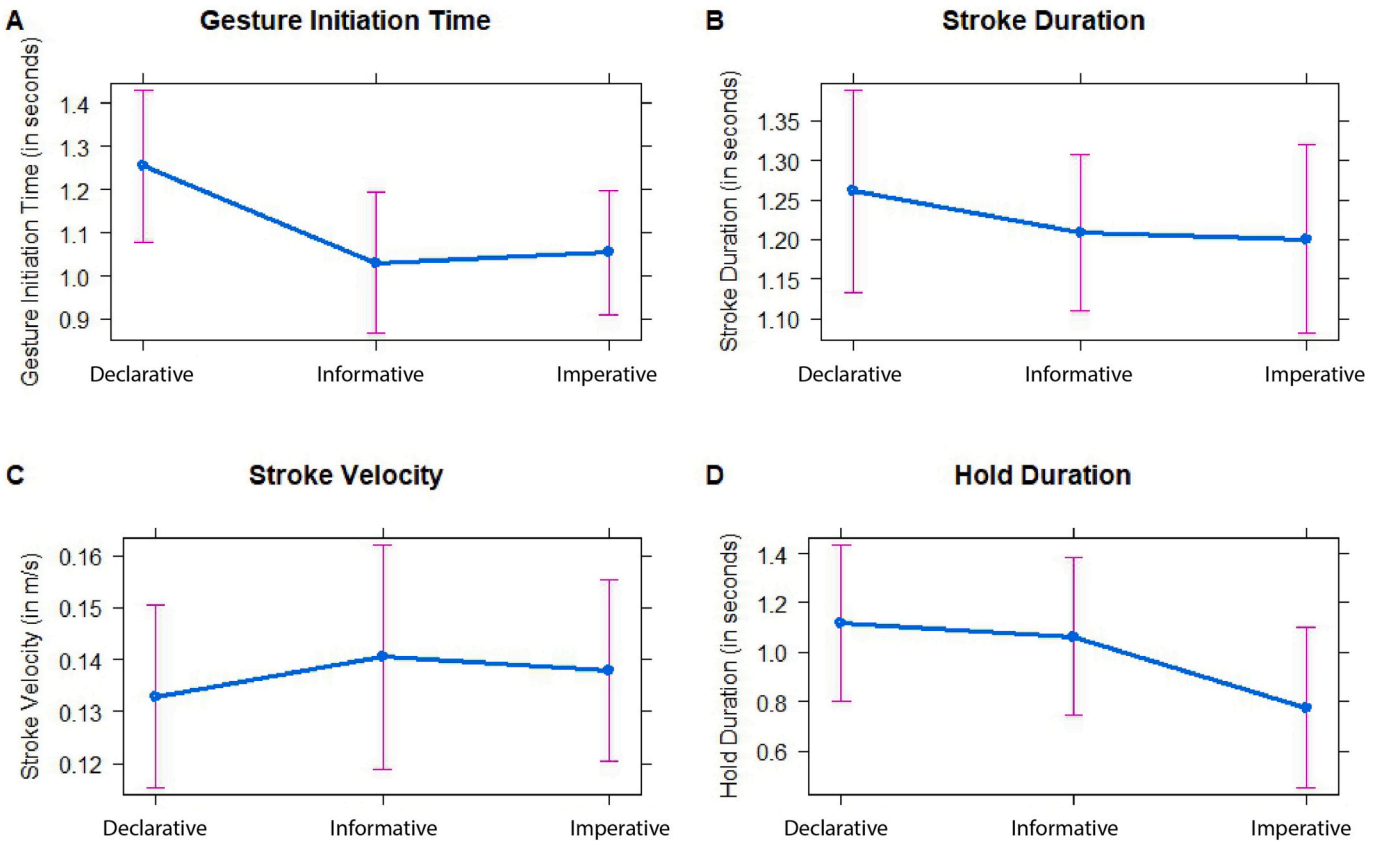

Fig. 4. The regression model's coefficients for the main effect of experimental condition for participants' (A) gesture initiation time, (B) stroke duration, (C) stroke velocity, and (D) hold duration. Error bars represent the models' standard error.

correlating the kinematic and the electrophysiological data.

In sum, as illustrated in Fig. 5, the Imperative condition (pink line) yielded an electrophysiological response that was significantly less positive in amplitude compared to both the Declarative (red line) and the Informative condition (green line). Visually, compared to these latter two conditions, it moved closer towards the non-communicative condition (blue line).

4. Discussion

Human communication is first and foremost a joint action in which the intention to communicate initiates the exchange of meaningful streams of words, gestures, and facial expressions (Bosker & Peeters, 2021; Clark, 1996; Holler & Levinson, 2019; Levelt, 1989; Perniss, 2018; Vigliocco, Perniss, & Vinson, 2014). As such, we have the capacity to translate our thoughts into multi-faceted messages that can be conveyed through a variety of channels (e.g., mouth, face, hands). Pointing gestures are a core component of our capacity to communicate as they allow for bringing a person, object, or event into our addressee's focus of attention, in the presence or absence of concomitant speech (Cooperrider, 2020; Kita, 2003; Tomasello, 2008). Interestingly, this seemingly simple hand gesture may be used to express a variety of different types of underlying intentions. In this study, we investigated whether and how speakers modulate the kinematic parameters of their pointing gesture as a function of their (declarative, informative, and imperative) intentions and whether these different types of intentions translated into distinguishable electrophysiological activity prior to the onset of the gestural movement.

In a nutshell, we observed that different types of socio-communicative intentions are associated with different

