Communicative intent modulates production and comprehension of actions and gestures: A Kinect study

James P. Trujillo[a,b,*], Irina Simanova[a], Harold Bekkering[a], Asli Özyürek[b,c]

[a] Donders Institute for Brain, Cognition and Behaviour, Radboud University Nijmegen, The Netherlands
[b] Centre for Language Studies, Radboud University Nijmegen, The Netherlands
[c] Max Planck Institute for Psycholinguistics, Wundtlaan 1, 6525XD Nijmegen, The Netherlands

ABSTRACT

Actions may be used to directly act on the world around us, or as a means of communication. Effective communication requires the addressee to recognize the act as being communicative. Humans are sensitive to ostensive communicative cues, such as direct eye gaze (Csibra & Gergely, 2009). However, there may be additional cues present in the action or gesture itself. Here we investigate features that characterize the initiation of a communicative interaction in both production and comprehension.

We asked 40 participants to perform 31 pairs of object-directed actions and representational gestures in more- or less- communicative contexts. Data were collected using motion capture technology for kinematics and video recording for eye-gaze. With these data, we focused on two issues. First, if and how actions and gestures are systematically modulated when performed in a communicative context. Second, if observers exploit such kinematic information to classify an act as communicative.

Our study showed that during production the communicative context modulates space–time dimensions of kinematics and elicits an increase in addressee-directed eye-gaze. Naïve participants detected communicative intent in actions and gestures preferentially using eye-gaze information, only utilizing kinematic information when eye-gaze was unavailable.

Our study highlights the general communicative modulation of action and gesture kinematics during production but also shows that addressees only exploit this modulation to recognize communicative intention in the absence of eye-gaze. We discuss these findings in terms of distinctive but potentially overlapping functions of addressee directed eye-gaze and kinematic modulations within the wider context of human communication and learning.

1. Introduction

Our hands may be used in a variety of ways to interact with the world around us. Two such interactions are object-directed actions, in which the hands interact with a physical object (e.g., to open a jar), and representational gestures (Kendon, 2004; McNeill, 1994), in which the hands are used to simulate an interaction or visually represent a non-present object (hands move as if opening a jar). What is specific to humans is that both categories of movements can be recruited for the purpose of communication, allowing us to teach through demonstration (Campisi & Özyürek, 2013; Southgate, Chevallier, & Csibra, 2009) or convey the intention for an observer to act in response (Tomasello, 2010).

Characteristic of communicative acts is the accompanying addressee-directed eye-gaze (Brand, Shallcross, Sabatos, & Massie, 2007).

Humans in particular seem inherently sensitive to ostensive communicative cues, such as direct eye gaze and eyebrow raise (Csibra & Gergely, 2009). Direct eye-gaze is particularly powerful, displaying a willingness to interact (Cary, 1978), as well as altering cognitive processing and behavioural response (Senju & Johnson, 2009). For example, a recent study by Innocenti et al. investigated the impact of eye-gaze on a requesting gesture, e.g. reaching out and grasping an empty glass with the implied request to have it filled. The study showed that both the speed and size of a communicative gesture and addressee-directed eye-gaze affected kinematics of the response act. Therefore, the mere presence of direct eye-gaze induced a measurable effect on the response of the addressee (Innocenti, de Stefani, Bernardi, Campione, & Gentilucci, 2012).

For communication in general, there are at least two main requirements: the communicator must make his or her intention to

---

communicate recognizable, and they must represent the semantic information they wish to be received by the observer. The first step in communicating using actions or gestures is thus for the communicator to make the action or gesture recognizable as being a communicative act. In doing so the communicator might use kinematic modulation (see, for example, (Becchio, Cavallo, et al., 2012)) as well as addressee-directed eye-gaze (Kampe, Frith, & Frith, 2003; Schilbach et al., 2006). Secondly the communicator's cues need to be picked up by addressee in order to interpret actions or gestures as communicative. Here, again, both the kinematics of the manual acts and the ostensive cues, or the interaction of both, can play a role. In the present study, we address the overall profile of communicative actions and gestures within the larger context of production and comprehension. We compare for the first time actions and gestures in more-communicative versus less communicative contexts to see if they are subject to similar kinematic modulations and are coupled by ostensive cues. We then investigate whether and how these cues are in turn interpreted by addressees. To quantify kinematic modulation effects we use the Kinect device to obtain a non-intrusive, objective and precise measure of action and gesture. The next few paragraphs summarize the current literature on the kinematic modulation and on the perception of actions and gestures in communicative context.

2. Production of communicative actions and gestures

At the basic motor control level, actions are thought to follow a principle of motor efficiency (Todorov & Jordan, 2002). In this framework, control of an action is a balance between reducing cost and achieving the goal of the action. While this framework explains action control in a neutral setting, there is evidence that other contextual or cognitive domains influence these dynamics. The intention to communicate affects the velocity of reach-to-grasp movements (Sartori, Becchio, Bara, & Castiello, 2009), and can modulate the trajectory of such movements to make a target more predictable to a co-actor (Sacheli, Tidoni, Pavone, Aglioti, & Candidi, 2013). Furthermore, child-directed communicative actions are marked by several kinematic modulations, including an increased range-of-motion and punctuality (Brand, Baldwin, & Ashburn, 2002). At the level of cognitive and neural implementation of motor control, this indicates a top-down influence on action production that is theorized to facilitate interactions by balancing the initial efficiency principle with the additional factor of disambiguating the end-goal for an observer (Pezzulo, Donnarumma, & Dindo, 2013). In line with the account by Pezzulo and colleagues, we suggest that the kinematic modulation from a communicative context can be summarized as an optimization of space-time dimensions (Pezzulo et al., 2013). In this account, communicative modulation is an effort to present the optimal amount of visual information to disambiguate the act (optimization of space) within an efficient amount of time (optimization of time). We extend this framework by investigating specific kinematic cues, and testing how ostensive eye-gaze is implemented together with kinematic modulation in both actions and gestures. As actions are not inherently communicative, and indeed less likely to be interpreted as communicative by observers (Kelly, Healey, Özyürek, & Holler, 2015; Novack, Wakefield, & Goldin-Meadow, 2016), it may be that direct eye-gaze is an important communicative cue for actions. An additional open question is whether similar communication modulations occur not only in actions, but also in representational gestures.

Although the motor efficiency/optimization principle does not specifically refer to gestures, they too are manual acts with a specific extrinsic goal. Often, this goal to change the internal state of an observer, but gestures may also be performed without communicative intention. For instance, in the context of co-thought gestures, one uses gestures while trying to solve complex visuospatial tasks (Chu & Kita, 2011). Additionally, clinicians often use pantomime production tasks as a clinical measure in aphasia (Goldenberg, Hartmann, & Schlott, 2003;

Hermsdörfer, Li, Randerath, Goldenberg, & Johannsen, 2012). Gestures then are likely to also follow an initial efficiency principle which may further be modulated depending on the goal or intention. Like actions, gestures are also influenced by a communicative context. For example, when meant to be more informative to an observer, pointing gestures are made slower than when the gesture will not be used by an observer (Peeters, Chu, Holler, Hagoort, & Özyürek, 2015). Furthermore, during a demonstration or explanation, a gap in common knowledge between speaker and addressee leads to gestures that are larger (Bavelas, Gerwing, Sutton, & Prevost, 2008; Campisi & Özyürek, 2013), more complex or precise (Galati & Brennan, 2014; Gerwing & Bavelas, 2004; Holler & Beattie, 2005) and are produced higher in space (Hilliard & Cook, 2016). Whether these kinematic modulations are comparable to those observed in actions in similar communicative settings, has not been assessed. This is of interest because gestures are reliant on kinematics to convey meaning, whereas actions can utilize the object (manipulation) to convey meaning. We could then expect the two modalities to differ in the way they are made more communicative. For example, because gestures are more inherently communicative, the strong direct eye-gaze signal may be less important for gestures compared to actions. Therefore, an interesting open question is whether the same kinematic and eye-gaze features are modulated when the two modalities are performed in a more communicative context.

3. Perception of communicative actions and gestures

Although communicative intent driven modulation is present during the production of actions and gestures, as shown above, it is less clear whether and how this modulation is seen or used by observers. Studies show that children prefer actions marked by increased range of motion and exaggerated movement boundaries (Brand et al., 2002), which leads to increased visual attention in infants (Brand & Shallcross, 2008), and more frequent imitation of a demonstrated action in children (Williamson & Brand, 2014). In regard to intention recognition, a study on social actions by Manera et al., showed that observers are able to distinguish between cooperative and competitive actions using only the kinematics (point-light-displays) (Manera, Becchio, Cavallo, Sartori, & Castiello, 2011). This suggests that kinematic modulation, at least in regard to child-directed actions and social context, is noticed by observers.

With regard to perception of the communicativeness of gestures, a recent study by Novack et al. shows that movements in the presence of objects are seen as representations of actions, while the *same* movements made in the absence of objects are described as being movement for its own sake (Novack et al., 2016). This suggests that even though kinematics clearly affects the way the action or gesture is perceived, observers rely strongly on situational constraints to understand the underlying intention. Further evidence comes from a study on body orientation and iconic gesture use (Nagels, Kircher, Steines, & Straube, 2015). Nagels and colleagues found that when a speaker is oriented toward an addressee and gestures during speech, the addressee feels more addressed, thereby indicating a better recognition of communicative intent. Interestingly, both the condition with the speaker orientated towards the addressee but not using iconic gestures as well as the condition with the speaker oriented away from the addressee but using iconic gesture were also rated as being more communicative than the condition in which the speaker faced away and did not use gestures (Nagels et al., 2015). These studies indicate that, at least for iconic gestures, both eye-gaze directed to the addressee and gestures can convey a communicative intent. It is important to note that although iconic gesture use contributed to the feeling of being addressed, the kinematics of gestures themselves were not modified in that study. To date, there are no studies that have investigated kinematic modulation of gestures in comprehension. Therefore the question remains of how such a modulation will impact the perceived communicativeness of the gesture or the action. Furthermore, previous research comparing the

integration of speech with actions and gestures suggests that, at the level of multi-modal integration, observers are more responsive to the communicativeness of gestures compared to actions, (Kelly et al., 2015). This is in line with previous claims that gestures, due to their nature as being representational and generalizable, have a special role in learning and communication (Goldin-Meadow, 2017). Therefore an open and interesting question is whether communicative cues are recognized, when indexing communicative intent, similarly in gestures compared to actions.

