Can you handle this? The impact of object affordances on how co-speech gestures are produced

Ingrid Masson-Carro, Martijn Goudbeek & Emiel Krahmer

Routledge
Taylor & Francis Group

Can you handle this? The impact of object affordances on how co-speech gestures are produced

Ingrid Masson-Carro, Martijn Goudbeek and Emiel Krahmer

Tilburg Centre for Cognition and Communication (TiCC), University of Tilburg, Tilburg, The Netherlands

ABSTRACT

Hand gestures are tightly coupled with speech and with action. Hence, recent accounts have emphasised the idea that simulations of spatio-motoric imagery underlie the production of co-speech gestures. In this study, we suggest that action simulations directly influence the iconic strategies used by speakers to translate aspects of their mental representations into gesture. Using a classic referential paradigm, we investigate how speakers respond gesturally to the affordances of objects, by comparing the effects of describing objects that afford action performance (such as tools) and those that do not, on gesture production. Our results suggest that affordances play a key role in determining the amount of representational (but not non-representational) gestures produced by speakers, and the techniques chosen to depict such objects. To our knowledge, this is the first study to systematically show a connection between object characteristics and representation techniques in spontaneous gesture production during the depiction of static referents.

Hand gestures produced in conversation convey meaning that is co-expressive with the content of speech (McNeill, 1992). This is particularly true for imagistic or representational gestures (Kendon, 2004; McNeill, 1992), which depict aspects of the objects or scenes they refer to. For instance, when speaking about an eagle, we may spread our arms away from the body symbolising the wings of the eagle, whereas when referring to a house, we may use our index finger to trace an inverted "v", symbolising its roof, and if we speak about our new piano, we may mime the action of playing the piano. These examples highlight how different referents may elicit the use of noticeably different gestural representation techniques such as drawing, imitating an action, etc. (Müller, 1998). Gestures occurring alongside speech are assumed to be spontaneous, i.e. produced without conscious awareness of the speaker (Goldin-Meadow, 2003; McNeill, 1992), and speakers seem to combine the use of these iconic strategies effortlessly (and successfully) when describing referents to an interlocutor. Identifying the factors that influence the choice and combination of representation techniques used by speakers to convey meaning is a central (but understudied) issue in gesture research, and one that may shed light on the nature of the conceptual representations that become active at the moment of speaking. Furthermore, speakers do not gesture about every idea they express in speech. While the amount of gestures produced by speakers is influenced by factors such as the communicative context (for instance, speakers often gesture to highlight information that is new for their addressees, Gerwing & Bavelas, 2004), it could be the case that certain features of objects are naturally more salient to speakers, and thus more likely to be gestured about. In this paper, we argue that the type of imagery that is activated upon perception of different object characteristics plays a role in determining (a) how frequently speakers gesture, and also (b) what manual techniques they may use in representing referents. Particularly, we focus on the effect of object affordances (i.e. action possibilities that objects allow for, Gibson, 1986) as a possible gesture predictor.

Affordances, object recognition, and language production

Affordances (Gibson, 1986) have been defined as potential actions that objects and other entities allow for. For example, a handle affords gripping, just like a doorknob affords twisting or a button affords pressing. According to Gibson (1986), humans are predisposed to pay attention to the affordances of objects. This attentional bias towards graspable or manipulable objects (see, e.g. Handy, Grafton, Shroff, Ketay, & Gazzaniga, 2003) has

led researchers to study the role of action affordances as facilitators of object recognition and categorisation, mainly using neuroimaging techniques and visuomotor priming paradigms. These studies have revealed activation in the premotor areas of the brain (presumably involved in the planning of movement), when participants are presented with manipulable objects during the completion of categorisation tasks (e.g. Gerlach, Law, & Paulson, 2002), laterality effects in motor response to affordance perception (e.g. Tucker & Ellis, 1998), and handshape-affordance congruency effects (e.g. Bub, Masson, & Bukach, 2003; Ellis & Tucker, 2000). Most importantly, these experiments challenge the view that motor planning requires a conscious intention to act.

Object affordances have also been acknowledged to influence language comprehension (Glenberg & Robertson, 2000; for a review see Fischer & Zwaan, 2008). In an experiment in which participants had to make sensibility judgements (i.e. identifying whether a sentence is sensible or not), Glenberg and Kaschak (2002) detected a compatibility effect between grammatical constructions and action understanding. Sentences such as "Andy delivered the pizza to you" were judged faster if the motion performed by the participant during the task (e.g. towards or away from body) would match the direction implied by the sentence. This facilitation effect suggests that processing language entails a certain degree of motor simulation (but note that other accounts have attributed these effects to linguistic, and not necessarily embodied, factors—see, for instance, Louwerse, 2011; or Louwerse & Jeuniaux, 2010, for further discussion). Strengthening these findings, several neuroimaging studies have shown that listening to sentences describing actions triggers the activation of the premotor brain areas related to the body parts involved in such actions (Hauk, Johnsrude, & Pulvermüller, 2004; Tettamanti et al., 2005). Similarly, reading the names of objects that can be grasped (e.g. a grape) or manipulated (e.g. pliers) triggers simulations of grasping and of specific hand configurations (Bub, Masson, & Cree, 2008; Glover, Rosenbaum, Graham, & Dixon, 2004).

In sum, the finding that the processing of action-related visual stimuli and language can evoke appropriate motor responses is relevant for the field of gesture studies: it is conceivable that such affordance-evoked motor responses may be partly responsible for the production of co-speech representational gestures, as has been recently suggested by Hostetter and Alibali (2008).