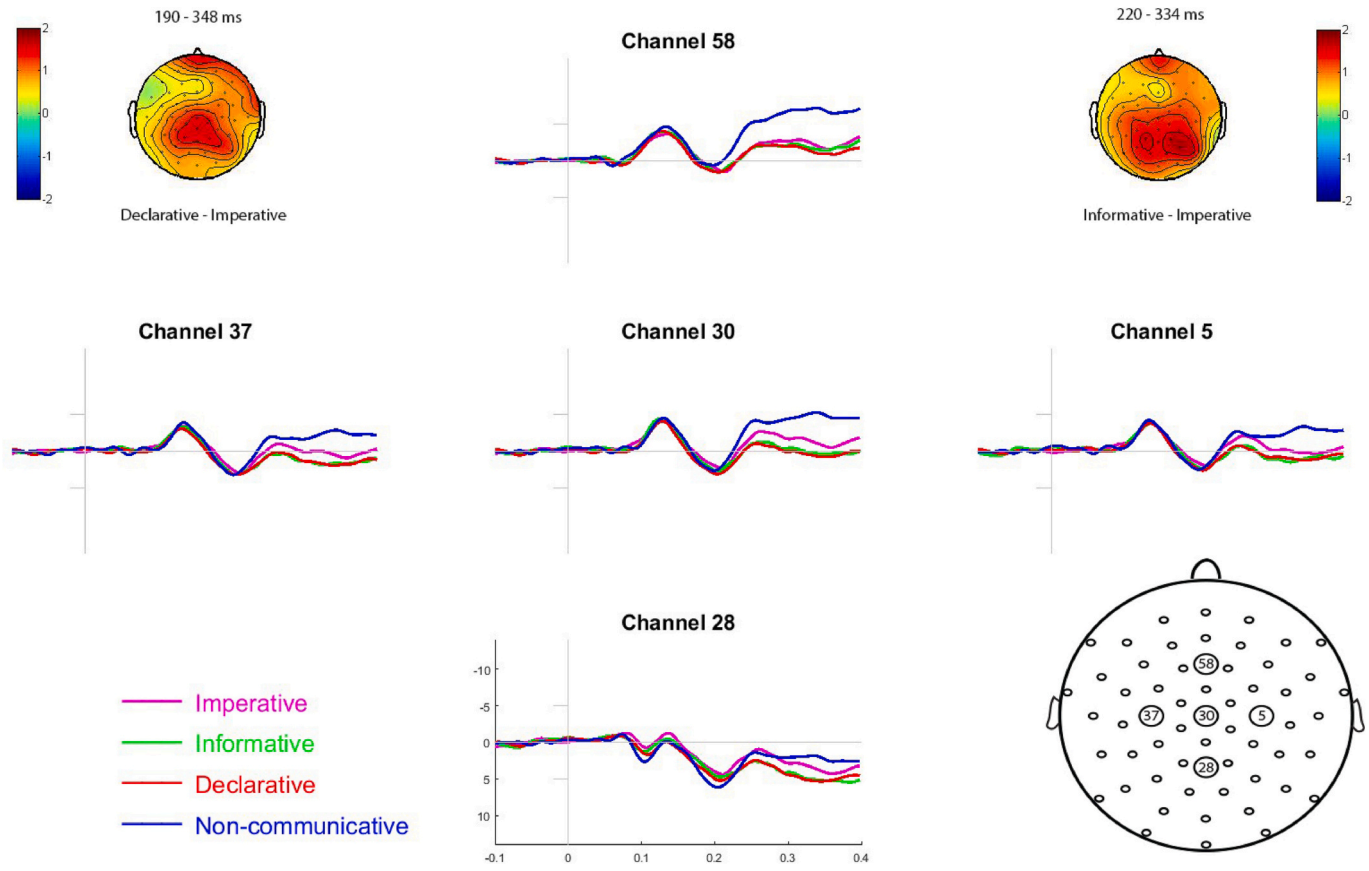

Fig. 5. Event-related potentials time-locked to the onset of stimuli across the three critical (declarative, informative, imperative) communicative conditions, and the non-communicative condition for visual reference, across five electrode channels. The image in the bottom right panel indicates the position of these five channels across the scalp. The topoplots in the top left and top right panel show that the observed differences were widespread over the scalp.

electrophysiological and kinematic markers at different stages during the planning and production of the gesture. In early planning stages, both declarative and informative intentions yielded a similar enhanced electrophysiological positivity compared to the imperative condition. More specifically, imperative intentions led to an electrophysiological response that was visually relatively closer to the brain response observed when participants prepared a largely non-communicative pointing movement. As such, at very early stages prior to the performance of a socio-communicative act, the brain is here found to make an initial distinction between a movement that is used to *share information and attention* with another person (as in our declarative and informative conditions) versus a movement that mainly uses that person as a means to perform an action (as in our imperative condition). Our data show that the first distinction the brain makes here is based on whether the upcoming gesture will either provide its addressee with information about a to-be jointly attended referent or be produced as an imperative instruction for the addressee to act.

Our ERP findings are conceptually in line with an evolutionary view that proposes that *imperative* pointing gestures may have phylogenetically derived from (initially non-communicative) reaching movements, for instance through a process of ontogenetic ritualization (Cochet & Vauclair, 2010; Tomasello & Call, 1997; Tomasello, Carpenter, & Liszkowski, 2007). Based on the observed differences in electrophysiological activity at early stages before the onset of the movement, we would very tentatively suggest that planning declarative and informative gestures might be more taxing for the theory-of-mind network prior to the start of the movement, as these two conditions require to take into account the mental knowledge state of one's addressee when sharing attention and information, whereas planning an imperative or non-communicative gestural movement typically does so only to a smaller extent. This is not to say that declarative and informative communication is uniquely human, as recent evidence suggests that both declarative and imperative communication is also present in great apes, for instance in their manual pointing behavior, both in captivity and in the wild (Krause, Udell, Leavens, & Skopos, 2018).

At subsequent stages of planning and performing the gestures, we found two notable behavioral differences across conditions. First, with respect to gesture initiation time, participants took more time before starting to execute their gesture when it had a declarative intention as opposed to when it had an informative or imperative intention. Earlier work has suggested that a prolonged gesture initiation time may reflect a higher cognitive load on the part of the communicator (Murillo Oosterwijk et al., 2017). This account hence raises the question to what extent our observed difference in gesture initiation time may be due to differences in task difficulty across the three conditions rather than to participants' differences in intentions. As we observed no difference in response times between a declarative and an informative condition in a non-pointing control experiment (Appendix B) that was otherwise identical to the main experiment reported here, it seems that differences in task difficulty and cognitive load cannot account for the observed effect. Likewise, the results of the control experiment render it unlikely that potential differences in perceptual stimulus processing or differences in reference selection difficulty across conditions may explain the prolonged duration of the stage prior to gesture onset in the declarative condition. Below, we will tentatively suggest that deliberately *not* moving the hands, prior to and during the gesture, may rather reflect a communicative signal in itself.

Second, with respect to the hold duration of the gestures, participants held their pointing finger still for a longer period of time when pointing declaratively or informatively compared to when the gesture