4. Current study

The current study seeks to link previous findings on communicative manual acts by investigating the characteristic features that facilitate the initiation of a communicative interaction, taking into account both production and comprehension. Specifically, we ask if communicative intent modulates the kinematics of, and eye-gaze behavior accompanying both actions and gestures, and if observers use kinematic modulation and/or eye gaze to recognize the communicative intention of the action and gesture. Previous studies have shown that communicative intent may modulate different aspects of actions and/or gestures. However, these two modalities have not been investigated in a single design, utilizing the same communicative context and considering both production and comprehension. To address these questions, we used three experiments: one for production and two for comprehension.

In the first experiment, two groups of participants performed a set of everyday actions, as well as the corresponding representational gestures. One group of participants performed in a more communicative context, and the other in a less-communicative, or self-serving context. In order to provide a non-intrusive, naturalistic setting, we did not specifically instruct participants to "be communicative", but used a subtle manipulation of the context in which they performed the task. We used high-definition video recordings for manual coding of eye-gaze behavior. Furthermore, we used the Microsoft Kinect to collect full-body 3D joint tracking data. Use of the Kinect allows tracking of the participants' 3-dimensional movements, allowing streamlined, quantitative coding of kinematic features. We chose this approach as opposed to the more traditionally used optical tracking as the Kinect does not require markers or calibration. This supports the naturalistic aspect of our experiment, while maintaining high quality motion capture performance (Chang et al., 2012; Fernández-Baena, Susín, & Lligadas, 2012). Although relatively new in the field of research, the Kinect has successfully been implemented for gesture (Biswas & Basu, 2011; Paraskevopoulos, Spyrou, & Sgouropoulos, 2016) and sign-language recognition (Pedersoli, Benini, Adami, & Leonardi, 2014) and was shown to be a reliable tool for measuring kinematics. In the second experiment, we showed a selection of single acts to a new set of participants in order to understand how these features are used by an addressee. These participants were asked to classify each act as either communicative or non-communicative. We then assessed which features contributed to an observer's context classification. In the third experiment, the same subset of videos was modified to obscure the eye-gaze information. The clips were then shown to a group of naïve participants, replicating the Experiment II, to further distinguish relative contribution of the kinematic modulation and eye-gaze in the detection of the communicative intend.

In sum, this study aims to elucidate the profile of communicative action and gesture, and place this profile in the larger frame of production and recognition. We ask which kinematic features are modulated by communicative interactions on the production side, and how this modulation facilitates comprehension of the communicative intent.

5. Methods – Experiment I

5.1. Participants

Forty participants were included in this study, recruited from the Radboud University. Participants were selected on the criteria of being aged 18–35, right-handed, healthy and fluent in the Dutch language. Additionally, one confederate also participated in all experiments. The confederate was a 23 year old, female, native Dutch speaker. The experimental procedure was in accordance with a local ethical committee.

5.2. Context settings

Participants were divided into two groups: more communicative (n = 20, 13 females, mean age = 23.6 years) and less communicative (n = 20, 13 females, mean age = 23.8 years). For the more-communicative group, the confederate was introduced as having the task of watching the experiment through the camera placed in front of the participant and learning the participant's actions/gestures. In the less-communicative group, the confederate was introduced as having the task of watching the experiment through the camera and learning the general experimental set-up. Critically, this means that in both groups the confederate was considered to be watching and learning, but only in the communicative group was the confederate stated to be learning directly from the participant's manual acts. The paradigm therefore aimed to create a continuum of behavior, extending from less communicative, self-serving behavior, to highly communicative behavior that was highly oriented towards the addressee. This novel paradigm builds on designs using confederates to control feedback while eliciting an interactive setting (eg. Holler & Wilkin, 2011; Sartori et al., 2009). Crucially, our context manipulation aims to influence the intentional stance of the participant towards the addressee, similar to Peeters et al. (2015), while keeping all other (eg. presence of confederate and instructions to participant) factors equal. Participants were pseudo-randomly assigned to groups, with consideration only being given to a relatively equal distribution of males and females to each group.

5.3. Items

The full set of actions/gestures contained 31 item sets, most of which consisted of two objects. Auditory instructions accompanied each item set and were recorded by a female, native Dutch speaker. Items were presented in random order for each participant and modality (action and gesture). All instructions were similarly constructed in a simplistic way as to indicate the object(s) and a verb (e.g. The participant may be given a pitcher of water and an empty glass, with the accompanying instructions "Giet het water in het glas", *pour the water into the glass*). A full list of the instructions used for these items can be found in Appendix A.

5.4. Modality

Both groups executed the full list of items in each of two conditions, reflecting two modalities of movement: action and gesture. For the action condition, participants were simply instructed to follow the auditory instructions using the items on the table. In the gesture condition, participants were instructed to follow the instructions as if they were using the objects, but without actually touching them. The order of modalities was counterbalanced across subjects. An overview of the design with example frames taken from each factor (modality × context) can be seen in Fig. 1.

5.5. Procedure

For both groups, we used the following procedure: the participant entered the experiment room and was briefly introduced to a
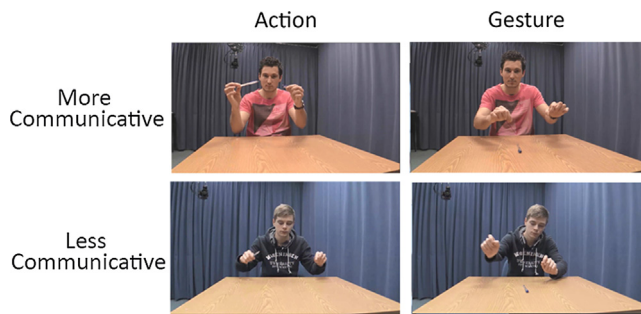
Fig. 1. Overview of the experimental design. Each image depicts an example frame taken from a video of the corresponding factor. In each image, the action or pantomime being performed is 'remove the cap from the pen'. The x-axis displays modality (action vs. gesture as a within factor) and the y-axis depicts context (more communicative vs. less-communicative as a between factor).

confederate, as described above. After the brief introduction, the confederate moved to an adjoining room. The participant was then seated at a table with a camera hanging directly in front of the table, facing the participant at approximately eye-level. The participant

was shown two areas marked on the table to designate the starting point for his/her hands and instructed on the experimental procedure. After asking both the participant and the confederate if they are ready to begin, the door separating the participant from the confederate was closed. Each item began with (an) object(s) being placed in front of the participant in the middle of the table. After the experimenter was out of sight from the participant, and both hands were resting on the designated starting points, auditory instructions were played indicating what action/gesture should be executed. After the instructions were played, a short interval followed before a bell sound was played, indicating the participant may begin executing the action/gesture. Participants were told that they must not begin acting until they hear the bell sound, at which point the camera would begin recording. When the action/gesture was completed, the participant returned his/her hands to the indicated starting places. At the end of the first block (modality), the experimenter explained the instructions for the second block and again asked for verbal confirmation from the confederate if their task was still going well. After this, the door was again closed and the second condition began. During both conditions, after the 10th and 20th item, the experimenter also briefly asked the confederate and the participant if their respective tasks were going well. This was done in order to enforce the idea that another participant was present throughout the experiment. At the end of the second block, the participant was debriefed regarding the purpose of the experiment and the presence of the confederate.

5.6. Data collection

In order to optimize and streamline analysis of kinematic features, we employed the Microsoft Kinect V2 to collect 3D joint tracking data. The Kinect utilizes single-camera motion tracking and allows automatic, markerless tracking of 25 joints on the human body. For the purpose of this study, we collected data from all 25 joints, although the hips and legs were not used for any analysis. For a graphic overview of the joints utilized in this study, see Fig. 2A. Although relatively new in the field of research, studies have shown that the Kinect offers hand and arm tracking performance with accuracy comparable to that of high performance optical motion tracking systems such as the OptiTrack (Chang et al., 2012). Data was collected at 30 frames per second (fps). Film data was collected at 25 fps by a camera hanging at approximately eye-level, directly in front of the participant.

Due to technical problems, Kinect data was not collected for seven recording acquisitions: for one less-communicative and one more-communicative participant no Kinect data was acquired, and for two

less-communicative and one more-communicative participant no Kinect data for the Action modality was acquired.

5.7. Data processing

All kinematic analyses were carried out in MATLAB 2015a (The MathWorks, Inc., Natick, Massachusetts, United States) using in-house developed scripts. To account for the noise inherent in Kinect recordings, we first applied a Savitsky-Golay filter with a span of 15 and degree of 5.

The following kinematic features were calculated individually for each item: *Distance* was calculated as the total distance travelled by both hands in 3D space over the course of the item. *Peak velocity* was calculated as the greatest velocity achieved with the right (dominant) hand. *Maximum amplitude* refers to the maximum vertical height, as indexed by six categories (see supplementary Fig. 1 for a visual representation of these categories), achieved by either hand in relation to the body. *Hold time* was calculated as the total time, in seconds, counting as a hold. Holds were defined as an event in which both hands and arms are still for at least 0.3 s. *Submovements* were calculated as the number of individual ballistic movements made, per hand, throughout the item. Our approach was based on the description given by Meyer and colleagues (Meyer, Abrams, Kornblum, Wright, & Smith, 1988). For a more detailed description of how the individual features were calculated, see Appendix B.

In order to allow comparisons between items with relatively different kinematic profiles, we first standardized all kinematic features. Each feature was transformed into a z-score, per item, by subtracting the mean (n = 40) for that item-feature and dividing by the same item-feature's standard deviation. This allowed us to keep any variability between subjects, while removing between-item variability.

We additionally calculated the overall duration of each item. The duration of the item was calculated as the total time between the beginning and end of the item. The beginning of the item was marked by the bell sound, which indicated the beginning of the trial for the participant, which occurred approximately 500 ms before the participant began to move his or her hands from the starting points; the end of the item was defined as approximately 500 ms after the participants' hands returned to the starting points, when the second bell sound was played. The 500 ms windows before and after hand movements were approximate in nature due to the fact that they are linked to the bell sound that was manually played by the experimenter. Participants tended to respond approximately 500 ms after hearing the sound, but if the participant waited more than 1000 ms or less than 250 ms, this window was given a duration of 500 ms. The bell was likewise played approximately 500 ms after both hands were resting on the table, but the duration was set to the bell sound (which could vary due to a variable response by the experimenter) in order to only capture the time-frame within which participants believed they were visible to the confederate. We transformed the durations into z-scores, per item, using the same method as described for the kinematic features.