Affordances and gestures

Gesture and speech seem suited to convey different types of information (Beattie & Shovelton, 2002; Cook &

Tanenhaus, 2009). Gestures occur often with content that is highly imageable (Hadar & Butterworth, 1997), and particularly so when speakers depict events that underlie spatial and motoric information (Chu & Kita, 2008; Feyereisen & Havard, 1999; Hostetter & Alibali, 2008). In these cases, gestures might help get across a meaning that is hard to encode linguistically but that is relatively easy to visualise. For example, Feyereisen and Havard (1999) conducted a series of interviews where they asked specific questions to elicit the activation of motor imagery (e.g. could you explain how to change the wheel of a car or to repair the tire of a bicycle?), of visual imagery (e.g. could you describe your favourite painting or sculpture?), or of no imagery (e.g. do you think more women should go into politics?). They found that speakers produced the highest amount of gestures when speaking about information related to action, and the lowest amount of gestures when speaking about abstract topics that, in principle, did not evoke imagery directly. Indeed, gestures are often depictive of (one's own) motoric experiences, and we could say that the gestures we perform daily reveal something about how we have acquired knowledge (Streeck, 2009).

In light of findings such as the above, Hostetter and Alibali propose their Gestures as Simulated Action (GSA) framework (Hostetter & Alibali, 2008). This framework contends that the gestures that speakers produce stem from the perceptual and motor simulations that underlie thinking and speaking. According to the GSA, one of the chief factors that determine whether a gesture will be produced by a speaker is the strength of activation of the simulated action (p. 503). This rests on the assumption that different types of mental imagery can be organised along a continuum determined by the extent to which they are tied to action simulation. In practice, this implies that simulations of motor imagery (e.g. a person imagines herself performing an action) and of spatial imagery (e.g. a person imagines what an object will look like if perceived from a different angle) have a stronger action component than simulations of visual imagery (e.g. a person mentally visualises a famous painting in detail), and will culminate into higher representational gesture rates.

Two studies investigated the differences in gesture rate when speakers were induced to simulate motor and spatial imagery, as compared with a visual imagery control condition (Hostetter & Alibali, 2010; Hostetter, Alibali, & Bartholomew, 2011). Hostetter and Alibali (2010) showed that speakers gestured more while describing visual patterns that they had manually constructed with matches than while describing patterns they had only viewed. In the second study, Hostetter et al. (2011) presented speakers with sets of arrow patterns, and

asked them to describe the patterns either in the position in which they were presented, or imagining them as they would appear if they were rotated. In this case, too, higher gesture rates were observed when speakers had to simulate rotation, as opposed to when they directly viewed the patterns. Thus, both studies supported the notion that co-speech gestures are produced more frequently following spatial or motoric simulations. Nevertheless, in both studies, speakers still gestured to a fair extent in the (no simulation) control conditions. The authors suggest that visual imagery may in some cases trigger a certain degree of action simulation. For example, in Hostetter and Alibali (2010), participants might have simulated the action of arranging the matches by hand to form the visual patterns they attended to. Similarly, in Hostetter et al. (2011), the stimuli consisted of arrows, which may thus have generated simulations of motion. Taking this into account, it becomes apparent that a clear-cut distinction cannot be made between types of mental imagery, with various types of imagery sometimes becoming simultaneously active.

The first question that we address in this paper relates to whether the perception of objects with a manual affordance (such as tools) will elicit simulations of object use and, hence, result in higher gesture rates. Typically, the perception of static scenes where no animate character or actor is involved should activate simulations of visual imagery, but the motor cognition literature has extensively shown that viewing objects with affordances may generate simulations of object manipulation and object use (e.g. Bub et al., 2003; Ellis & Tucker, 2000; Glover et al., 2004). A handful of recent studies have asked whether objects that afford action performance elicit higher gesture rates during description tasks similar to the experiment reported in the present study (Hostetter, 2014; Pine, Gurney, & Fletcher, 2010) but also during a mental rotation task and a subsequent motion depiction task (Chu & Kita, 2015). In an experiment designed to examine the intrapersonal function of gestures, Pine et al. (2010) presented speakers with pictures of praxic (e.g. scissors, stapler) and non-praxic objects (e.g. fence, chicken), and measured their gesture rates while describing these objects to a listener under different visibility conditions. Their results showed that people produced more gestures in trials corresponding to praxic objects, regardless of whether they could directly see their addressee or not. Using a similar paradigm, Hostetter (2014) asked speakers to describe a series of nouns, and found more gesturing accompanying the descriptions of the items that had been rated highest in a scale of manipulability, also regardless of visibility. Both studies conclude that the likelihood of producing representational gestures is co-determined by the semantic

properties of the words they accompany—specifically, by the motoric component evoked by such words.

While these findings are suggestive, both studies have some limitations which we try to address in the current paper. First of all, in both studies, participants were not allowed to name the objects being described. It is likely that this type of instruction may have biased the speakers' descriptions towards including information about the function of objects when possible, perhaps as the easiest communicative strategy to describe objects. This would make questionable the extent to which speakers gesture more about manipulable objects because of the action simulation that may underlie the representation of such objects, perhaps arguing in favour of an account where function is simply a more salient (and easier to gesturally depict) attribute, that leads to more successful identification.