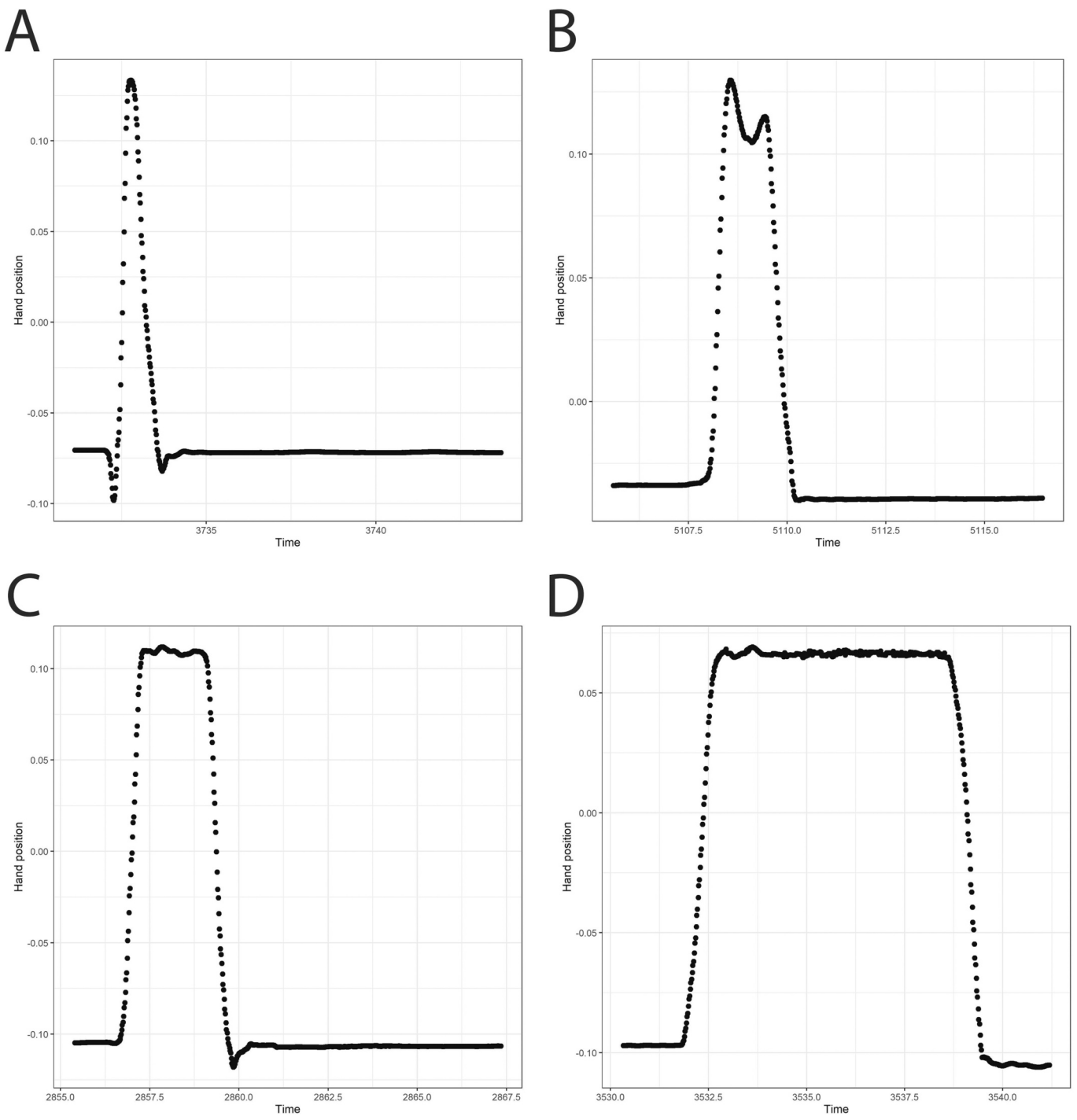

Fig. 6. Illustration of some of the variability in pointing gesture kinematics across four different trials in our overall dataset. Time is plotted on the x-axis, while the y-axis represents the position in space of the pointing index-finger moving forward and backward. These trials illustrate that some pointing gestures may be retracted directly when reaching apex (panel A), while others may reach apex twice (panel B) or show a relative short (panel C) or long hold duration (panel D).

was imperative in nature (see also Fig. 6). Although different intentions led to gestural strokes that had similar mean durations and velocity across conditions, we thus observed that once the gestures reached their apex position, participants waited significantly longer to retract their arm when they had a declarative or informative intention compared to when they had an imperative intention. As such, our participants allowed their (virtual) addressee more time to look at their extended arm and index-finger, and thus at the vector it created towards a specific referent, when they were sharing information about themselves (as in the declarative condition) or about somebody else (as in the informative condition) compared to the (imperative) situation in which they merely aimed to use their addressee as a means to carry out an action. Interestingly, also the analysis of toddlers' pointing behavior has shown that "declarative gestures lasted longer than imperative ones, which might reflect infants' wish to maintain interactions in the declarative situation" (Cochet & Vauclair, 2010, p. 438). Our findings hence complement and confirm these observations from an adult population perspective. It remains to be seen as to what extent these observations are driven or modulated by differences across conditions in how difficult it was to select which given referent to point at (see Appendix B and below).

Broadly speaking, our results confirm the general observation that people modulate the kinematic properties of their pointing gestures as a function of their socio-communicative intentions (Chu & Hagoort, 2014; Cleret de Langavant et al., 2011; Liu et al., 2019; Murillo Oosterwijk et al., 2017; Peeters et al., 2015). While previous behavioral work in adults compared pointing gestures across situations that were deemed either more or less communicative (Cleret de Langavant et al., 2011; Liu et al., 2019; Murillo Oosterwijk et al., 2017; Peeters et al., 2015; Winner et al., 2019), the present study provides more fine-grained insights about how different theoretically motivated types of socio-communicative intentions (declarative, informative, imperative) lead to pointing gesture that have a different kinematic profile over time. Indeed, while the gesture and action literatures suggest that movements may be exaggerated when they are explicitly used to communicate or demonstrate something to someone (McEllin et al., 2018; Sacheli et al., 2013; Trujillo et al., 2018; Vesper & Richardson, 2014), our study provides further subtlety to this claim. Specifically, when comparing our results across the four behavioral measures of interest, it becomes clear that the underlying intent of our participants translated not so much into the *movement* parameters of their pointing gestures (e.g., the duration and velocity of the stroke), but rather into the *stationary* properties of their gestures (i.e., how long they kept their arm and finger still).

Indeed, behavioral timing differences were observed both prior to the start of the movement (as reflected in the differences in gesture initiation time across conditions) and when the hand had reached its apex position (as reflected in the differences in hold duration across conditions). Thus, at least in the case of pointing gestures produced in the lab, specific types of theoretically motivated socio-communicative intentions translated into differences in the duration of the interval when the hands were actually *not* moving. These findings have consequences for theories of communication that posit that addressees may use gestural *movement* properties to derive a speaker's intent, as they naturally raise the question of whether addressees may use not only motion parameters to aid in deriving a speaker's intention, but also the duration of the intervals preceding and during the execution of the gesture when the hand and finger are deliberately not in motion. This idea is in line with the study of other communicative cues in face-to-face communication. For instance, it has been observed that also the duration of *eye blinks* can be used to provide one's interlocutor with meaningful information, for example about the extent to which an incoming message has been understood. Indeed, also in the case of blinking it seems that communicative information is conveyed through variability in the duration of keeping the eyes closed and as such via the *non-moving* part of the signal (Hömke et al., 2018). Similarly, silent *pauses* in ongoing speech may have meaningful communicative implications (Rochester, 1973). In general terms, one could argue that the temporary absence of a dynamic communicative signal in an ongoing stream of meaningful information can be taken to be a meaningful communicative signal in itself. Indeed, the prolonged temporal duration, "by virtue of it being longer than necessary and thus instrumentally dysfunctional, ostensively marks the action as communicative" (Liu et al., 2019, p. 20). A future perception experiment may go beyond existing work and investigate to what extent addressees attempt to derive their interlocutor's intent not just from the moving but also from the non-moving stages of their pointing gestures. In addition, our findings can be taken as an encouragement for the broader (non-communicative) action literature to move beyond a focus on *movement informativeness* and also theoretically and experimentally consider the stationary phases prior to and during a movement in light of the actor's intent (Becchio et al., 2012; Koul, Soriano, Tversky, Becchio, & Cavallo, 2019).

The present study extends previous work in avoiding the relatively artificial manipulation of eliciting "more communicative" versus "less communicative" or "non-communicative" pointing gestures in lab settings (Cleret de Langavant et al., 2011; Liu et al., 2019; Murillo Oosterwijk et al., 2017; Peeters et al., 2015; Winner et al., 2019) by focusing on three types of socio-communicative intentions that we know people have when pointing in everyday life (Tomasello, 2008). Nevertheless, it must be acknowledged that we operationalized declarative, informative, and imperative intent in one specific way in our experiment, and other operationalizations would have been possible. For instance, while our imperative condition created a (cognitively relatively straightforward) situation in which the participant used their interlocutor purely "as a tool" to perform an action, in some everyday imperative situations a substantially higher amount of perspective taking and mind-reading may take place. For instance, when asking someone via a pointing gesture to open a window, one may consider from the other person's perspective whether that person is actually capable and willing of doing so at that given moment. Likewise, before pointing at the salt shaker to receive it from one's addressee, one may consider from their perspective whether they would perhaps like to first season their food themselves. These considerations make clear that the distinction between different types of (declarative, informative, imperative) intent is not always clear-cut, that different types of intent may be combined in the same act, that participants' task of selecting a referent in our imperative block may have been cognitively slightly easier than in the other two blocks (cf. Appendix B), and that future work may therefore investigate the kinematic consequences of different types of intent *within* in addition to *across* declarative, informative, and imperative situations.