Eye gaze was manually coded on a frame-by-frame basis using the video annotation software ELAN (www.lat-mpi.eu/tools/elan/). Eye-gaze was coded by taking the amount of time between the beginning and end (as calculated for our duration measure) in which the participant looked directly at the camera, in milliseconds, and divided by the total duration of the item. This provided a general measure of the proportional gaze time, indicating the percent of the overall item duration in which eye-contact was made with the camera. Including the 500 ms included before initial hand movement and after final hand movement was done in order to incorporate gaze cues immediately preceding or following an action, during the time in which participants thought they were being observed or recorded.
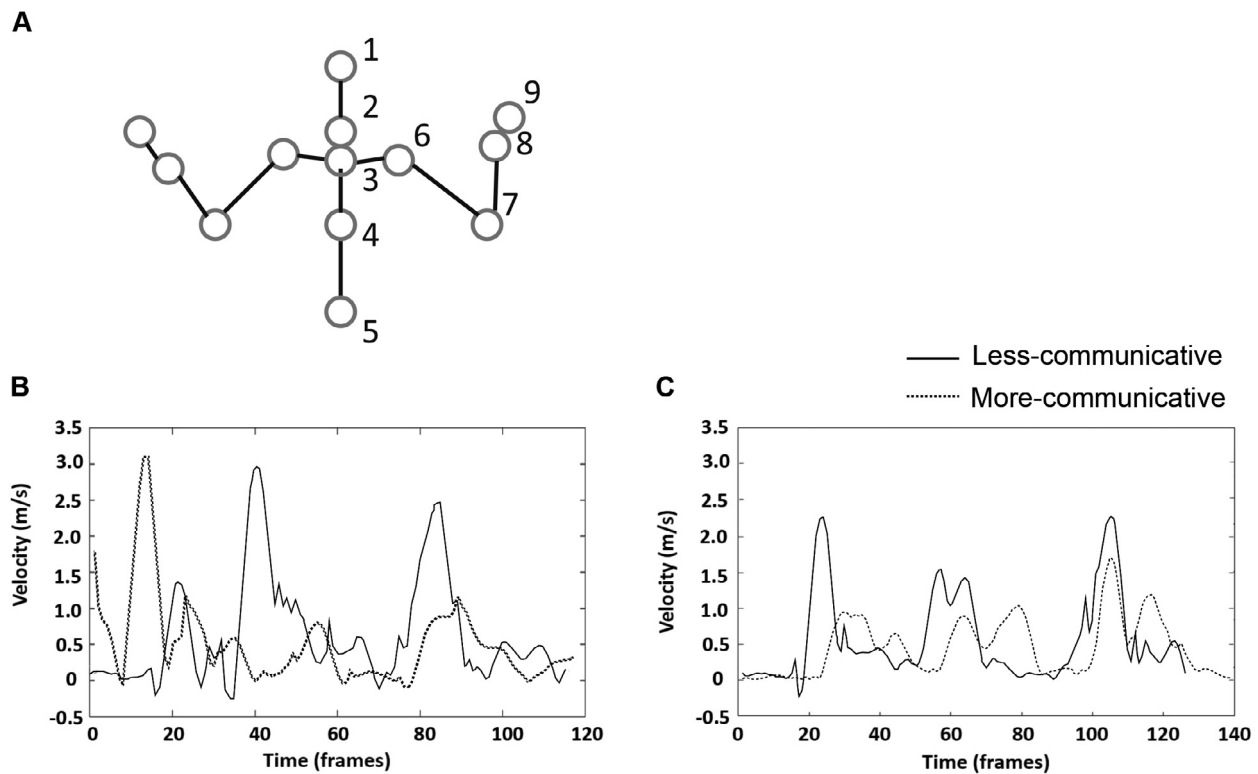
Fig. 2. Illustration of the tracked skeleton (A) and comparison of velocity profiles (B, C). Panel A illustrates the joints tracked by the Kinect for analysis of kinematics. The circles represent each individual joint: 1. Top of head 2. Neck 3. Spine – upper 4. Spine – middle 5. Spine – lower 6. Shoulder 7. Elbow 8. Wrist 9. Hand. Panels B and C depict two representative velocity profiles (measured from the right hand), taken from the same item ("Place the apple in the bowl"), shown overlaid for comparison. Panel B depicts items from the Action modality, while panel C depicts items from the Gesture modality. The green line corresponds to a more-communicative act, while the blue line corresponds to a less-communicative act. The x-axis represents time, given in frames. The y-axis represents velocity, given in meters per second (m/s).

5.8. Data analysis

In order to determine whether the two contexts could be differentiated on the basis of kinematic features, we utilized a linear mixed model. This was done in order to incorporate all of the data variance into our analyses. This analysis was perfomed using R (R Core Team, 2013) and lme4 (Bates, Mächler, Bolker, & Walker, 2014). We created six linear mixed-effects models, each with one of the features of interest (distance, maximum amplitude, submovements, hold-time, peak velocity, gaze, duration) as the dependent variable, with context as a fixed-effect, and a random intercept for the item factor. To test the significance of these models, we used chi-square tests to compare the models of interest with a null model, thereby comparing whether the variable of interest, context, explains significantly more of the variance than the random-intercept-only model. In order to account for potential correlations between kinematic features and eye-gaze, as well as the increased type-I error rate associated with multiple comparisons, we used Simple Interactive Statistical Analysis (http://www.quantitativeskills.com/sisa/calculations/bonfer.htm) to calculate an adjusted Bonferroni correction using the mean correlation between the six tested features (action r = 0.12; gesture r = 0.16), which led to a Bonferroni adjusted alpha value of to $p < 0.011$ for gestures and $p < 0.010$ for actions.

No statistical comparisons were performed between actions and gestures. This is due to the z-transformation of the kinematic values, which normalizes the data between items, but results in similar distributions for actions and gestures. Any difference in the mean of these two distributions is therefore due to an uneven distribution of data around the mean, rather than a difference of the mean itself.

6. Results – Experiment I

In the action modality, the communicative context was associated with an increased proportion of addressee-directed eye-gaze of $4\% \pm 0.53\%$ of the total video duration ($\chi^2(1) = 54.61$, $p < 0.001$), as well as an increase of 0.21 ± 0.04 SDs in *distance* ($\chi^2(1) = 26.94$, $p < 0.001$), an increase of 0.18 ± 0.06 SDs in *submovements* ($\chi^2(1) = 10.10$, $p = 0.001$) and an increase of 0.16 ± 0.06 SDs in *maximum amplitude* ($\chi^2(1) = 7.21$, $p = 0.007$) and near-significant increase of 0.11 ± 0.01 SDs in *peak velocity* ($\chi^2(1) = 5.99$, $p = 0.014$). *Holdtime* was not significantly different between the two contexts ($\chi^2(1) = 0.16$, $p = 0.691$). More communicative actions were found to be 0.15 ± 0.05 SDs longer in overall duration when compared to less-communications actions ($\chi^2(1) = 7.73$, $p = 0.005$).

In the gesture modality, the communicative context was estimated to increase the proportion of addressee-directed eye-gaze by $7\% \pm 0.82\%$ of the total video duration ($\chi^2(1) = 61.01$, $p < 0.001$), as well as *distance* by 0.24 ± 0.05 SDs ($\chi^2(1) = 19.57$, $p < 0.001$), *peak velocity* by 0.31 ± 0.06 *SDs* ($\chi^2(1) = 30.97$, $p < 0.001$), *submovements by* 0.28 ± 0.06 SDs ($\chi^2(1) = 23.36$, $p < 0.001$) and *maximum amplitude by* 0.36 ± 0.06 SDs ($\chi^2(1) = 37.43$, $p < 0.001$). *Holdtime was increased* by 0.12 ± 0.06 SDs, which was not significant with the adjusted alpha threshold ($\chi^2(1) = 4.42$, $p = 0.011$). More-communicative gestures were found to be 0.22 ± 0.06 SDs longer in duration when compared to less-communicative gestures ($\chi^2(1) = 15.15$, $p < 0.001$).