Secondly, both studies provide no data about the occurrence of other non-representational gesture types (e.g. rhythmic gestures such as beats) in relation to manipulable objects. While it is true that both the study by Pine et al. (2010) and the GSA (Hostetter, 2014; Hostetter & Alibali, 2008) are specific to representational gestures, it may be the case that the activation evoked by descriptions with a strong action component is not restricted to the production of representational gestures, but that it primes gesturing in general. This could support what we may term a general activation account, by means of which the motoric activation evoked by action-related language may lower the speaker's gesture threshold (Hostetter & Alibali, 2008, p. 503) enough to allow for other hand movements to be produced. However, whether both representational and non-representational gestures depend on the same threshold height is not specified in any gesture model to date, and remains to be investigated.

A recent study by Chu and Kita (2015) extends previous research by suggesting that gestures may arise in response to action potential independently of the content of speech, as evidenced by the increase in the number of gestures both while solving a mental rotation task ("co-thought" gestures) and during depictions of motion events (co-speech gestures), where the affordance component of the object presented (in this case, mugs with handles) was task-irrelevant. Furthermore, their study featured a condition in which the affordances of the mugs were obscured, by presenting participants with mugs covered in spikes (minimising grasping potential). In both co-speech and co-thought conditions, participants were less likely to gesture about the mugs in the spiky condition, exposing a fine-grained sensitivity to the affordance of objects in speakers, even when these are task-irrelevant.

So far the few studies that have examined gesture production about objects that afford action performance have mostly looked at the frequency of gesturing. However, gesture rate may not be the only aspect of gesture production influenced by perceiving affordances. Here, we argue that the representation technique chosen to depict a referent (e.g. Kendon, 2004; Müller, 1998; Van Nispen, van de Sandt-Koenderman, Mol, & Krahmer, 2014; Streeck, 2008, 2009) might be susceptible to such influence too. If we think of representational gestures as being abstract materialisations of (selective) mental representations that are active at the moment of speaking, one can think that the techniques chosen to represent these images may reveal something about the nature and quality of the information being simulated by a speaker. Müller (1998) recognises four main representation modes employed by speakers in the construction of meaning. These gestures are perceivably different, and imply varying degrees of abstraction with respect to the referent they represent. These modes include *imitation*, which is by and large the most common technique associated to first-person (enacting) gestures, and consists of miming actions associated to an object; *portrayal*, where the hand represents an object or character, for example the hand pretending to be a gun; *drawing*, where a speaker traces a contour, typically with an extended finger; and *moulding*, where the speaker moulds a shape in the air, as if palpating it. Very little is known about what drives the use of one technique over another and, in general, about what determines the physical form that representational gestures adopt (Bavelas, Gerwing, Sutton, & Prevost, 2008; Krauss, Chen, & Gottesman, 2000).

One factor known to influence gestural representation modes is action observation (e.g. seeing a character perform an action, Parrill, 2010) or action performance (e.g. Cook & Tanenhaus, 2009). For instance, Cook and Tanenhaus (2009) had speakers solve the Tower of Hanoi problem and describe its solution to a listener. Solving this task consists of moving a stack of disks from one peg to another one, using an auxiliary middle peg. Half of the speakers performed the task with real disks, whereas the other half performed the task on the computer, by dragging the disks with the mouse. While no changes were observed in the speech and number of gestures in these two conditions, gestures were qualitatively different. When speakers had performed the actions with real disks, they were more likely to use grasping handshapes, i.e. imitating the action that they just performed. Speakers who solved the task on the computer tended to use drawing gestures, i.e. tracing the trajectory of the mouse on the screen. This suggests that the type of action simulation may have an impact on

the particular representation techniques used by speakers. However, it could also be that these results stem from priming effects, whereby speakers simply "reproduced" the action they had just performed.

Chu and Kita (2015) also suggest a connection between affordance and representation technique. Although their study only included one object type (mugs), their results show that speakers were more likely to use grasping gestures to solve the rotation task when the mugs were presented with a smooth surface (affordance enhanced) as opposed to when the mugs appeared covered in spikes (affordance obscured). Hence, both of these studies highlight the importance of investigating not only the number of gestures produced by speakers, if we are really to understand why we produce gestures at all—as has been emphasised by recent studies on gesture production (e.g. Bavelas & Healing, 2013; Galati & Brennan, 2014; Hoetjes, Koolen, Goudbeek, Krahmer, & Swerts, 2015). Limiting ourselves to annotating the number of gestures produced can be compared to doing speech studies in which only the number of words—but not the content of speech—is analysed.

The present study

In sum, it seems that action simulation plays a role in eliciting gesture production, with recent studies suggesting that higher gesture rates may be evoked by visual inspection of objects that afford action performance, such as tools. Nevertheless, previous research has mainly focussed on analysing gesture rates; therefore, we have little knowledge of how object characteristics influence the strategies that gesturers employ in communicating about them.

The aim of this study is to assess the effects of perceiving objects with different (high and low) affordance degrees, on the production of speech-accompanying gestures during a communication task, focussing on the gestural techniques employed by speakers in the representation of objects. We predict that affordance will determine the number of gestures produced by speakers, with more gestures accompanying the descriptions of manipulable objects, in line with previous research (Chu & Kita, 2015; Hostetter, 2014; Pine et al., 2010). Currently, the predictions made by the GSA (Hostetter & Alibali, 2008) are specific to representational gestures. In this study, we will also annotate the occurrence of non-representational gestures. On the one hand, it is conceivable, given that gestures are seen as outward manifestations of specific imagery simulations, that only the number of representational gestures is influenced by our condition. On the other hand, however, it

could be possible that the activation evoked by action-related language primes the production of hand gestures in general, including non-representational types.