In everyday life, pointing gestures often occur in the presence of concomitant speech (Cooperrider, 2020; Enfield et al., 2007; Kendon, 2004). In this respect, our results align well with a general framework of multimodal demonstrative reference that theoretically analyzed human manual pointing behavior in the context of demonstratives such as *this* and *that* (Peeters, Krahmer, & Maes, 2021). First and foremost, this framework describes how different types of physical, psychological, and referent-intrinsic variables may jointly and concurrently influence whether a speaker will use one type of demonstrative (e.g., *this*) or another (e.g., *that*) in their spoken referential expression. For instance, the physical variable *visibility of the referent* may influence the choice of referring expression, in that speakers of English use more proximal demonstratives (e.g., *this*) for referents that are visible compared to for referents that are invisible (Coventry, Griffiths, & Hamilton, 2014). The psychological variable *joint attention* influences what type of demonstrative a speaker of Turkish will use, in that the demonstratives *bu* or *o* are commonly preferred when reference is made to entities that are in joint attention already, whereas *şu* is typically used for referents that are not yet in joint attention between speaker and addressee (Küntay & Özyürek, 2006). Also referent-intrinsic variables, such as a referent's relative size, may influence whether speakers use one demonstrative form or another (Rocca, Tylén, & Wallentin, 2019).

In line with this general framework, the same types of physical, psychological, and referent-intrinsic variables may influence the form

and kinematics a pointing gesture takes (Peeters et al., 2021). For instance, in various speech communities, the physical variable *location of the referent* may influence whether speakers point with their thumb (e.g., for referents behind them; Kendon & Versante, 2003), index finger (e.g., for referents in front of them), or whole hand (e.g., for invisible referents when giving directions; Flack, Naylor, & Leavens, 2018). The current study suggests that the specific socio-communicative intention a person has will play a role at the psychological level of this framework in shaping the hold duration of the pointing gesture. Future work may explore whether the type of spoken demonstrative people use (e.g., *this* vs. *that*) or their acoustic properties (such as indeed their duration) are also modulated as a function of a speaker's underlying intentions (cf. Yoon & Brown-Schmidt, 2019).

Finally, the current study adds to a growing body of literature that successfully uses immersive virtual reality to experimentally study aspects of human communication in the lab (e.g., Pan & Hamilton, 2018; Peeters, 2020; Tromp et al., 2018; Zappa et al., 2019), and shows that it is methodologically feasible and valuable to concomitantly record kinematic and electrophysiological data in rich and dynamic 3D environments. This virtual reality approach circumvents the intrinsic limitations human confederates have in replicating the subtleties of their verbal and non-verbal behavior across participants and research labs by using virtual agents instead (Kuhlen & Brennan, 2013; Pan & Hamilton, 2018; Peeters & Dijkstra, 2018). Indeed, we here showed that it is possible to establish a "referential triangle" between human participant, virtual addressee, and virtual referent, and that participants adapt their behavior to the assumed mental state of their virtual addressee. These findings confirm earlier work showing that experimental participants treat virtual interlocutors as if these are real (e.g., Heyselaar, Hagoort, & Segaert, 2017; Peeters, 2019). It is perhaps not surprising that participants may ascribe mental states to life-size three-dimensional virtual agents, as they typically even ascribe agency and mental states to animated two-dimensional circles and triangles moving around on a screen (Heider & Simmel, 1944). Future work may build on the current study and continue to re-create naturalistic situations in virtual and immersive lab environments to investigate the fundamental question of how our intentions shape our actions.

CRediT authorship contribution statement

Renuka Raghavan: Conceptualization, Methodology, Formal analysis, Investigation, Data curation, Writing – original draft, Visualization. **Limor Raviv:** Formal analysis, Data curation, Visualization, Writing - review & editing. **David Peeters:** Conceptualization, Methodology, Formal analysis, Writing – original draft, Writing – review & editing, Visualization, Supervision, Funding acquisition.

Declaration of Competing Interest

The authors have no competing interests to declare.

Data availability

Data and scripts are available on OSF via the link provided in the article.

Acknowledgments

The authors would like to thank Asli Özyürek for valuable advice during early stages of this project, Birgit Knudsen and Janniek Wester for help in collecting the data for the control experiment, and Albert Russel for technical support. DP was supported by a Veni grant (275-89-037) awarded by *De Nederlandse organisatie voor Wetenschappelijk Onderzoek* (NWO, the Dutch Research Council). The funding source had no involvement in the study.

Appendix A Pre-test

The goal of the pre-test was to select picture triplets, to be used in the main experiment, that were matched for salience, visual complexity, and familiarity of the depicted entities to the participant.

By matching stimuli on these measures within each triplet, we avoided that any choice made by a participant in the main experiment was driven by intrinsic stimulus differences rather than by our manipulation of interest (i.e., type of socio-communicative intent).

Pre-test participants

Fourteen participants (mean age = 23.9, age range = 18–30, all female), who did not take part in the main experiment, were recruited to participate in the pre-test. They were right-handed, native speakers of Dutch, and born in the Netherlands. They reported no history of neuropsychological conditions, dyslexia, or speech problems, and had normal or corrected-to-normal vision. They gave written informed consent and received monetary compensation for their participation. Data from two of these participants was not analyzed, because they did not finish the pre-test within the allotted time of 90 min.

Pre-test stimuli

Stimulus pictures were selected from online sources. Three stimulus categories were defined: food, clothing, and home lifestyle. For each category, fifteen representative types of entity were chosen. For instance, pictures falling under the clothing category included a skirt, shoes, or a dress. Food pictures included for example a pizza, pasta, or a type of dessert. Home lifestyle pictures included for instance a chair, a clock, or a sofa.

For the main experiment, we aimed for 45 trials per condition consisting of 15 trials per stimulus category. The pictures present in each experimental condition were the same. For the pre-test, three alternative triplets were selected for each trial in the main experiment. For example, for the food item "pizza", there were three triplets of three different pizzas, out of which we aimed to select one triplet for the main experiment. This led to 135 trials in the pre-test.

Pre-test design and procedure

After providing informed consent, participants were handed a response booklet and entered a soundproof experimental booth. Pre-test stimuli were presented to each participant on one of two laptops (HP, screen resolution = 1920 × 1080, refresh rate = 60 Hz). A little above eye level, on the wall in front of them, an illustration of the rating scale was provided for reference. Each participant was allowed a maximum of 90 min to complete the

pre-test.

During each of the 135 trials, a unique combination of three pictures was presented on a computer screen. Participants were instructed to carefully inspect each picture triplet and answer a number of questions in a response booklet. They first rated each picture in each triplet for familiarity (by indicating on a Likert scale of 1–5 how often they had encountered the item in the picture in the past) and visual complexity (by indicating on a Likert scale of 1–5 how visually complex and detailed they found the item in the picture). For each triplet, they then indicated whether one of the three pictures appeared to be more salient, by answering (yes vs. no) whether one of the pictures stood out. If they responded yes, they were further asked to indicate which of the three images was most salient, by indicating the position of the image (left, middle, right). For each trial, participants were then asked to provide a single name to describe the three items in a triplet. This allowed us to verify whether participants considered the three tokens of each type of entity (e.g., a pizza) as indeed representative of that entity. Finally, participants were asked for the 45 triplets of food items whether one of the items in the pictures could clearly be considered healthy, and if so, which one. For the 45 triplets of clothing items, they were asked whether one of the items in the pictures could be considered posh/stylish, and if so, which one. For the 45 triplets of home lifestyle items, they were asked whether one of the items in the pictures could be considered antique, and if so, which one.