An illustrative example of the kinematic profile from sample cases of actions and gestures can be seen in Fig. 2, and overview of the eye-gaze and kinematic results can be seen in Fig. 3.
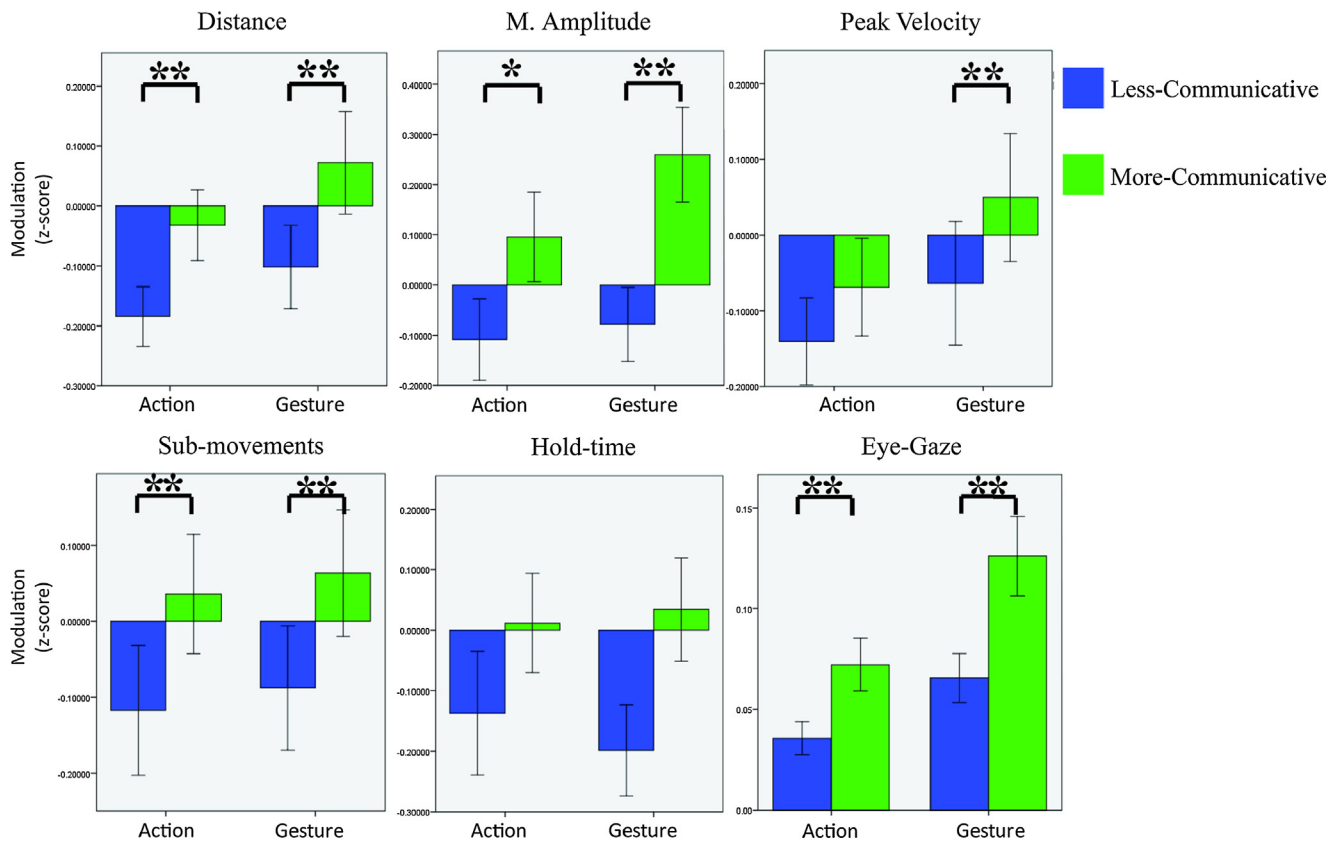
Fig. 3. Comparison of more-communicative and less-communicative kinematic features and eye-gaze. Features are displayed in separate plots. Action and gesture are separated on the x-axis. For kinematic features, the y-axis displays the standardized value (positive values therefore indicate higher-than-average features, while negative values indicate lower-than-average features); for eye-gaze, the y-axis represents proportional addressee-directed eye-gaze. Blue bars depict less-communicative average values, while green bars represent more-communicative average values. M. Amplitude = Maximum Amplitude. $^*p < 0.05$; $^{**}p < 0.001$.

7. Conclusion and discussion – Experiment I

The aim of our first experiment was to quantify the kinematics and eye-gaze behavior of actions and gestures produced in more or less communicative setting. We found that both modalities were modulated in regards to the overall size, number of submovements, and maximum amplitude, with gestures also showing an increase in peak velocity in the communicative context. Furthermore, both modalities elicited more addressee-directed eye-gaze in the communicative context. We also showed this to be the case for a variety of items.

At a motor control level, actions are performed in a manner that optimally balances the successful completion of the action with energy cost, fine control of the movement (Todorov & Jordan, 2002), and environmental constraints (Gergely & Csibra, 2003). Although this explains action control in a neutral setting, previous studies have shown an effect of social context on action kinematics (Becchio, Sartori, Bulgheroni, & Castiello, 2008; Sartori et al., 2009). In these studies, the velocity of movements is differentially modulated dependent on whether or not the actor is attempting to communicate, or whether the action is being performed in a competitive or a cooperative setting. Our findings confirm and expand upon these studies by showing that multiple aspects of movement kinematics are modulated by a communicative context across a wide selection of manual acts. The results indicate a top-down, or context-driven modulation of the motor control system (Friston, 2011). We additionally show that a similar pattern of kinematic modulation is seen both for object-directed actions as well as for the corresponding representational gesture.

Although highly similar, gestures differed from actions in that gestures also had a faster peak-velocity and a subtle increase in hold-time. These features may be more subtle, or they may result from the

additional presence of objects during action production, which provides an extra constraint. These two features fit well with the idea of communicative acts being produced with more punctuation, with the difference between modalities suggesting that this may not always be possible when acting with an object. Furthermore, we show that both actions and gestures are accompanied by more addressee-directed eye-gaze. This compliments previous studies examining communicative modulation of kinematics, showing that actions are also accompanied by more addressee-directed gaze, and also shows that although gestures may be more inherently communicative, they too show an up-regulation of this gaze behavior in the more-communicative context.

We suggested that the communicative context enhances communication efficiency by optimizing space-time dimensions. We found that more-communicative acts covered more visual space and involved more submovements than less-communicative acts, although this was at the cost of requiring more time to produce. The increase in size may optimize the overall amount of information available (ie. Providing more visual sampling of that movement within the same time-frame), while the increase in submovements may indicate a more detailed representation within the presented information. The fact that these increases are produced at the cost of affecting the overall duration provides support for computational accounts of modulations occurring as an optimization of space-time dimensions (Pezzulo et al., 2013). In other words, the amount of utilized visual space increases, but this is balanced against how much time the overall act requires to produce. This is in line with the rather minimal difference in standardized durations (more communicative actions were 0.15 standard deviations larger than less-communicative actions, while more communicative gestures were 0.22 standard deviations larger than less-communicative gestures). Our finding of a heightened peak-velocity in the gesture

modality is also mirrored in a study by Vesper and Richardson, where a cooperative context elicits increased size and peak-velocity during a joint-tapping task (Vesper & Richardson, 2014). This finding can also be interpreted as an optimization of space-time parameters, with the larger movement providing more information and the faster peak-velocity reducing the overall time to produce the act. Although we do not specifically investigate differences between individual manual acts, our study provides experimental evidence that this kinematic optimization may be a signature of more communicative acts in general, regardless of what the specific act is.

Communicative acts are inherently designed for a second person with whom the actor wishes to interact. Although movement kinematics are modulated by the communicative context, it must still be determined what the effect of this modulation is on the observer. For example, although end-goal intentions also modulate the initial phases of an action, a study by Naish and colleagues showed that this information cannot be read by an observer (Naish, Reader, Houston-Price, Bremner, & Holmes, 2013). The role kinematic modulation plays must still be investigated in order to understand their importance in communicative signaling relative to eye-gaze, which is a well-known cue in social interaction (de C Hamilton, 2016).

The aim of our second experiment was therefore to determine if any of the aforementioned features of communicative manual acts are as important for signaling the intention to communicate as addressee-directed eye-gaze. To this end, we used a selection of the videos produced in our first experiment and asked a new set of participants to classify each video as communicative or non-communicative.

8. Methods – Experiment II

8.1. Participants

Twenty participants were included in this study, recruited from the Radboud University. Participants were selected on the criteria of being aged 18 – 35, right-handed, healthy, native Dutch-speakers, and without having participated in the previous experiment. The experimental procedure was in accordance with a local ethical committee.

8.2. Materials

Eighty videos (of the 2480) recorded from experiment I were selected for inclusion in this experiment. To provide a representative sampling of each of the two groups, all individual items from all subjects included in the previous experiment were ranked according to eye-gaze and overall kinematics (z- scores). The two groups were ordered such that items in the more communicative context with high communicative context with low eye-gaze and kinematic values were ranked higher than those with low values. This placed all items on a continuum that ranks how representative their features are of their respective groups. This was done due to the observation that, due to the subtle manipulation of context in Experiment I, there was considerable overlap of behavior in the lower ends of each spectrum (i.e. Some participants in the more communicative context showed behavior more similar to those of the less communicative context, and vice versa). Due to the necessarily restricted number of videos to be included in this experiment, we chose to include items which represented a spectrum of eye-gaze and kinematic features representative of their respective context. It should be noted that although this method allowed a more clear separation of the contexts, our further selection procedure (described below) ensured that items were included across a wide range of this ranked continuum. Included items were therefore not the extreme ends only, as shown in Fig. 4.

After creating the ranked continuum of items, inclusion moved from highest to lowest ranked items. Each of the 31 items, as defined in Methods I – Items, was included a minimum of two times and maximum of three times across the entire selection, while ensuring that each item

also appeared at least once as an action and once as a gesture, and at least once in more-communicative context and once in the less-communicative context. This was done to ensure an equal representation of each item across modalities and contexts. One action and one gesture video was included from each participant in Experiment I. This ensured that when watching the videos participants of Experiment II would be less likely to learn the context of any given actor (Experiment I participant).

8.3. Procedure

Before beginning the experiment, participants were given a brief description of the task in order to inform them of the nature of the stimuli. This ensured that participants knew to expect both actions and gestures, and that this was not relevant for their task. Participants were seated in front of a 24″ Benq XL2420Z monitor with a standard keyboard for responses. Stimuli were presented at a frame rate of 29 frames per second, with a display size of 1280 × 720. During the experiment, participants would first see a fixation cross for a period 1000 ms with a jitter of 250 ms. One of the item videos was then displayed on the screen, after which the first question appeared: "Was the action performed for the actor self or for you?" Participants could respond with the 0 (self) or 1 (you) keys on the keyboard. Actions classified as being performed for the actor self were considered non-communicative, while those classified as being performed for "you" (in this case, the participant) were considered communicative. Immediately after answering, participants received the next question prompt: "How certain are you about your decision?" Participants could then respond with the 0 – 5 number keys, representing a range from "very uncertain" (0) to "very certain" (5), as was also indicated on the screen. After 40 items, participants were informed via the computer screen that they were halfway through the experiment, and were allowed to take a short break if needed. Probe trials were presented every 7–9 trials, in which participants were additionally asked what had made them more or less certain about their judgment. For this question, free response typed answers were recorded. These trials were not used for statistical analysis. Context judgments were recorded for each trial, as well as the accuracy of the response.

8.4. Data analysis

Overall performance reflected the accuracy of classifying less-communicative videos as being performed for the actor self, and more-communicative videos being performed for the participant. Before any analyses were performed, we removed outliers in two steps. First, we determined whether there were any participants with outlying performance accuracy, reflected by mean accuracy of less than 2.5 SDs below the mean. After removing any outlying participants, we then calculated mean RT across all participants and excluded any single trials where RT was less than 2.5 SDs below the mean. In order to determine the overall accuracy of performance, a one-sample t-test with test-value = 50 was performed to test if accuracy was greater than chance. Chi-square tests were used to determine if accuracy was equal in both modalities, as well as in both contexts (i.e. To test whether context judgment was more difficult for actions or gestures, or for discriminating one context over the other).