When we look specifically at the presentation of gestures, we expect that communicating about objects that afford actions will trigger more imitation gestures (e.g. where the speaker mimes the function associated with the object) than tracing or moulding gestures (e.g. where the speaker traces or "sculpts" an object's shape), given that the gestures should reflect the type of imagery being simulated at the moment of speaking. Conversely, we do not expect the occurrence of imitation gestures accompanying descriptions of objects that are non-manipulable (although they can occur—e.g. pretending to eat, when describing a table), but a predominance of moulding or tracing gestures.

Method

Participants

Eighty undergraduate students from Tilburg University ($M = 21$; SD = 2; 50 female) took part in this experiment, in exchange for course credit points. All participants were native speakers of Dutch, and carried out the experimental task in pairs.

Material and apparatus

Our stimuli set was composed of pictures of 28 objects: 14 with a high-affordance degree (e.g. whisk), and 14 with a low-affordance degree (e.g. plant) (see Appendix 1 for the complete list of objects). We defined objects with a high-affordance degree simply as *manipulable* objects operated exclusively with the hands, whose operation may induce a change in the physical world. For instance, the use of a pair of scissors typically results into the division of a sheet of paper into smaller units. Conversely, *non-manipulable* objects could not be directly operated using the hands, and we minimised the possibility for any object in our dataset to induce motor simulation. For instance, if an object might contain handles or knobs, we either chose a visual instance of the object without such features, or the features were digitally erased from the picture.

To validate the stimuli, we conducted a pre-test where we asked questions about the objects to 25 Dutch-speaking naïve judges uninvolved in the actual experiment, using Crowdflower (an online crowdsourcing service; http://www.crowdflower.com/). In this questionnaire, participants were asked to name each object in Dutch (we later computed whether the name was correct, and assigned it either a 0—incorrect or 1—correct),

and also rated the manipulability, and degree of perceived visual complexity of each object on a scale from 0 to 100 (being 0 the least manipulable/complex and 100 the most). Our aim was to make sure that participants could name the objects correctly in Dutch, and that these objects were rated similarly in visual complexity, to ensure that the speakers' gesturing rate would not be affected by anything other than our affordance manipulation.

The percentage of correctly named objects ranged between 90% and 100% for the selected items (M_{HIGH} = 94.35, SD = 2.24, M_{LOW} = 93.14, SD = 2.14), and fell below 35% for their perceived visual complexity (M_{HIGH} = 29.01, SD = 2.46, M_{LOW} = 26.74, SD = 2.39). Most importantly, the scores did not differ between the high- and low-affordance items for complexity ($t(24) = 1.51$, $p = .14$). The manipulability ratings for both affordance groups were statistically significant, as intended (M_{HIGH} = 74.47, SD = 11.96, M_{LOW} = 41.4, SD = 21.42) ($t(24) = 9.53$, $p < .001$).

Procedure

The experiment introduced participants to a fictive scenario in which participant A (the speaker) was relocating, but due to an injury could not go by himself to the department store to buy utensils and furniture. Participant B (the listener) would go in his place, but for this to be possible they would have to agree beforehand on the items to be purchased. Thus, the speaker's task was to briefly describe each of the items, in such a way that the listener would be able to visually identify them. The stimuli that the speaker would describe were displayed on a 13 in. laptop screen, placed on a table to the left side of the speaker. All picture items were compiled into a presentation document, where high- and low-affordance objects were mixed at random. Each object fully occupied the screen. Each object was preceded by a slide indicating the trial number (see Figure 1), to ease the coordination between the speaker and the listener's tasks. The listener was given a paper brochure, in which pictures of all objects appeared forming a grid, each item accompanied by a letter. Next to it, the listener was given an answer sheet with two columns: one with the trial numbers, and the other with blanks to fill in the letters corresponding to the items described. Thus, the listener's task was to identify each object in the brochure she was given, and annotate the letter corresponding to such object on her answer sheet.

Each pair received written instructions, and had the chance to do a practice round before the actual experiment began, with an item that was not part of the
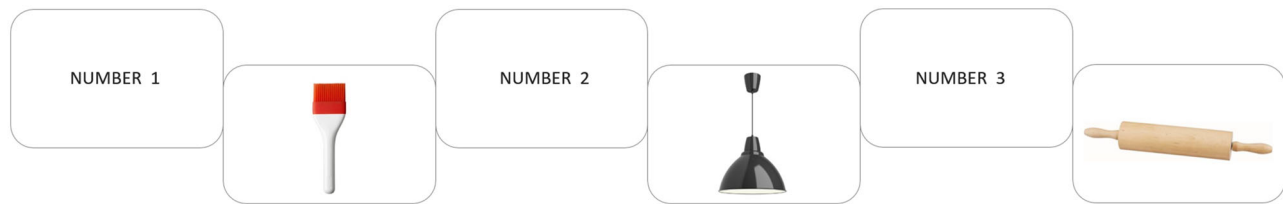
Figure 1. Example of the stimuli presentation as seen by the speaker. Each object is embedded in one slide, occupying it fully, always preceded by a slide presenting the item number.

stimuli set. Speakers and listeners were allowed to speak freely, and had no restrictions with respect to the way they designed their descriptions—for example, naming the objects was not prohibited. A digital video camera was placed behind the listener, to record the speaker's speech and gestures.