Pre-test analysis

In a step-wise procedure, we selected the best 45 picture triplets (15 per category) for the main experiment from the 135 triplets used in the pre-test. First, only triplets were retained for which >8 out of 12 participants showed consensus on that there was one picture in the triplet that identified as most healthy (for food), most posh/stylish (for clothing), or most antique looking (for home lifestyle). This left us with 122 triplets. Next, we removed triplets for which at least 50% of participants indicated that one of the three pictures was more salient than the other two. For the remaining 93 triplets, we then averaged the familiarity and visual complexity ratings across participants. For each unique picture, the deviation from the overall mean in familiarity and visual complexity was calculated. For both familiarity and visual complexity, the sum of the deviation values of the three pictures within each triplet was calculated. Lower values here are indicative of larger overlap in terms of familiarity and visual complexity within a triplet. Therefore, for each of the 45 trials in the main experiment, we selected the triplet (out of a maximum of three candidate triplets) with the smallest sum value in terms of familiarity and visual complexity as measured in the pre-test. This led to 45 triplets of pictures to be used in each condition of the main experiment.

Appendix B Control experiment

To what extent are the observed differences across the three conditions indeed due to participants' *intent* and not to any other potential task-intrinsic differences across the three experimental blocks? Although the order of blocks was counterbalanced across participants and the visual stimuli were the same in every condition, the same pictures could have been *perceived* differently as a function of the social affordances of the pictures in the block at hand. As such, any differences that we attribute to the relation between intent and action might be influenced by differences in perception. In addition, the decision participants had to take (i.e., which referent to select) in each block is substantially different: In the declarative block a referent is selected based on a personal preference, in the informative block a choice is made based on the stimuli's intrinsic characteristics in relation to another person's preferences, and in the imperative block a referent is selected for further close inspection. Although one could argue that these differences may form an intrinsic part of what it means to have a specific type of (declarative, informative, imperative) intention, it is equally fair to assume that these different tasks are supported by cognitive operations that may differ in their inherent difficulty. To establish to what extent the differences in Gesture Initiation Time, the observed electrophysiological effects, and potentially even the observed Hold Duration differences in the main experiment may not have to do with the relation between intent and *pointing*, but rather with different task difficulties and their downstream consequences across the three conditions, we carried out a control experiment that was very similar to the main experiment but required participants to select a picture referent by pressing a button on a button box rather than by manually pointing at it.

Control Experiment

Method

Participants. Thirty-six native speakers of Dutch (mean age = 23.0, age range = 19–30, all female), who did not take part in the main experiment, participated in the control experiment. As in the main experiment, they were all right-handed (Oldfield, 1971) and Dutch was their single native language. They had normal or corrected-to-normal vision and had no history of neuropsychological disorder, dyslexia, or speech problems. They gave written informed consent and received monetary compensation for their participation.

Design, stimuli, apparatus, procedure. The control experiment was identical to the main experiment in design, stimuli, apparatus, and procedure except for one important difference in instruction and apparatus used. Participants in the control experiment were not asked to select one of the three visual picture stimuli on every trial by pointing at it, but rather by pressing one of three buttons on a three-button button box. They were instructed that the left button corresponded to the left picture, the middle button to the middle picture, and the right button to the right picture presented in the virtual environment. This set-up made the Wizard-of-Oz procedure redundant (see Appendix C). As in the main experiment, a non-communicative familiarization block was followed by the three target blocks (declarative, informative, imperative) that were presented in counterbalanced order across participants. We analyzed whether response times (RTs) as recorded by the button box would differ across the three conditions using a procedure and statistical model identical to the Gesture Initiation Time analysis reported in the main text. Across blocks, the response deadline was set to 2 s to be able to elicit responses in a similar time-window compared to the Gesture Initiation Times elicited in the main experiment, allowing for a fair comparison.

Results

The collected raw dataset consisted of 4860 data points (36 participants × 3 conditions × 45 trials). After removal of trials on which no button press was recorded before the response deadline, a dataset of 4110 data points entered the statistical analysis. Numerically, the Declarative condition (M = 1414, SD = 0.30) and the Informative condition (M = 1400, SD = 0.30) elicited a substantially longer RT compared to the Imperative condition (M = 1336, SD = 0.31), which was confirmed by the statistical model reported in Table A1.

Table A1
Outcome of the regression model statistically comparing the log RTs across the Declarative, Imperative, and Informative conditions. The significant difference is marked in bold. Result pattern did not differ when RTs (rather than log RTs) were used as the dependent variable.

Dependent variable		Estimate	Std.Error	df	*t* value	*p* value
Reaction Time	(Intercept)	0.31423	0.02075	34.95828	15.14440	0.00000
	Declarative vs. Informative	0.01520	0.01856	35.39973	0.81908	0.41822
	Informative vs. Imperative	0.04788	0.01811	35.23858	2.64443	**0.01214**

Combining the Gesture Initiation Time data from the main experiment with the button press RT data from the control experiment into one dataset allowed for carrying out an overall regression analysis that included Experiment (Main experiment, Control experiment) and Experimental Condition (Declarative, Informative, Imperative) in the same statistical model. A significant interaction effect between Experiment and Experimental Condition (contrast: Declarative vs. Informative) confirmed that the control experiment yielded an RT pattern that was statistically different from the Gesture Initiation Time results from the main experiment (β = −0.203, SE = 0.07, t = −2.860, p = 0.00558) in the comparison of the Declarative to the Informative condition. The absence of a significant interaction effect (β = 0.086, SE = 0.08, t = 1.018, p = 0.31) between Experiment and Experimental Condition (contrast: Informative vs. Imperative) indicated that the results from the two experiments did not statistically differ in the comparison of the Informative to the Imperative condition.

Conclusion

The results from the control experiment indicate a significantly shorter RT in the Imperative condition compared to the Informative condition, and given our coding scheme, also compared to the Declarative condition. No difference in RT was observed between the Declarative and the Informative condition in the control experiment. These outcomes mean that (i) the difference in Gesture Initiation Time in the main experiment between the Declarative and Informative condition cannot be explained by a difference in stimulus perception or task difficulty, as these two conditions yielded statistically and numerically similar RTs in the control experiment, which was confirmed by an overall analysis statistically comparing Gesture Initiation Times (main experiment) and RT data (control experiment) in the same regression model, and (ii) the RT result pattern in the control experiment to some extent matches the result pattern for the Hold Duration results in the main experiment. The implications of these findings are further discussed in the main text.

Appendix C Control analysis trial length

During each experimental session in the main experiment, the experimenter monitored the participant via a window from an adjacent control room behind the virtual reality lab in which the experiment took place. On every trial in the main experiment, the participant pointed at one of three picture stimuli, after which virtual agent Sandra looked at that stimulus. To make sure Sandra would always look at the stimulus that was pointed at by the participant, unbeknownst to the participant, the experimenter by button press indicated which exact stimulus (left, middle, or right) the participant pointed at in a Wizard-of-Oz set-up. On every trial, which exact stimulus the participant pointed at became gradually clear during the stroke phase of the participant's pointing gesture, and was fully clear when apex was reached and during the hold phase of the gesture. As such, the information required for the experimenter to press the correct button became increasingly clear over time during each trial. The use of this procedure came at the risk of the participant's Hold Duration being affected by the latency of the experimenter's button press, as the onset of virtual agent Sandra's gaze shift towards the stimulus depended on it. In theory, if the participant would wait to retract their arm until the moment virtual agent Sandra looked at the stimulus that the participant pointed at, and the experimenter's button press latency would differ across experimental conditions, the Hold Duration differences observed in the main experiment could have been confounded by the button press latency. To be able to rule out this potential confound, we analyzed the experimenter's button press latency across the three experimental conditions.