To assess the contribution of eye-gaze and kinematic features to the judgment of communicative context, we performed a two-step linear mixed-effects logistic regression, with context judgment as the dependent variable. Before building the models or differentiating between action and gesture, we tested all predictor variables (eye gaze and kinematic features) for multicollinearity by calculating the variance inflation factor (VIF) using the methodology of Zuur and colleagues (Zuur, Ieno, & Elphick, 2010). Predictors with a VIF greater than three were excluded from all subsequent models.

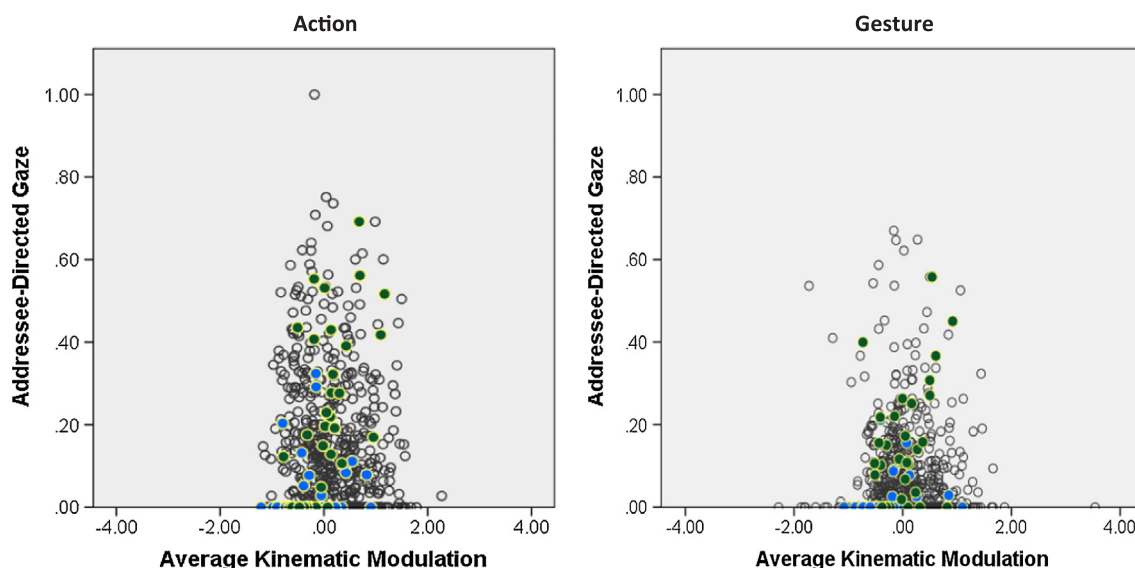Statistical models were assessed for actions and gestures in order to

Fig. 4. Selection of items used in Experiment II. The left plot shows Action items. The right plot shows Gesture items. In both plots, the x-axis represents the mean modulation of the five kinematic features from Experiment I (distance, maximum amplitude, hold-time, sub-movements, and peak velocity). The y-axis represents proportional addressee-directed eye-gaze. Filled blue circles depict the selected less-communicative items, while filled green circles depict the selected more-communicative items, and empty black circles depict the remaining non-selected items.

test for differences in relevant predictor variables, and utilized the modulation values described in Methods – Experiment I, *Data Processing*. We included both correct and incorrect judgments in our statistical model as we were most interested in the perceived context. In the first step of the regression we included eye-gaze as the predictor variable, as eye-gaze is recognized in the literature as a highly salient cue for communication (Csibra & Gergely, 2009). In the second step of the regression model we included all kinematic features that were not previously excluded due to multicollinearity, thereby ensuring the models for action and gesture were alike. We used a likelihood ratio test to compare the two steps of the model, thereby assessing the additional contribution of kinematics to the prediction of communicative context, over and beyond the (expected) contribution of eye-gaze. The contributions of individual predictors (ie. eye gaze and individual kinematic features) are additionally reported in order to show the relative weight of each predictor in the complete model. Random intercepts were included for actor and item at each step of the model.

Certainty was assessed in two domains: first, the effect of modality and context was determined using Welch's *t*-tests, as implemented by R. This approach corrects for (potential) inequalities of variance, thereby providing a more robust comparison of the means. Second, the contribution of eye-gaze and kinematic features on an observer's context judgment was determined using a linear mixed-effects regression. Following the same block procedure as described for the logistic regression we included certainty as the dependent variable, with eye-gaze in a first predictive step of the model and kinematic features (modulation values) in the second step. In order to test the significance of eye-gaze, we again used a likelihood ratio test comparing the model that included eye-gaze as a predictor against the model that only contained the random effects. For these models, random intercepts were again included for actor and item. We additionally modeled random slopes for judgment together with each predictor variable at both steps of the model. This was done because we predict that kinematic modulation and direct eye-gaze are positively associated with judging an act to be communicative, therefore the predictor variables should be positively associated with certainty when the video was judged to be communicative, but negatively associated with certainty when the video was judged to be less-communicative.

9. Results – Experiment II

One participant was excluded due to outlying classification accuracy, and an additional 43 trials were excluded due to slow RT. Analysis of multicollinearity revealed a VIF of 3.12 for Distance, leading us to discard this feature from all subsequent analyses. After removing Distance, the VIF of all remaining predictors was found to be less than two.

Overall performance in classifying context was 60.86%, which was significantly greater than the 50% chance level, $t(18) = 8.68$, $p < 0.001$. Performance was significantly better in recognizing less-communicative (67% accuracy) compared to more-communicative (57% accuracy) contexts, $t(35.97) = 2.49$, $p = 0.017$. We found only marginally higher accuracy in classifying gestures (M = 62.48%, SD = 0.06) compared to actions (M = 59.20%, SD = 0.08), $t (34.34) = -1.428$, $p = 0.16$.

Eye-gaze was a strong predictor for context judgment in both actions (parameter estimate = 7.87, error = 1.78, z = 4.41, $p < 0.001$) and gestures (parameter estimate = 8.48, error = 1.09, z = 7.72, $p < 0.001$). Adding kinematics did not contribute to the model for actions ($\chi^2(4) = 4.15$, $p = 0.39$) or gestures ($\chi^2(4) = 0.56$, $p = 0.97$). An overview of the model results can be seen in Table 1, including the parameter estimate, the standard error of the estimate, and the associated Z-score of each predictor in the full model. Z-scores with a p-value less than 0.05 are marked with a '*', and those with a p-value of less than 0.01 are marked with a '**'. We report here the statistics for eye-gaze from the first step of the model, and the statistics of the kinematics from the second step.

Certainty in the less-communicative context judgments (M = 3.53, SD = 0.69) was not significantly different than certainty in the more-communicative context (M = 3.64, SD = 0.50), $t(33.02) = 0.54$, $p = 0.592$. Certainty when judging actions (M = 3.65, SD = 0.56) was not significantly different compared to when judging gestures (M = 3.52, SD = 0.65), $t(35.36) = 0.65$, $p = .529$. In both actions and gestures, eye-gaze showed a linear relation with certainty (action: $\chi^2(3) = 8.17$, $p = 0.043$; gesture: $\chi^2(3) = 17.80$, $p < 0.001$), with increased direct eye-gaze changing certainty by 0.16 ± 1.65. This change was positive or negative depending on whether the video was judged to be communicative or non-communicative (see Supplementary Fig. 2). Including kinematics did not significantly improve this model

Table 1
Effect of eye-gaze and kinematic modulation on context judgments.

Model parameter	Action			Gesture		
	Parameter estimate	Std. error	Z	Parameter estimate	Std. error	Z
Eye-gaze	7.69	1.62	4.73[**]	8.31	1.37	6.07[**]
Max. Amplitude	0.35	0.21	1.72	0.09	0.19	0.46
Hold-time	0.16	0.19	0.83	0.06	0.13	0.47
Submovements	0.06	0.19	0.31	−0.16	0.23	−0.67
Peak Velocity	0.15	0.33	0.44	0.11	0.13	0.47

[*] $p < 0.05$.
[**] $p < 0.01$.

for actions ($\chi^2(16) = 6.86$, $p = 0.976$) or gestures ($\chi^2(16) = 2.97$, $p = 0.999$).

10. Conclusion and discussion – Experiment II

A communicative context is dependent upon interaction, and thus recognition of the communicative intention by the addressee. We therefore sought with our second experiment to examine the role of communicative acts from the standpoint of the addressee. The optimality principle of motor control (Todorov & Jordan, 2002), together with that of contextual efficiency (Gergely & Csibra, 2003), suggests a dynamic (ie. variable), yet effectively constrained system of action production. We suggested that a deviation from these efficiency principles would be noticeable by an observer, and thereby used as a signal of intention.

Contrary to our initial hypothesis we found that kinematics do not contribute to an observer's recognition of communicative intent. Instead, observers rely much more on addressee-directed eye-gaze. Our second experiment therefore lends additional evidence to the idea that eye-gaze cues may be the most important indicator of communicative intent for the addressee (Csibra & Gergely, 2009). Although the suggestion that intention can be read from kinematics (Ansuini, Cavallo, Bertone, & Becchio, 2014; Becchio, Manera, et al., 2012) finds support in the literature, it may be that eye-gaze is such an important cue for recognizing intentions that it overrides kinematic information when both are available. Rather than an interaction, the two cues may alternatively be seen as a hierarchy with regard to cue importance. To test this assumption we conducted the third experiment to determine whether detecting of intentions from kinematics could be limited to a particular modality (actions or gestures), or to situations where eye-gaze information is unavailable.

11. Methods – Experiment III

11.1. Participants

Twenty naïve participants were included in this study, recruited from the Radboud University. Participants were selected on the criteria of being aged 18 – 35, right-handed, healthy, native Dutch-speakers, and without having participated in either of the previous experiments.

The experimental procedure was in accordance with a local ethical committee.

11.2. Materials

The same selection of videos was used as in Experiment II, but with the faces of the actors obscured in order to remove the possibility of using eye-gaze information. In order to obscure the faces, we utilized the Mosaic feature in Adobe Premiere Pro to create a pixilated oval (pixel size = 80 × 80) which covered the entire face in each of the videos.

11.3. Procedure and data analysis

Experimental procedure and data analysis were carried out exactly as in Experiment II, except that eye-gaze is excluded from these models. Therefore, each kinematic model is compared for significance against a null model in which only the random factors for actor and item are present.