Data analyses

We transcribed all words produced by the speakers (until the listener would write down her response) and annotated all gestures, using the multimodal annotation tool Elan (Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands, http://www.lat-mpi.eu/tools/elan; Wittenburg, Brugman, Russel, Klassmann, & Sloetjes, 2006). We categorised gestures as representational and non-representational gestures. Representational gestures were defined as hand movements depicting information related to the semantic content of the ongoing speech. Examples of such gestures are tracing the contour of a house with the index finger, or repeatedly pushing down the air with the palm, simulating the bouncing of a basketball. The non-representational gestures mainly comprised rhythmic gestures used to emphasise words (beats—McNeill, 1992), and interactive or pragmatic gestures directed at the addressee (Bavelas, Chovil, Lawrie, & Wade, 1992; Kendon, 2004). We excluded from our annotation other non-verbal behaviours such as self-adaptors (e.g. fixing one's hair). Each gesture was annotated in its full length, from the preparation to the retraction phase (see McNeill, 1992). When a gesture stroke was immediately followed by a new gesture, we examined the fragment frame-by-frame, and set the partition at the exact moment where a change in hand shape, or movement type would take place.

Next, we annotated the techniques observed in the speakers' gestures. Representation technique was coded only for representational gestures, assigning always one technique to each gesture. We took as our point of departure Müller's four representation modes —imitating, drawing, portraying, and moulding (Müller, 1998), and expanded the list, further sub-categorising

some representation modes, based on the gestures we observed in our dataset after screening the first five videos, and adding an extra category: placing (see, e.g. Bergmann & Kopp, 2009). A detailed overview of the techniques annotated can be found in Appendix 2. While it is true that some representation modes are often associated to specific handshapes (for example, moulding is oftentimes associated with flat handshapes, and tracing is often performed with a single stretched finger), our main criterion in coding these representation modes was to ask "how the hands are used symbolically" (Müller, 1998, p. 323).

To validate the reliability of the annotations, a second coder, naïve to the experimental conditions and hypotheses, performed gesture identification in 40 descriptions (produced by 8 different speakers), and judged the representation technique used in a sample of 60 gestures produced by 12 different speakers (5 gestures per speaker). In total, 146 gestures from 20 different speakers (9.8% of all annotated gestures) were analysed by the second coder. Cohen's κ reveals substantial agreement with respect to the number of gestures produced by speakers ($\kappa = .71$, $p < .001$), and an almost perfect agreement with respect to the representation techniques ($\kappa = .84$ $p < .001$).

Design and statistical analyses

The effects of affordance on our dependent variables were assessed using linear mixed models for continuous variables (i.e. gesture rates), and logit mixed models for categorical variables (i.e. representation techniques) (see Jaeger, 2008). Mixed-effect models allow us to account for fixed as well as random effects in our data simultaneously, thereby optimising the generalisability of our results and eliminating the need to conduct separate F1 and F2 analyses. Thus, "affordance" (two levels: high and low) was the fixed factor in all of our analyses, and participants and items were included as random factors. In all cases, we started with a full random effects model (following the recommendation by Barr, Levy, Scheepers, & Tily, 2013). In case the model did not converge, we eliminated the random slopes with

the lowest variance. *P* values were estimated using the Likelihood Ratio Test, contrasting, for each dependent variable, the fit of our (alternative) model with the fit of the null model.

Results

The communication task elicited 1120 descriptions, 509 of which were accompanied by at least one gesture. A total of 1483 gestures were identified. Representational gestures accounted for 72% (1070) of the gestures annotated, the remaining 28% (413) consisting of non-representational gestures. Our first research question was concerned with whether perceiving objects that afford manual actions would result into the production of more gestures. We computed a *normalised* gesture rate measure, whereby the number of gestures produced per description is calculated relative to the number of words spoken (Gestures/Words × 100). Trials where no gestures were produced (null trials) were excluded, but in such a way that the ratio of descriptions for high and low manipulability objects for each speaker was preserved. Thus, we only excluded null trials for one condition if the same number of null trials could be excluded for the other condition, leading to the examination of gestures in 572 descriptions (286 per condition). We did this in order to reduce the variance in our dataset caused by the amount of 0-gesture trials, without either losing data or compromising our results. We computed the gesture rate two times, first for representational gestures and second for non-representational gestures. The results show that *affordance* influenced the representational gesture rate, which was higher for high-affordance objects ($M_{HIGH} = 9.76$, SD = 12.53) than for low-affordance objects ($M_{LOW} = 6.47$, SD = 7.82) ($\beta = -2.91$, SE = 1.17, $p = .004$). However, we found no effects of *affordance* on the non-representational gesture rate,

which did not differ between manipulable ($M_{HIGH} = 3.83$, SD = 6.33) and non-manipulable objects ($M_{LOW} = 3.71$, SD = 6.01) ($\beta = -.15$, SE = 0.74, $p = .72$).

Given that gesture rate is also dependent on the number of words produced by a speaker, it could be the case that the number of words is also sensitive to *affordance*, which could in turn have influenced gesture rate. Hence, we computed the effects of *affordance* on the number of words uttered by speakers, and found no statistically supported differences between manipulable ($M_{HIGH} = 23.29$, SD = 14.85) and non-manipulable objects ($M_{LOW} = 24.41.4$, SD = 15.2) ($\beta = .58$, SE = 2.78, $p = .1$).

In summary, our results suggest that speakers do gesture more when faced with an object that they can manipulate with their hands, but this effect is restricted to the production of representational gestures (Figure 2).

Analysis of representation techniques

Our results support the prediction that describing objects that afford manual action would elicit more gestures where the speaker pretended to execute the action associated to the object ($\beta = -4.46$, SE = 0.93, $p < .001$) ($M_{HIGH} = .39$, SD = .48; $M_{LOW} = .02$, SD = .15), or pretended to handle (grip) such object ($\beta = -3.34$, SE = 1.17, $p < .001$) ($M_{HIGH} = .16$, SD = .37; $M_{LOW} = .02$, SD = .14). In contrast, for objects in the low-affordance condition, speakers typically made use of moulding gestures in which the hands sculpted their shape ($\beta = 1.76$, SE = 0.34, $p < .001$) ($M_{HIGH} = .28$, SD = .45; $M_{LOW} = .66$, SD = .47), and of placing gestures where the hands expressed the spatial relation between different features of an object ($\beta = 3.007$, SE = 1.04, $p < .001$) ($M_{HIGH} = .02$, SD = .14; $M_{LOW} = .13$, SD = .33) (see Figure 3).