Figure A1 below depicts for every trial and every participant the duration between gesture onset and the end of each trial, corrected for built-in trial length differences across the three conditions. If this duration is constant across the three conditions, the kinematic properties of the participant's gesture cannot be influenced by the timing of the experimenter's button press, as it on average must have taken place at the same moment in time. Numerically, differences in average trial duration (here indicated in seconds) and variability (here indicated by the standard deviations and depicted in Fig. A1) across the declarative (M = 9.60, SD = 0.92), informative (M = 9.72, SD = 0.92), and imperative condition (M = 9.77, SD = 0.86) seemed negligible. Indeed, a linear mixed effects model, identical to the statistical models reported on the four dependent variables in the Results section but taking the current trial length duration as its dependent variable, did not yield any significant differences across the conditions (p's > 0.05). If anything, the time between gesture onset and the end of the trial was numerically slightly longer in the imperative condition compared to the other two conditions, while the observed Hold Duration in the data was actually shortest in the imperative condition. Together, these observations rule out that the timing of the experimenter's button press influenced the kinematic parameters of the participants' gestures.

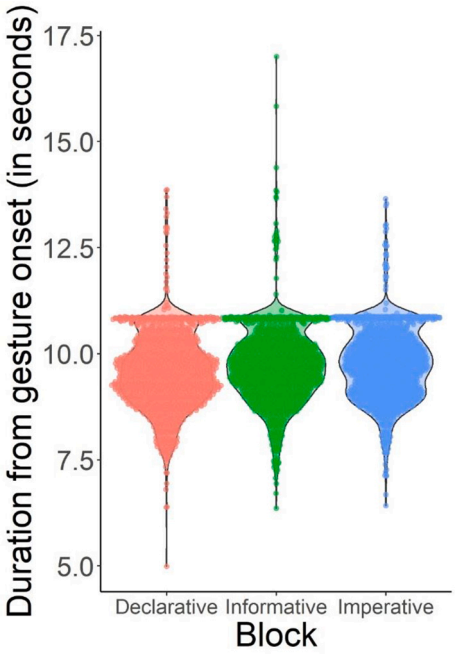

Fig. A1. Similar trial durations between gesture onset and the end of a trial across the three conditions.

Appendix D Brain-behavior correlations

As Stroke Duration and Stroke Velocity did not differ across the three conditions in the main experiment, we can be relatively sure that the two electrophysiological effects observed do not simply correspond to sensorimotor planning of different movements, regardless of participants' intent. As such, it is more likely that the observed electrophysiological differences reflect communicative/intentional rather than sensorimotor planning activity. But to what extent do the electrophysiological effects correlate with the kinematic effects observed in the present study?

There are many ways in which the multidimensional ERP data could be correlated with the multidimensional kinematic data. In a conservative approach, we first correlated the observed ERP *effects* with the observed kinematic *effects* across the conditions in the ERP analysis and the kinematic measures in the behavioral analysis that showed an effect. Across both time-windows in which an ERP effect was observed (190–348 ms and 220–334 ms after stimulus onset), we calculated for each of the three conditions per participant the elicited amplitude in microvolt averaged across the 59 electrodes and across the respective time-window. This resulted in one average ERP value per participant per condition per time-window. As the ERP effect in the comparison of the Declarative and the Imperative condition was observed in the 190–348 ms time-window, based on this information we then calculated for every participant the ERP difference between these two conditions (Declarative – Imperative) and tested for correlations (Pearson's *r*) with the behavioral difference in Gesture Initiation Time and Hold Duration per participant between these two conditions, as these were the two dependent variables that showed the kinematic effects. Similarly, the ERP effect in the comparison of the Informative and Imperative conditions in the 220–334 ms time-window was correlated with these two kinematic difference scores across participants. The four resulting correlation coefficients were not statistically significant (see Table A2).

Table A2
Pearson coefficients (r values) for the correlations between the ERP effects and the kinematic effects. Values between parentheses represent *p* values. Note that no corrections for multiple correlations were performed and no significant correlations were observed.

	Declarative – Imperative: Gesture Initiation Time	Declarative – Imperative: Hold Duration
Declarative – Imperative: ERP 190–348 ms	0.171 (0.325)	0.022 (0.890)
	Informative – Imperative Gesture Initiation Time	Informative – Imperative Hold Duration
Informative – Imperative: ERP 220–334 ms	0.114 (0.514)	0.107 (0.539)

In a more liberal, exploratory approach, we then correlated for each condition separately the average ERP amplitude in microvolt per participant per time-window with the corresponding average per participant for each of the four kinematic measures. As such, this approach tests whether the electrophysiological activity preceding the execution of the pointing gesture in a given condition correlates with any of the kinematic measures of the gesture in that same condition. As can be seen in Table A3 below, no statistically significant correlations were observed in this analysis either.

In sum, we refrain from drawing any strong conclusions on whether the observed kinematic differences are directly driven by the observed earlier electrophysiological activity. Future work may be capable of reliably analyzing electrophysiological activity *during* the execution of the gesture and observe stronger correlations between brain and behavior in doing so.

Table A3
Pearson coefficients (r values) for the correlations between average ERP amplitude and behavioral averages for the four kinematic dependent variables across the three conditions. Values between parentheses represent *p* values. Note that no corrections for multiple correlations were performed and no significant correlations were observed.

	Declarative ERP: 190–348 ms	Informative ERP: 190–348 ms	Imperative ERP: 190–348 ms
Declarative			
Gesture Initiation Time	-0.184 (0.289)		
Stroke Duration	-0.089 (0.609)		
Stroke Velocity	0.003 (0.987)		
Hold Duration	0.250 (0.148)		
Informative			
Gesture Initiation Times		-0.001 (0.996)	
Stroke Duration		0.012 (0.947)	
Stroke Velocity		-0.101 (0.566)	
Hold Duration		0.202 (0.245)	
Imperative			
Gesture Initiation Time			-0.060 (0.732)
Stroke Duration			0.097 (0.581)
Stroke Velocity			-0.196 (0.260)
Hold Duration			0.161 (0.356)

	Declarative ERP: 220–334 ms	Informative ERP: 220–334 ms	Imperative ERP: 220–334 ms
Declarative			
Gesture Initiation Time	-0.175 (0.316)		
Stroke Duration	-0.128 (0.464)		
Stroke Velocity	0.021 (0.904)		
Hold Duration	0.232 (0.181)		
Informative			
Gesture Initiation Time		-0.012 (0.947)	
Stroke Duration		-0.042 (0.813)	
Stroke Velocity		-0.073 (0.678)	
Hold Duration		0.194 (0.264)	
Imperative			
Gesture Initiation Time			-0.031 (0.859)
Stroke Duration			0.036 (0.834)
Stroke Velocity			-0.153 (0.379)
Hold Duration			0.150 (0.391)

References

Ansuini, C., Cavallo, A., Koul, A., Jacono, M., Yang, Y., & Becchio, C. (2015). Predicting object size from hand kinematics: A temporal perspective. *PLoS One, 10*(3). https://doi.org/10.1371/journal.pone.0120432. e0120432.

Bakeman, R., & Adamson, L. B. (1984). Coordinating attention to people and objects in mother-infant and peer-infant interaction. *Child Development, 55*(4), 1278–1289. https://doi.org/10.2307/1129997

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

Bara, B. G. (2010). *Cognitive pragmatics: The mental processes of communication*. MIT Press.

Bates, D. M., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bates, E., Camaioni, L., & Volterra, V. (1975). The acquisition of performatives prior to speech. *Merrill-Palmer Quarterly of Behavior and Development, 21*(3), 205–226.