12. Results – Experiment III

Due to a technical issue, 40 trials from one participant were lost from the initial dataset. One participant was excluded due to outlying performance accuracy, and an additional 26 trials were removed due to outlying RT. Multicollinearity test revealed distance to have a VIF of 3.21, leading us to exclude it from further analyses. After removing distance, the remaining predictors had VIFs of less than two.

Overall accuracy of context judgments was 52.47%, which was significantly above chance level, $t(18) = 2.99$, $p = 0.008$. We found no difference in accuracy when judging communicative (M = 51.61%, SD = 0.05) compared to less communicative (M = 53.18%, SD = 0.06) videos, $t(36.49) = 0.82$, $p = 0.419$. We similarly found no difference when judging actions (M = 52.51%, SD = 0.06) compared to gestures (M = 52.44%, SD = 0.05), $t(35.94) = 0.04$, $p = 0.967$.

In actions, kinematics contributed to a near-significant increase in the model fit, $\chi^2(4) = 9.42$, $p = 0.051$. In gestures kinematics contributed to a significant improvement to the model, $\chi^2(4) = 14.65$, $p = 0.005$. An overview of the parameter estimates, standard error, and z-scores for each predictor in the full model can be seen in Table 2. Z-

Table 2
Effect of (non-visible) eye-gaze and kinematic modulation on context judgments.

Model parameter	Action			Gesture		
	Parameter estimate	Std. error	Z	Parameter estimate	Std. error	Z
Max. Amplitude	0.19	0.13	1.50	0.32	0.10	3.14[**]
Hold-time	−0.14	0.12	−1.13	0.05	0.06	0.81
Submovements	0.15	0.12	1.23	0.17	0.14	1.17
Peak Velocity	−0.16	0.20	−0.77	0.01	0.12	0.09

[*] $p < 0.05$.
[**] $p < 0.01$.

scores with a p-value < 0.05 are marked with a '*', and those with a p-value of less than 0.01 are marked with a '**'.

When judging actions, kinematics did not influence certainty ($\chi^2(4) = 1.55$, $p = 0.818$). In gestures, we found that kinematic modulation influences certainty ($\chi^2(4) = 14.46$, $p = 0.009$), with maximum amplitude increasing certainty by 0.13 ± 0.05. The other kinematic features did not show strong associations with certainty. To make sure that the differences in the model outcomes between the Experiment II and III are not simply due to differences in the statistical models (eye-gaze included in the model of Experiment II but not III), we tested the same model with eye-gaze in Experiment III. The model produces highly comparable results, with identical overall outcomes (statistics not shown) in context judgment, but no significant relation between kinematics and certainty. Thus the difference in the models' outcomes reflects a difference in how participants perceived the videos in the two experiments.

13. Conclusion and discussion – Experiment III

The results of this study show a marginally better-than-chance recognition of more- compared to less-communicative actions and gestures, and also indicate that both modalities (actions and gestures) and contexts (more- and less-communicative) are recognized with similar levels of accuracy. We further show that while eye-gaze was not associated with context judgments in either modality, increased kinematic modulation was predictive of gestures being judged as more-communicative. Specifically, increasing maximum amplitude of a gesture leads to it being perceived as more communicative.

Accuracy of recognition was similar in the action modality, but without being significantly explained by kinematic modulation. This suggests that the kinematics of actions, although modulated similarly, may be interpreted differently by addressees, in a way that was not quantified by the present study. For example, gestures are reliant primarily on movement in order to convey meaning, which could lead to a bias towards kinematic modulation of this movement being the primary salient cue of communicative intent. The objects being manipulated in manual actions, on the other hand, may be more salient for the observer than the fine-grained kinematics of the movements. In this way, besides the kinematic modulation that we found in Experiment I, participants in the more-communicative context may have modulated additional cues, such as orienting the object towards the observer to make the object manipulation more clear. This in turn could have driven the accuracy of context judgment in this experiment. As such features of action production were not *a priori* hypothesized and thus also not measured, additional research is needed in order to shed light on whether actions use different cues, such as the objects themselves, to convey communicative intent.

In gestures on the other hand, we see a strong relation between increased maximum amplitude and a higher rate of being perceived as more-communicative. That this effect is present in the gesture modality, despite low accuracy, suggests that participants were more receptive to the kinematic modulation in gestures, and more readily interpreted them as more communicative. Although speculative, this would be in line with theories by Goldin-Meadow and colleagues suggesting that gestures have a special role in communication that is stronger than actions, as the latter is not by default used for communication (Goldin-Meadow, 2017). As such may be more likely to be interpreted as intended for someone besides the actor (Kelly et al., 2015; Novack et al., 2016).

These results highlight the difficulty of recognizing communicative context from kinematics alone. However, the results also indicate that, at least in gestures, kinematic modulation may play a role in guiding this recognition process. This finding is intriguing given that the kinematic modulation in the present stimuli set was highly subtle, with a large overlap between the less- and more-communicative contexts.

14. General discussion

In this study we set out to characterize the initiation of a communicative interaction in both production and comprehension. To do this, we first used motion-tracking and automatic feature calculation to quantify spatial and temporal kinematic features and accompanying eye-gaze behavior of communicative actions and gestures (production), and then assessed the contribution of kinematic modulation and addressee-directed eye-gaze to the judgment of communicative context by addressees (comprehension). Overall, our results show that space-time dimensions of both action and gesture kinematics are modulated by a communicative context. Addressee-directed eye-gaze is also increased in the communicative context and is the best determinant of an observer's classification of an act as being communicative, although kinematic modulation plays a role when eye-gaze information is unavailable.

Results from our first experiment showed that in a more-communicative context both actions and gestures are made larger, with greater vertical amplitude and with a more complex movement pattern when compared to a less-communicative context. Additionally, we find increased addressee-directed eye-gaze in the more-communicative context. This finding is in agreement with previous studies showing increased addressee-directed gaze in more communicative contexts, and further supports the notion that this effect is not simply reliant on the participant being watched (as was true in both the more- and less-communicative contexts of our experiment), but that it is directly related to the communicativeness of the context. Our finding of kinematic modulation is in line with research on infant-directed gestures. Infant directed actions show evidence for 'motionese', a form of kinematic modulation which is argued to help sustain attention in infants as well as to make action intentions more legible (Brand et al., 2002). Specifically, this kinematic modulation includes a greater range of motion as well as increased 'punctuality', a qualitative measure of fluidity versus segmented movement. While range of motion can be seen as a parallel of the distance measure in our study, punctuality may also reflect our quantification of submovements and holds. We similarly found more submovements and, at least in gestures, a trend-level increase in communicative holds, which may reflect the more segmented movement profile described by Brand and colleagues. This similarity provides support for our results, as motionese can be seen as an exaggeration of communicative gestures in general. Our finding of kinematic modulation may therefore be a functionally similar exaggeration. For communication with adults, we exaggerate the kinematics of our movements; for communication with children, we exaggerate kinematics even more. In addition to showing this exaggeration occurs in both actions and gestures, we additionally expand the fundamental framework in which these modulations can be seen by proposing that kinematic modulation is an extension of motor efficiency that optimizes the space–time dimension of communicative acts. This work therefore bridges earlier behavioural studies (Brand et al., 2002; Campisi & Özyürek, 2013) with computational models (Pezzulo et al., 2013) using modern motion tracking and automatic feature quantification to define specific kinematic features relating to the spatial and temporal characteristics of actions and gestures.

Results from our second experiment showed that addressee-directed eye-gaze remains the most salient cue for recognizing an act as being communicative. While previous studies have suggested that a communicative intention can be read from kinematics (Ansuini, Santello, Massaccesi, & Castiello, 2006; Becchio, Manera, et al., 2012), our study suggests that kinematics are not a primary source of information for this classification and thus may be part of a hierarchy, with eye-gaze being the primary cue for reading communicative intent.

Our third experiment attempted to disentangle eye-gaze from kinematics by occluding facial information. Results from this experiment showed that, at least in gestures, spatial information can act as a cue to communicative intent. Although the correlation between kinematic

features and intention recognition did not hold for actions, we speculate that this may be related to the magnitude of the effect. Upon visual inspection of the production data from Experiment I, vertical amplitude is the most strongly modulated kinematic feature, and this appears more pronounced in gesture than actions. Similarly, vertical amplitude in gestures is the only feature that is found to be a significant predictor of intention recognition in Experiment III.

As eye-gaze is known to have a strong impact on attention and cognitive processing (Calder et al., 2002), these results suggest that kinematics are simply lower in a hierarchy for intention recognition. The dominance of eye-gaze as a signal for communicative intention does not mean kinematic modulation is entirely useless to the addressee, as it can also be used as a cue for intention when more primary social cues are obscured. One possibility is that the modulation that participants exhibited in our study (Experiment I) may be specific to their (perceived) visibility. In our study, participants were aware that they were being watched via a video camera, which may have influenced their communicative strategy. For example, while studies have shown kinematic modulation in face-to-face interactions (McEllin, Knoblich, & Sebanz, 2017; Sartori et al., 2009), others have shown that individuals playing the role of 'leader' in a joint action, which can be considered a form of communication, only modulate their kinematics when they know that their movements are fully visible to their partner (Sartori et al., 2009; Vesper & Richardson, 2014). This suggests that communicative cues are used dynamically based on situational constraints. In our case, since eye-gaze is the strongest cue, kinematics may have indeed been modulated less because the cue is less salient. If participants were made aware that their eye-gaze would not be visible, kinematic modulation may therefore be further exaggerated. However, the primary role of kinematic modulation may lie elsewhere in the communicative interaction, such as in clarifying the semantic content being communicated.