Discussion

The experiment reported in this paper was designed to examine the impact of a core property of objects, namely their degree of *action affordance*, on the production of co-speech gestures. Particularly, we sought to elucidate (a) whether perceiving objects that afford manual actions (without attending to explicit action demonstrations) sufficed to increase the production of (representational) gestures and (b) whether the action component intrinsic to these objects would be reflected in the representation techniques used to gesture.

Our analyses yielded a number of noteworthy results. First, our results suggest that merely describing objects with manual affordances (e.g. tools), as opposed to objects whose daily function is not primarily executed
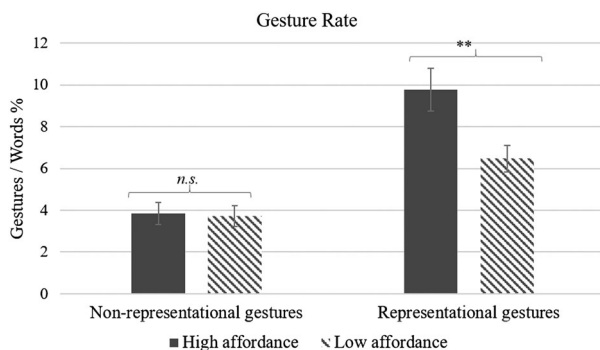

Figure 2. Gesture rates for non-representational gestures (left) and representational gestures (right). The bars represent the mean number of gestures per 100 words, and the error bars represent the 95% confidence intervals. **Significant at $p < .005$.
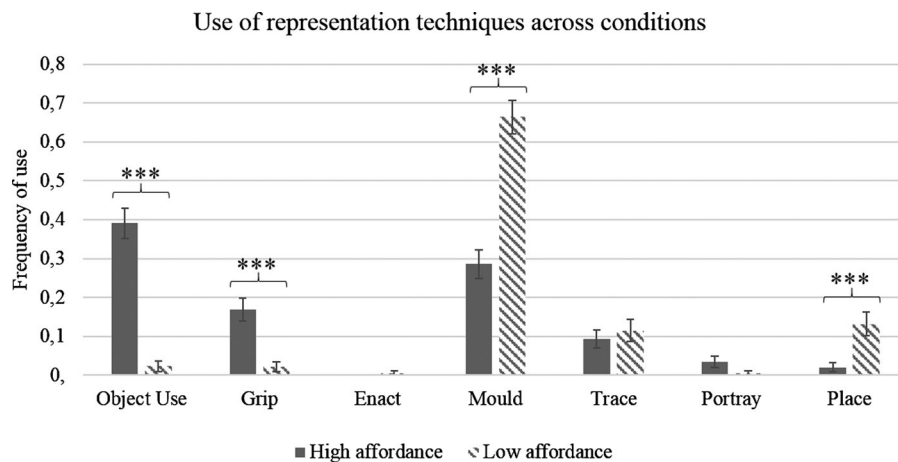
Figure 3. Frequency of use of each representation technique (annotated only for representational gestures). The error bars represent the 95% confidence intervals. ***Significant at $p < .001$.

with the hand, is indeed enough to elevate the rate of co-speech gestures produced by speakers. This result, however, was only found for representational gestures. This is consistent both with previous research (Pine et al., 2010), and with the *GSA* (Hostetter & Alibali, 2008), which specifically predicts more representational gestures accompanying stronger simulations of motor imagery. Currently, the GSA framework accounts solely for the production of representational gestures. Our study contributes to a possible instantiation of such framework, by showing that this effect does not extend to the production of non-imagistic gestures such as *pragmatic* gestures directed to the addressee (Kendon, 2004) and *beats* (McNeill, 1992)—which constituted most of the gestures annotated as non-representational in our study. The relationship between representational and non-representational gestures has been largely ignored in currently available gesture models, in terms of the mechanisms underlying the production of both gesture types, with nearly all accounts limiting their scope to representational gesture. This fact suggests that, although produced together in talk, both types of gestures may have their origin in different cognitive processes (Chu & Kita, 2015) and relate to imagistic and linguistic content in different ways. Our results emphasise this difference by showing that the activation caused by our stimuli was restricted to representational gestures, thereby suggesting that the response to the perception of affordances does not generate simple movement activation (going against what we earlier termed a "general activation" account), but that it seems to recruit motor responses that are specific to the features of the represented referents. The extent to which the production of affordance-congruent gestures is semantically mediated, or whether these gestures emerge from a more "direct" visual route to action is a question that requires further investigation.