Becchio, C., Manera, V., Sartori, L., Cavallo, A., & Castiello, U. (2012). Grasping intentions: From thought experiments to empirical evidence. *Frontiers in Human Neuroscience, 6*. https://doi.org/10.3389/fnhum.2012.00117

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

Bosker, H. R., & Peeters, D. (2021). Beat gestures influence which speech sounds you hear. *Proceedings of the Royal Society B: Biological Sciences, 288*(1943), 20202419. https://doi.org/10.1098/rspb.2020.2419

Brunetti, M., Zappasodi, F., Marzetti, L., Perrucci, M. G., Cirillo, S., Romani, G. L., … Aureli, T. (2014). Do you know what I mean? Brain oscillations and the understanding of communicative intentions. *Frontiers in Human Neuroscience, 8*. https://doi.org/10.3389/fnhum.2014.00036

Butterworth, G. (2003). Pointing is the royal road to language for babies. In S. Kita (Ed.), *Pointing: Where language, culture, and cognition meet* (pp. 17–42). Psychology Press.

Carpenter, M., Nagell, K., & Tomasello, M. (1998). Social cognition, joint attention, and communicative competence from 9 to 15 months of age. *Monographs of the Society for Research in Child Development, 63*(4), i–174. https://doi.org/10.2307/1166214

Cavallo, A., Koul, A., Ansuini, C., Capozzi, F., & Becchio, C. (2016). Decoding intentions from movement kinematics. *Scientific Reports, 6*(1), 37036. https://doi.org/10.1038/srep37036

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

Clark, H. H. (1996). *Using language*. Cambridge University Press.

Clark, H. H., Schreuder, R., & Buttrick, S. (1983). Common ground at the understanding of demonstrative reference. *Journal of Verbal Learning and Verbal Behavior, 22*(2), 245–258. https://doi.org/10.1016/S0022-5371(83)90189-5

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). https://doi.org/10.1371/journal.pone.0017719

Cochet, H., & Vauclair, J. (2010). Features of spontaneous pointing gestures in toddlers. *Gesture, 10*(1), 86–107. https://doi.org/10.1075/gest.10.1.05coc

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. (2017). Foreground gesture, background gesture. *Gesture, 16*(2), 176–202. https://doi.org/10.1075/gest.16.2.02coo

Cooperrider, K. (2020). *Fifteen ways of looking at a pointing gesture* [preprint]. *PsyArXiv*. https://doi.org/10.31234/osf.io/2vxft

Cooperrider, K., Fenlon, J., Keane, J., Brentari, D., & Goldin-Meadow, S. (2021). How pointing is integrated into language: Evidence from speakers and signers. *Frontiers in Communication, 6*, 567774. https://doi.org/10.3389/fcomm.2021.567774

Coventry, K. R., Griffiths, D., & Hamilton, C. J. (2014). Spatial demonstratives and perceptual space: Describing and remembering object location. *Cognitive Psychology, 69*, 46–70. https://doi.org/10.1016/j.cogpsych.2013.12.001

Eichert, N., Peeters, D., & Hagoort, P. (2018). Language-driven anticipatory eye movements in virtual reality. *Behavior Research Methods, 50*(3), 1102–1115. https://doi.org/10.3758/s13428-017-0929-z

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

Enrici, I., Adenzato, M., Cappa, S., Bara, B. G., & Tettamanti, M. (2011). Intention processing in communication: A common brain network for language and gestures. *Journal of Cognitive Neuroscience, 23*(9), 2415–2431. https://doi.org/10.1162/jocn.2010.21594

Flack, Z. M., Naylor, M., & Leavens, D. A. (2018). Pointing to visible and invisible targets. *Journal of Nonverbal Behavior, 42*(2), 221–236. https://doi.org/10.1007/s10919-017-0270-3

Goldin-Meadow, S. (2007). Pointing sets the stage for learning language—And creating language. *Child Development, 78*(3), 741–745. https://doi.org/10.1111/j.1467-8624.2007.01029.x

Grice, H. P. (1975). Logic and conversation. *Speech Acts*, 41–58. https://doi.org/10.1163/9789004368811_003

Heider, F., & Simmel, M. (1944). An experimental study of apparent behavior. *The American Journal of Psychology, 57*(2), 243–259.

Henderson, L. M., Yoder, P. J., Yale, M. E., & McDuffie, A. (2002). Getting the point: Electrophysiological correlates of protodeclarative pointing. *International Journal of Developmental Neuroscience, 20*(3), 449–458. https://doi.org/10.1016/S0736-5748(02)00038-2

Herbort, O., & Kunde, W. (2016). Spatial (mis-)interpretation of pointing gestures to distal referents. *Journal of Experimental Psychology: Human Perception and Performance, 42*(1), 78–89. https://doi.org/10.1037/xhp0000126

Heyselaar, E., Hagoort, P., & Segaert, K. (2017). In dialogue with an avatar, language behavior is identical to dialogue with a human partner. *Behavior Research Methods, 49*(1), 46–60. https://doi.org/10.3758/s13428-015-0688-7

Holler, J., & Levinson, S. C. (2019). Multimodal language processing in human communication. *Trends in Cognitive Sciences, 23*(8), 639–652. https://doi.org/10.1016/j.tics.2019.05.006

Hömke, P., Holler, J., & Levinson, S. C. (2018). Eye blinks are perceived as communicative signals in human face-to-face interaction. *PLoS One, 13*(12). https://doi.org/10.1371/journal.pone.0208030. e0208030.

Huizeling, E., Peeters, D., & Hagoort, P. (2022). Prediction of upcoming speech under fluent and disfluent conditions: Eye tracking evidence from immersive virtual reality. *Language, Cognition and Neuroscience, 37*(4), 481–508. https://doi.org/10.1080/23273798.2021.1994621

Kelly, S., Healey, M., Özyürek, A., & Holler, J. (2015). The processing of speech, gesture, and action during language comprehension. *Psychonomic Bulletin & Review, 22*(2), 517–523. https://doi.org/10.3758/s13423-014-0681-7

Kelly, S. D., Barr, D. J., Church, R. B., & Lynch, K. (1999). Offering a hand to pragmatic understanding: The role of speech and gesture in comprehension and memory. *Journal of Memory and Language, 40*(4), 577–592. https://doi.org/10.1006/jmla.1999.2634

Kendon, A. (2004). *Gesture: Visible action as utterance*. Cambridge University Press.

Kendon, A., & Versante, L. (2003). Pointing by hand in "Neapolitan". In S. Kita (Ed.), *Pointing: Where language, culture, and cognition meet* (pp. 109–138). Psychology Press.

Kita, S. (2003). *Pointing: Where language, culture, and cognition meet*. Psychology Press.

Koul, A., Soriano, M., Tversky, B., Becchio, C., & Cavallo, A. (2019). The kinematics that you do not expect: Integrating prior information and kinematics to understand intentions. *Cognition, 182*, 213–219. https://doi.org/10.1016/j.cognition.2018.10.006

Krause, M. A., Udell, M. A. R., Leavens, D. A., & Skopos, L. (2018). Animal pointing: Changing trends and findings from 30 years of research. *Journal of Comparative Psychology, 132*(3), 326–345. https://doi.org/10.1037/com0000125

Krishnan-Barman, S., Forbes, P. A. G., & de Hamilton, A. F. 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

Küntay, A. C., & Özyürek, A. (2006). Learning to use demonstratives in conversation: What do language specific strategies in Turkish reveal? *Journal of Child Language, 33*(2), 303–320. https://doi.org/10.1017/S0305000906007380

Leavens, D. A., Hopkins, W. D., & Bard, K. A. (2005). Understanding the point of chimpanzee pointing: Epigenesis and ecological validity. *Current Directions in Psychological Science, 14*(4), 185–189. https://doi.org/10.1111/j.0963-7214.2005.00361.x

Legault, J., Zhao, J., Chi, Y.-A., Chen, W., Klippel, A., & Li, P. (2019). Immersive virtual reality as an effective tool for second language vocabulary learning. *Languages, 4*(1), 13. https://doi.org/10.3390/languages4010013

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. MIT Press.