Communication requires both the recognition of the intention to communicate as well as comprehension of the semantic content being conveyed. We suggest that kinematic modulation occurs in order to enhance the saliency or legibility of the semantic content being communicated (i.e. the specific movements or their meanings). In this view, eye-gaze signals the intention to communicate, while the kinematics are modulated in order to make the message more easily understood. While speculative, this theory is in line with the interpretation of kinematic modulation in motionese as enhancing action legibility (Brand et al., 2002). In this view, larger, more punctuated actions are thought to make the semantic content more legible. Although legibility was not directly tested by Brand et al., later studies showed that mothers begin exaggerating their action kinematics when infants are capable of learning the action (Fukuyama et al., 2015), infants prefer watching actions featuring motionese (Brand & Shallcross, 2008), and children are more likely to reproduce these actions (Williamson & Brand, 2014). Furthermore, studies in joint actions in adults also reveals actions that direct the attention of the addressee to a certain object using "an exaggerated manner or conspicuous timing" (Clark, 2005) which may be analogous to spatial and temporal modulation of kinematics. Robotics research, which combines theory-based robotic production of gestures or actions with validation through human comprehension experiments, supports the notion that exaggeration of kinematics improves semantic interpretation of a manual act (Dragan & Srinivasa, 2014; Holladay, Dragan, & Srinivasa, 2014). This theory has also been explored in the framework of computational modeling, where movement trajectories are modulated to disambiguate the end-goal (Pezzulo et al., 2013). Together, these findings suggest that kinematic modulation may play a role in learning and communication when semantic content needs to be made clear. By modulating the kinematics to be optimally unambiguous, the communicator is thus able to optimize the space–time dimensions of the interaction.

On the other hand, eye-contact is a strong social cue (Calder et al., 2002; Senju & Csibra, 2008) that initiates a pedagogical stance even

early in life (Csibra & Gergely, 2009; Senju & Csibra, 2008; Williamson & Brand, 2014). Although cognitively separate from the processing of action semantics (Rizzolatti et al., 2014), this initiation of interaction may therefore be necessary to prepare the addressee to benefit from kinematic modulation. We speculate that kinematic modulation likely serves another purpose in human communication, i.e., to enhance the saliency or legibility of the semantic content being communicated, but can also serve as a cue for intention recognition when more primary cues, such as eye-gaze, are not available. Future studies are, however, needed in order to bring further light to this hypothesis.

14.1. Strengths and limitations

Our study provides novel insights into the kinematics of communicative actions and gestures. Using robust motion-tracking technology we were able to automatically quantify several kinematic features, which relate to different spatial and temporal components of the act's kinematic profile. This lends precision to our results and may provide a framework for future studies examining kinematic features of actions or gestures. Furthermore, the naturalistic elicitation of more- and less-communicative contexts provides ecological validity to our results, in that participants performed ordinary, everyday acts, such as pouring water or slicing bread) without the use of physical markers being placed on the body. Our study is also the first to examine actions and gestures within the same framework of communicative contexts and manual acts, providing a novel investigation of the similarities and differences between the two modalities. Especially in regard to using the same manual acts in both communicative contexts, we are able to attribute kinematic differences to the context itself, while avoiding differences due to different motor end goals intentions(van Elk, van Schie, & Bekkering, 2014). Finally, the relatively large sample size (n = 40) and variety of action/gesture pairs used (n = 31) provides evidence for the external validity of our findings.

While the naturalistic setting of our study provides ecological validity, we recognize that this comes at the cost of some control over experimental variables. As participants were never specifically asked to be communicative, we rely on the assumption that the subtle manipulation of instructions elicited genuinely communicative behavior. Given the significant performance in context judgment in the second experiment, however, we believe that our context distinction is valid. Lastly, our study was limited in its ability to directly compare actions and gestures statistically due to the methodology used. While this methodology allowed investigation of many different acts, and thus allows generalization of these findings to other acts, it also hindered us from making between-modality comparisons. The difference in significant results between actions and gestures, however, allows some conclusions to be drawn regarding the differences in kinematic modulation. Finally, the subtle elicitation of the more communicative context may have led to kinematic differences between the two contexts that are too subtle or variable to entirely separate.

15. Conclusion

In summary, we examined the features characterizing the initiation of a communicative interaction, examining both the production and comprehension of actions and gestures. We found that a communicative context elicits kinematic modulation of both actions and gestures, together with an increase in addressee-directed eye-gaze. While eye-gaze strongly contributes to the recognition of communicative contexts, kinematic modulation only serves this purpose in gestures when eye-gaze information is unavailable. We suggest that eye-gaze is primarily responsible for initiating the interaction, while kinematics may contribute to enhancing the legibility of the movement, potentially facilitating transmission of the semantic content of the communicative act.

Data archiving

Final data and analysis scripts are available in the following repository: http://hdl.handle.net/11633/di.dcc.DSC_2016.00227_393.

Acknowledgements

Appendix A. Production instructions

Original (Dutch)	English
doe de apple in de kom	Place the apple in the bowl
borstel je haar met de borstel	Brush your hair with the brush
veeg het papier af	Brush off the paper
kreukel het papier	Crumple the paper
snij het brood met de mes	Cut the bread with the knife
knip het papier doormidden	Cut the paper in half
wis de figuur met de gom	Erase the figure with the eraser
vouw het papier doormidden	Fold the paper in half
sla de spijkers met de hammer	Hammer the nails with the hammer
meet het papier met het meetlint	Measure the paper with the measuring tape
open het potje	Open the jar
open het slot met de sleutel	Open the lock with the key
pel de banaan	Peel the banana
doe het dopje op de pen	Put the pencap on the pen
giet het water in het glas	Pour the water in the glass
doe de hoed op	Put on the hat
doe de ring aan	Put on the ring
verwijder het kurkje van de fles	Remove the cork from the bottle
verwijder het dopje van de pen	Remove the pencap from the pen
schrob het bureau met de spons	Scrub the desk with the sponge
schud de kaarten door elkaar	Shuffle the cards
pers de citroen uit	Squeeze the lemon
stapel de blokken op elkaar	Stack the blocks on top of each other
stempel het papier	Stamp the paper
niet de papieren samen	Staple the papers together
dompel het theezakje in het water	Steep the teabag in the water
roer de thee met de lepel	Stir the tea with the spoon
doe de zonnebril op	Put on the sunglasses
scheur het papier doormidden	Tear the paper in half
gooi de dobbelstenen	Roll the dice
schrijf je naam op het papier met de pen	Write your name on the paper with the pen

Appendix B. Calculation of kinematic features

Spaces for the Vertical Amplitude feature were dynamically defined in equal distances between the midline of the torso, base of the neck, and top of the head at each frame of acquisition. This yielded a total of 5 heights that were dependent on the height of the participant and their current body position. For a visual depiction of the spaces defined, see Supplementary Fig. 1.

Submovements were defined by using the velocity profile of a given hand. Following the description by Meyer and colleagues (Meyer et al., 1988), submovements were operationalized as movements that exceed a given velocity threshold, with the beginning and end marked by either the crossing of a near-zero velocity threshold (going from static to moving) or showing a secondary acceleration (reversal from deceleration to acceleration). We used a standard peak analysis to determine the total number of peaks within the velocity profile of each hand that can be considered submovements. For our study, we assigned a minimum velocity threshold of 0.2 m per second, a minimum distance between peaks of 8 frames, and a minimum peak height and prominence of 0.2 m.

Appendix C. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.cognition.2018.04.003.

References

Ansuini, C., Cavallo, A., Bertone, C., & Becchio, C. (2014). The visible face of intention: Why kinematics matters. *Frontiers in Psychology, 5*, 815. http://dx.doi.org/10.3389/fpsyg.2014.00815.

Ansuini, C., Santello, M., Massaccesi, S., & Castiello, U. (2006). Effects of end-goal on hand shaping. *Journal of Neurophysiology, 95*(4), 2456–2465. http://dx.doi.org/10.1152/jn.01107.2005.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). *Fitting Linear Mixed-Effects Models using lme4*. Retrieved from < http://arxiv.org/abs/1406.5823 > .

Bavelas, J., Gerwing, J., Sutton, C., & Prevost, D. (2008). Gesturing on the telephone: Independent effects of dialogue and visibility. *Journal of Memory and Language, 58*(2), 495–520. http://dx.doi.org/10.1016/j.jml.2007.02.004.

Becchio, C., Cavallo, A., Begliomini, C., Sartori, L., Feltrin, G., & Castiello, U. (2012). Social grasping: From mirroring to mentalizing. *NeuroImage, 61*(1), 240–248. http://dx.doi.org/10.1016/j.neuroimage.2012.03.013.

Becchio, C., Manera, V., Sartori, L., Cavallo, A., & Castiello, U. (2012). Grasping intentions: From thought experiments to empirical evidence. *Frontiers in Human Neuroscience, 6*(May), 1–6. http://dx.doi.org/10.3389/fnhum.2012.00117.

Becchio, C., Sartori, L., Bulgheroni, M., & Castiello, U. (2008). Both your intention and mine are reflected in the kinematics of my reach-to-grasp movement. *Cognition, 106*(2), 894–912. http://dx.doi.org/10.1016/j.cognition.2007.05.004.

Biswas, K. K., & Basu, S. K. (2011). Gesture recognition using Microsoft Kinect®. *The 5th international conference on automation, robotics and applications* (pp. 100–103). IEEE. http://dx.doi.org/10.1109/ICARA.2011.6144864.

Brand, R. J., Baldwin, D. A., & Ashburn, L. A. (2002). Evidence for "motionese": Modifications in mothers' infant-directed action. *Developmental Science, 5*(1), 72–83. http://dx.doi.org/10.1111/1467-7687.00211.

Brand, R. J., & Shallcross, W. L. (2008). Infants prefer motionese to adult-directed action. *Developmental Science, 11*(6), 853–861. http://dx.doi.org/10.1111/j.1467-7687.2008.00734.x.

Brand, R. J., Shallcross, W. L., Sabatos, M. G., & Massie, K. P. (2007). *Fine-grained analysis of motionese: Eye gaze, object exchanges, and action units in infant-versus adult-directed action.* Wiley Blackwell.

Calder, A. J., Lawrence, A. D., Keane, J., Scott, S. K., Owen, A. M., Christoffels, I., & Young, A. W. (2002). Reading the mind from eye gaze. *Neuropsychologia, 40*(8), 1129–1138. http://dx.doi.org/10.1016/S0028-3932(02)00008-8.

Campisi, E., & Özyürek, A. (2013). Iconicity as a communicative strategy: Recipient design in multimodal demonstrations for adults and children. *Journal of Pragmatics, 47*(1), 14–27. http://dx.doi.org/10.1016/j.pragma.2012.12.007.

Cary, M. S. (1978). The role of gaze in the initiation of conversation. *Social Psychology, 41*(3), 269. http://dx.doi.org/10.2307/3033565.