Despite our finding that more gestures were produced while describing high-affordance objects, still a high amount of gestures were produced while describing low-affordance items. We hypothesise that objects in the low-affordance category may have evoked action simulations as well, but of a different kind. For instance, many of these objects had large flat surfaces, which may have activated haptic ("touching") simulations in the speaker (e.g. a ball-shaped lamp affords to be palpated and its structure affords to be moulded with both hands; a flat surface affords running our palms over it, etc.). This explanation is supported by the predominant use of moulding gestures (mainly associated with flat handshapes) in the description of low-affordance objects. In addition, we observed a tendency in speakers to represent the objects in the low-affordance condition following a piecemeal strategy. That is, whereas for high-affordance objects speakers could mime the performance of an action in one gesture, for low-affordance objects speakers tended to represent separately, in sequential gestures, the shape of different salient features of the object. For instance, it was common that a speaker would describe a shelves rack by first moulding its overall shape, then moulding the shape of one shelf (showing its horizontality, flatness, and size) and then producing several placing gestures, indicating the location of the remaining individual shelves with respect to one another. Such detailed descriptions occurred very often in our dataset, and they may partly be due to the fact that our speakers had to describe pictures of objects, rich in visual detail, and not verbal items (as in Hostetter, 2014). It is therefore likely that speakers will produce even less gestures

accompanying the descriptions of non-manipulable objects when the targets are not presented visually. Further studies comparing the production of gestures in response to both types of stimuli presentation (written versus pictorial) should clarify this issue.

Representation modes in gestural depiction

While the use of different techniques to gesturally represent concepts has been described in the literature (see, e.g. Kendon, 2004; Müller, 1998; Streeck, 2008, 2009; van Nispen et al., 2014), it has received little scholarship thus far. If gestures stem from the imagery that underlies thought, we can conceive representational hand gestures as visible materialisations of certain aspects of a speaker's mental pictures. In other words, there is a degree of isomorphism between mental representations and hand gestures, and therefore it is worthwhile investigating the iconic strategies that allow for the "transduction" of imagery into movement. In this study, we originally looked at four representation modes that have their root on daily activity as well as on artistic expression modes: imitating, moulding, tracing, and portraying (Müller, 1998). Our results show that objects that afforded manual actions were mostly represented through imitating gestures (particularly, object use and gripping gestures), whereas low-affordance objects were mostly represented with moulding and placing gestures. The remaining categories did not reveal significant differences (e.g. tracing), mostly because of the low frequency with which they occurred (e.g. enacting, portraying).

In summary, it is likely that high-affordance objects evoked simulations of action in the speakers, and that this was manifested not only in the amount of gesturing, but also in the features of the referents that these gestures represented. The fact that most high-affordance objects were represented through imitating gestures (object use, grip) supports the notion that viewing objects triggers the activation of the motor processes associated with physically grasping those objects (e.g. Bub et al., 2003; Ellis & Tucker, 2000). Conversely, it is likely that the moulding gestures accompanying low-affordance objects stemmed from simulations of touching, having their root on everyday exploring through *haptic perception* (Lederman & Klatzky, 1987). We have hypothesised about the connection between imitating and moulding gestures, and different types of simulated action. Nevertheless, we wonder about the cognitive origin of other representation modes, such as tracing, or portraying. One noteworthy aspect of the gestural techniques we analysed is that they display different degrees of abstraction or schematicity (see, e.g. Perniss

& Vigliocco, 2014). For instance, miming the performance of an object is close to daily sensorimotor experience, and seems relatively "unfiltered" in comparison with drawing a contour, which implies the abstraction of a series of features into a shape, ultimately traced by the finger. It becomes apparent that these gestures also vary in terms of their cognitive complexity (e.g. Bartolo, Cubelli, Della Sala, & Drei, 2003), and it is therefore likely that different gestural techniques originate in different processes. Thus, future research should address how cognitive and communicative aspects constrain the use of representation techniques, which will, in our opinion, inform greatly the creation of more comprehensive co-speech gesture models.

In conclusion, this study showed that (action) affordances influence gestural behaviour, by determining both the amount of representational gestures produced by speakers, and the gestural techniques chosen to depict such objects. The present findings thus support and expand the assumptions of the GSA framework (Hostetter & Alibali, 2008) and are compatible with previous research in the field of motor cognition showing specific handshape-affordance congruency effects during visual and language tasks (e.g. Bub & Masson, 2006; Bub et al., 2003; Tucker & Ellis, 1998). In addition, to our knowledge, this is the first study to have systematically shown a connection between object properties and gestural representation techniques during referential communication. The insight gained by looking at such techniques highlights the importance of adopting a more qualitative approach to gesture research, as means to comprehend in depth the processes that give rise to gesture production.

References

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*, 255–278.

Bartolo, A., Cubelli, R., Della Sala, S., & Drei, S. (2003). Pantomimes are special gestures which rely on working memory. *Brain & Cognition*, *53*, 483–494.

Bavelas, J., & Healing, S. (2013). Reconciling the effects of mutual visibility on gesturing: A review. *Gesture*, *13*(1), 63–92. doi:10.1075/gest.13.1.03bav

Bavelas, J. B., Chovil, N., Lawrie, D. A., & Wade, A. (1992). Interactive gestures. *Discourse Processes*, *15*, 469–489.

Bavelas, J. B., Gerwing, J., Sutton, C., & Prevost, D. (2008). Gesturing on the telephone: Independent effects of dialogue and visibility. *Journal of Memory and Language*, *58*, 495–520.

Beattie, G., & Shovelton, H. (2002). What properties of talk are associated with the generation of spontaneous iconic hand gestures?. *British Journal of Psychology*, *41*, 403–417.

Bergman, K., & Kopp, S. (2009). Increasing expressiveness for virtual agents: Autonomous generation of speech and gesture. In K. Decker, J. Sichman, C. Sierra, & C. Castelfranchi (Eds.), *Proceedings of the 8th international conference on autonomous agents and multiagent systems* (pp. 361–368). Ann Arbor, MI: IFAAMAS.

Bub, D. N., & Masson, M. E. J. (2006). Gestural knowledge evoked by objects as part of conceptual representations. *Aphasiology*, *20*, 1112–1124.