Levelt, W. J. M., Richardson, G., & La Heij, W. (1985). Pointing and voicing in deictic expressions. *Journal of Memory and Language, 24*(2), 133–164. https://doi.org/10.1016/0749-596X(85)90021-X

Liszkowski, U., Carpenter, M., Henning, A., Striano, T., & Tomasello, M. (2004). Twelve-month-olds point to share attention and interest. *Developmental Science, 7*(3), 297–307. https://doi.org/10.1111/j.1467-7687.2004.00349.x

Liu, R., Bögels, S., Bird, G., Medendorp, W. P., & Toni, I. (2019). *Hierarchical integration of communicative and visuospatial perspective-taking demands in sensorimotor control of referential pointing* [preprint]. *PsyArXiv*. https://doi.org/10.31234/osf.io/htvqa

Manera, V., Becchio, C., Cavallo, A., Sartori, L., & Castiello, U. (2011). Cooperation or competition? Discriminating between social intentions by observing prehensile movements. *Experimental Brain Research, 211*(3), 547–556. https://doi.org/10.1007/s00221-011-2649-4

Maris, E. (2012). Statistical testing in electrophysiological studies. *Psychophysiology, 49*(4), 549–565. https://doi.org/10.1111/j.1469-8986.2011.01320.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

McEllin, L., Sebanz, N., & Knoblich, G. (2018). Identifying others' informative intentions from movement kinematics. *Cognition, 180*, 246–258. https://doi.org/10.1016/j.cognition.2018.08.001

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

Murillo Oosterwijk, A., de Boer, M., 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

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia, 9*(1), 97–113. https://doi.org/10.1016/0028-3932(71)90067-4

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience, 2011, 1-9*. https://doi.org/10.1155/2011/156869. e156869.

Pan, X., & Hamilton, A. F. C. (2018). Why and how to use virtual reality to study human social interaction: The challenges of exploring a new research landscape. *British Journal of Psychology, 109*(3), 395–417. https://doi.org/10.1111/bjop.12290

Parsons, T. D. (2015). Virtual reality for enhanced ecological validity and experimental control in the clinical, affective and social neurosciences. *Frontiers in Human Neuroscience, 9*. https://doi.org/10.3389/fnhum.2015.00660

Peeters, D. (2019). Virtual reality: A game-changing method for the language sciences. *Psychonomic Bulletin & Review, 26*(3), 894–900. https://doi.org/10.3758/s13423-019-01571-3

Peeters, D. (2020). Bilingual switching between languages and listeners: Insights from immersive virtual reality. *Cognition, 195*, 104107. https://doi.org/10.1016/j.cognition.2019.104107

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

Peeters, D., Chu, M., Holler, J., Özyürek, A., & Hagoort, P. (2013). Getting to the point: The influence of communicative intent on the kinematics of pointing gestures. In *Proceedings of the 35th annual meeting of the cognitive science society* (pp. 1127–1132).

Peeters, D., & Dijkstra, T. (2018). Sustained inhibition of the native language in bilingual language production: A virtual reality approach. *Bilingualism: Language and Cognition, 21*(5), 1035–1061. https://doi.org/10.1017/S1366728917000396

Peeters, D., Krahmer, E., & Maes, A. (2021). A conceptual framework for the study of demonstrative reference. *Psychonomic Bulletin & Review, 28*(2), 409–433. https://doi.org/10.3758/s13423-020-01822-8

Perniss, P. (2018). Why we should study multimodal language. *Frontiers in Psychology, 9*. https://doi.org/10.3389/fpsyg.2018.01109

R Core Team. (2018). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Rocca, R., Tylén, K., & Wallentin, M. (2019). This shoe, that tiger: Semantic properties reflecting manual affordances of the referent modulate demonstrative use. *PLoS One, 14*(1). https://doi.org/10.1371/journal.pone.0210333. e0210333.

Rochester, S. R. (1973). The significance of pauses in spontaneous speech. *Journal of Psycholinguistic Research, 2*, 51–81.

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

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

Sartori, L., Becchio, C., & Castiello, U. (2011). Cues to intention: The role of movement information. *Cognition, 119*(2), 242–252. https://doi.org/10.1016/j.cognition.2011.01.014

Searle, J. R. (1969). *Speech acts: An essay in the philosophy of language*. Cambridge University Press.

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

Tomasello, M., & Call, J. (1997). *Primate cognition*. Oxford University 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

Tromp, J., Peeters, D., Meyer, A. S., & Hagoort, P. (2018). The combined use of virtual reality and EEG to study language processing in naturalistic environments. *Behavior Research Methods, 50*(2), 862–869. https://doi.org/10.3758/s13428-017-0911-9

Trujillo, J. P., & Holler, J. (2021). The kinematics of social action: Visual signals provide cues for what interlocutors do in conversation. *Brain Sciences, 11*(8), 996. https://doi.org/10.3390/brainsci11080996

Trujillo, J. P., Simanova, I., Bekkering, H., & Özyürek, A. (2018). Communicative intent modulates production and comprehension of actions and gestures: A Kinect study. *Cognition, 180*, 38–51. https://doi.org/10.1016/j.cognition.2018.04.003

Trujillo, J. P., Vaitonyte, J., Simanova, I., & Özyürek, A. (2019). Toward the markerless and automatic analysis of kinematic features: A toolkit for gesture and movement research. *Behavior Research Methods, 51*(2), 769–777. https://doi.org/10.3758/s13428-018-1086-8

Vesper, C., & Richardson, M. J. (2014). Strategic communication and behavioral coupling in asymmetric joint action. *Experimental Brain Research, 232*(9), 2945–2956. https://doi.org/10.1007/s00221-014-3982-1

Vigliocco, G., Perniss, P., & Vinson, D. (2014). Language as a multimodal phenomenon: Implications for language learning, processing and evolution. *Philosophical*

Transactions of the Royal Society B: Biological Sciences, 369(1651), 20130292. https://doi.org/10.1098/rstb.2013.0292

Willems, R. M., de Boer, M., de Ruiter, J. P., Noordzij, M. L., Hagoort, P., & Toni, I. (2010). A dissociation between linguistic and communicative abilities in the human brain. *Psychological Science, 21*(1), 8–14. https://doi.org/10.1177/0956797609355563

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

Wittgenstein, L. (1955). *Philosophical investigations*. Basil Blackwell.

Yoon, S. O., & Brown-Schmidt, S. (2019). Contextual integration in multiparty audience design. *Cognitive Science, 43*(12). https://doi.org/10.1111/cogs.12807. e12807.

Zappa, A., Bolger, D., Pergandi, J.-M., Mallet, P., Dubarry, A.-S., Mestre, D., & Frenck-Mestre, C. (2019). Motor resonance during linguistic processing as shown by EEG in a naturalistic VR environment. *Brain and Cognition, 134*, 44–57. https://doi.org/10.1016/j.bandc.2019.05.003

Zeileis, A., & Hothorn, T. (2002). *Diagnostic checking in regression relationships. 5*.