Chang, C.-Y., Lange, B., Zhang, M., Koenig, S., Requejo, P., Somboon, N., & Ellipsis Rizzo, A. (2012). Towards pervasive physical rehabilitation using microsoft kinect. *6th International conference on pervasive computing technologies for healthcare* (pp. 159–162). . http://dx.doi.org/10.4108/icst.pervasivehealth.2012.248714.

Chu, M., & Kita, S. (2011). The nature of gestures' beneficial role in spatial problem solving. *Journal of Experimental Psychology: General, 140*(1), 102–116. http://dx.doi.org/10.1037/a0021790.

Clark, H. H. (2005). Coordinating with each other in a material world. *Discourse Studies, 7*(4–5), 507–525. http://dx.doi.org/10.1177/1461445605054404.

Csibra, G., & Gergely, G. (2009). Natural pedagogy. *Trends in Cognitive Sciences, 13*(4), 148–153. http://dx.doi.org/10.1016/j.tics.2009.01.005.

de C Hamilton, A. F. (2016). Gazing at me: The importance of social meaning in understanding direct-gaze cues. *Philosophical Transactions of the Royal Society of London, Series B, Biological Sciences, 371*(1686), 20150080. http://dx.doi.org/10.1098/rstb.2015.0080.

Dragan, A., & Srinivasa, S. (2014). Integrating human observer inferences into robot motion planning. *Autonomous Robots, 37*(4), 351–368. http://dx.doi.org/10.1007/s10514-014-9408-x.

Fernández-Baena, A., Susín, A., & Lligadas, X. (2012). Biomechanical validation of upper-body and lower-body joint movements of Kinect motion capture data for rehabilitation treatments. *Fourth international conference on intelligent networking and collaborative systems* (pp. 656–661). IEEE. http://dx.doi.org/10.1109/iNCoS.2012.66.

Friston, K. (2011). What is optimal about motor control? *Neuron, 72*(3), 488–498. http://dx.doi.org/10.1016/j.neuron.2011.10.018.

Fukuyama, H., Qin, S., Kanakogi, Y., Nagai, Y., Asada, M., & Myowa-Yamakoshi, M. (2015). Infant's action skill dynamically modulates parental action demonstration in the dyadic interaction. *Developmental Science, 18*(6), 1006–1013. http://dx.doi.org/10.1111/desc.12270.

Galati, A., & Brennan, S. E. (2014). Speakers adapt gestures to addressees' knowledge: Implications for models of co-speech gesture. *Language, Cognition and Neuroscience, 29*(4), 435–451. http://dx.doi.org/10.1080/01690965.2013.796397.

Gergely, G., & Csibra, G. (2003). Teleological reasoning in infancy: The naïve theory of rational action. *Trends in Cognitive Sciences, 7*(7), 287–292. http://dx.doi.org/10.1016/S1364-6613(03)00128-1.

Gerwing, J., & Bavelas, J. (2004). Linguistic influences on gesture's form. *Gesture, 4*(2), 157–195. http://dx.doi.org/10.1075/gest.4.2.04ger.

Goldenberg, G., Hartmann, K., & Schlott, I. (2003). Defective pantomime of object use in left brain damage: Apraxia or asymbolia? *Neuropsychologia, 41*, 1565–1573. http://dx.doi.org/10.1016/S0028-3932(03)00120-9.

Goldin-Meadow, S. (2017). Using our hands to change our minds. *Wiley Interdisciplinary Reviews: Cognitive Science.* http://dx.doi.org/10.1002/wcs.1368.

Hermsdörfer, J., Li, Y., Randerath, J., Goldenberg, G., & Johannsen, L. (2012). Tool use without a tool: Kinematic characteristics of pantomiming as compared to actual use and the effect of brain damage. *Experimental Brain Research, 218*(2), 201–214. http://dx.doi.org/10.1007/s00221-012-3021-z.

Hilliard, C., & Cook, S. W. (2016). Bridging gaps in common ground: Speakers design their gestures for their listeners. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 42*(1), 91–103. http://dx.doi.org/10.1037/xlm0000154.

Holladay, R. M., Dragan, A. D., & Srinivasa, S. S. (2014). Legible robot pointing. *The 23rd IEEE international symposium on robot and human interactive communication, 2014 RO-MAN*.

Holler, J., & Beattie, G. (2005). Gesture use in social interaction: how speakers' gestures can reflect listeners' thinking. *2nd conference of the international society for gesture studies (ISGS): interacting bodies* (pp. 1–12). Retrieved from < http://gesture-lyon2005.ens-lyon.fr/article.php3?id_article=259 > .

Holler, J., & Wilkin, K. (2011). An experimental investigation of how addressee feedback affects co-speech gestures accompanying speakers' responses. *Journal of Pragmatics, 43*(14), 3522–3536. http://dx.doi.org/10.1016/j.pragma.2011.08.002.

Innocenti, A., de Stefani, E., Bernardi, N. F., Campione, G. C., & Gentilucci, M. (2012). Gaze direction and request gesture in social interactions. *PLoS ONE, 7*(5), e36390. http://dx.doi.org/10.1371/journal.pone.0036390.

Kampe, K. K. W., Frith, C. D., & Frith, U. (2003). "Hey John": Signals conveying communicative intention toward the self activate brain regions associated with "mentalizing", regardless of modality. *The Journal of Neuroscience, 23*(12), 5258–5263. http://dx.doi.org/10.1002/hbm.21164.

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. http://dx.doi.org/10.3758/s13423-014-0681-7.

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

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*(3–4), 547–556. http://dx.doi.org/10.1007/s00221-011-2649-4.

McEllin, L., Knoblich, G., & Sebanz, N. (2017). Distinct kinematic markers of demonstration and joint action coordination? Evidence from virtual xylophone playing. *Journal of Experimental Psychology. Human Perception and Performance.* http://dx.doi.org/10.1037/xhp0000505.

McNeill, D. (1994). *Hand and mind: What gestures reveal about thought. Leonardo. Vol. 27* Chicago: University of Chicago Press http://dx.doi.org/10.2307/1576015.

Meyer, D. E., Abrams, R.a., Kornblum, S., Wright, C. E., & Smith, J. E. (1988). Optimality in human motor performance: Ideal control of rapid aimed movements. *Psychological Review, 95*(3), 340–370. http://dx.doi.org/10.1037/0033-295X.95.3.340.

Nagels, A., Kircher, T., Steines, M., & Straube, B. (2015). Feeling addressed! The role of body orientation and co-speech gesture in social communication. *Human Brain Mapping, 36*(5), 1925–1936. http://dx.doi.org/10.1002/hbm.22746.

Naish, K. R., Reader, A. T., Houston-Price, C., Bremner, A. J., & Holmes, N. P. (2013). To eat or not to eat? Kinematics and muscle activity of reach-to-grasp movements are influenced by the action goal, but observers do not detect these differences. *Experimental Brain Research, 225*(2), 261–275. http://dx.doi.org/10.1007/s00221-012-3367-2.

Novack, M. A., Wakefield, E. M., & Goldin-Meadow, S. (2016). What makes a movement a gesture? *Cognition, 146*, 339–348. http://dx.doi.org/10.1016/j.cognition.2015.10.014.

Paraskevopoulos, G., Spyrou, E., & Sgouropoulos, D. (2016). A real-time approach for gesture recognition using the Kinect sensor. *Proceedings of the 9th Hellenic conference on artificial intelligence – SETN '16* (pp. 1–4). New York, USA: ACM Press. http://dx.doi.org/10.1145/2903220.2903241.

Pedersoli, F., Benini, S., Adami, N., & Leonardi, R. (2014). XKin: An open source framework for hand pose and gesture recognition using Kinect. *The Visual Computer, 30*(10), 1107–1122. http://dx.doi.org/10.1007/s00371-014-0921-x.

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. http://dx.doi.org/10.1162/jocn_a_00865.

Pezzulo, G., Donnarumma, F., & Dindo, H. (2013). Human sensorimotor communication: A theory of signaling in online social interactions. *PLoS ONE, 8*(11), e79876. http://dx.doi.org/10.1371/journal.pone.0079876.

R Core Team (2013). *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing. URL < http://www.R-project.org/ > .

Rizzolatti, G., Cattaneo, L., Fabbri-Destro, M., Rozzi, S., Abdollahi, R., Jastorff, J., ... Roland, P. (2014). Cortical mechanisms underlying the organization of goal-directed actions and mirror neuron-based action understanding. *Physiological Reviews, 94*(2), 655–706. http://dx.doi.org/10.1152/physrev.00009.2013.

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. http://dx.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. http://dx.doi.org/10.1016/j.concog.2009.06.004.

Schilbach, L., Wohlschlaeger, A. M., Kraemer, N. C., Newen, A., Shah, N. J., Fink, G. R., & Vogeley, K. (2006). Being with virtual others: Neural correlates of social interaction. *Neuropsychologia, 44*(5), 718–730. http://dx.doi.org/10.1016/j.neuropsychologia.2005.07.017.

Senju, A., & Csibra, G. (2008). Gaze following in human infants depends on communicative signals. *Current Biology: CB, 18*(9), 668–671. http://dx.doi.org/10.1016/j.cub.2008.03.059.

Senju, A., & Johnson, M. H. (2009). The eye contact effect: Mechanisms and development. *Trends in Cognitive Sciences, 13*(3), 127–134. http://dx.doi.org/10.1016/j.tics.2008.11.009.

Southgate, V., Chevallier, C., & Csibra, G. (2009). Sensitivity to communicative relevance tells young children what to imitate. *Developmental Science, 12*(6), 1013–1019. http://dx.doi.org/10.1111/j.1467-7687.2009.00861.x.

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

Tomasello, M. (2010). *Origins of human communication.* Cambridge: MIT Press.

van Elk, M., van Schie, H., & Bekkering, H. (2014). Action semantics: A unifying conceptual framework for the selective use of multimodal and modality-specific object knowledge. *Physics of Life Reviews, 11*(2), 220–250. http://dx.doi.org/10.1016/j.plrev.2013.11.005.

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

Williamson, R. A., & Brand, R. J. (2014). Child-directed action promotes 2-year-olds' imitation. *Journal of Experimental Child Psychology, 118*.

Zuur, A. F., Ieno, E. N., & Elphick, C. S. (2010). A protocol for data exploration to avoid common statistical problems. *Methods in Ecology and Evolution, 1*(1), 3–14. http://dx.doi.org/10.1111/j.2041-210X.2009.00001.x.