Bub, D. N., Masson, M. E. J., & Bukach, C. M. (2003). Gesturing and naming: The use of functional knowledge in object identification. *Psychological Science*, *14*, 467–472.

Bub, D. N., Masson, M. E. J., & Cree, G. S. (2008). Evocation of functional and volumetric gestural knowledge by objects and words. *Cognition*, *106*, 27–58.

Chu, M., & Kita, S. (2008). Spontaneous gestures during mental rotation tasks: Insights into the microdevelopment of the motor strategy. *Journal of Experimental Psychology: General*, *137*, 706–723.

Chu, M., & Kita, S. (2015). Co-thought and Co-speech gestures are generated by the same action generation process. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. Advance online publication. doi:10.1037/xlm0000168

Cook, S. W., & Tanenhaus, M. K. (2009). Embodied communication: Speakers' gestures affect listeners' actions. *Cognition*, *113*, 98–104.

Ellis, R., & Tucker, M. (2000). Micro-affordance: The potentiation of components of action by seen objects. *British Journal of Psychology*, *91*(4), 451–471. doi:10.1348/000712600161934

Feyereisen, P., & Havard, I. (1999). Mental imagery and production of hand gestures while speaking in younger and older adults. *Journal of Nonverbal Behavior*, *23*, 153–171.

Fischer, M., & Zwaan, R. (2008). Embodied language: A review of the role of the motor system in language comprehension. *The Quarterly Journal of Experimental Psychology*, *61*(6), 825–850. doi:10.1080/17470210701623605

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

Gerlach, C., Law, I., & Paulson, O. B. (2002). When action turns into words: Activation of motor-based knowledge during categorization of manipulable objects. *Journal of Cognitive Neuroscience*, *14*, 1230–1239.

Gerwing, J., & Bavelas, J. (2004). Linguistic influences on gesture's form. *Gesture*, *4*, 157–195.

Gibson, J. J. (1986). *The ecological approach to visual perception*. New York: Psychology Press. doi:10.4324/9780203 767764

Glenberg, A. M., & Kaschak, M. P. (2002). Grounding language in action. *Psychonomic Bulletin & Review*, *9*, 558–565.

Glenberg, A. M., & Robertson, D. A. (2000). Symbol grounding and meaning: A comparison of high-dimensional and embodied theories of meaning. *Journal of Memory and Language*, *43*(3), 379–401. doi:10.1006/jmla.2000.2714

Glover, S., Rosenbaum, D. A., Graham, J., & Dixon, P. (2004). Grasping the meaning of words. *Experimental Brain Research*, *154*, 103–108.

Goldin-Meadow, S. (2003). *Hearing gesture: How our hands help us think*. Cambridge, MA: Harvard University Press.

Hadar, U., & Butterworth, B. (1997). Iconic gestures, imagery and word retrieval in speech. *Semiotica*, *115*, 147–72.

Handy, T. C., Grafton, S. T., Shroff, N. M., Ketay, S., & Gazzaniga, M. S. (2003). Graspable object grab attention when the potential for action is recognised. *Nature Neuroscience*, *6*, 421–427.

Hauk, O., Johnsrude, I., & Pulvermüller, F. (2004). Somatotopic representation of action words in human motor and pre-motor cortex. *Neuron*, *41*, 301–307.

Hoetjes, M., Koolen, R., Goudbeek, M., Krahmer, E., & Swerts, M. (2015). Reduction in gesture during the production of repeated references. *Journal of Memory and Language*, *79–80*, 1–17.

Hostetter, A. B. (2014). Action attenuates the effect of visibility on gesture rates. *Cognitive Science*, *38*(7), 1468–1481. doi:10.1111/cogs.12113

Hostetter, A. B., & Alibali, M. W. (2008). Visible embodiment: Gestures as simulated action. *Psychonomic Bulletin and Review*, *15*, 495–514.

Hostetter, A. B., & Alibali, M. W. (2010). Language, gesture, action! A test of the gesture as simulated action framework. *Journal of Memory and Language*, *63*, 245–257.

Hostetter, A. B., Alibali, M. W., & Bartholomew, A. E. (2011). Gesture during mental rotation. In L. Carlson, C. Hoelscher, & T. Shipley (Eds.), *Proceedings of the 33rd annual meeting of the cognitive science society* (pp. 1448–1454). Austin, TX: Cognitive Science Society.

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, *59*(4), 434–446. doi:10.1016/j.jml.2007.11.007

Kendon, A. (2004). *Gesture. Visible action as utterance*. Cambridge: Cambridge University Press. doi:10.1017/cbo9780511807572

Krauss, R. M., Chen, Y., & Gottesman, R. F. (2000). Lexical gestures and lexical access: A process model. In D. McNeill (Ed.), *Language and gesture* (pp. 261–283). New York, NY: Cambridge University Press. doi:10.1017/cbo9780511620850.017

Lederman, S. J., & Klatzky, R. L. (1987). Hand movements: A window into haptic object recognition. *Cognitive Psychology*, *19*, 342–368.

Louwerse, M. M. (2011). Symbol interdependency in symbolic and embodied cognition. *Topics in Cognitive Science*, *3*, 273–302.

Louwerse, M. M., & Jeuniaux, P. (2010). The linguistic and embodied nature of conceptual processing. *Cognition*, *114*, 96–104.

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

Müller, C. (1998). Iconicity and gesture. In S. Santi, I. Guatiella, C. Cave, & G. Konopczyncki (Eds.), *Oralité et Gestualité* (pp. 321–328). Montreal: L'Harmattan.

Parrill, F. (2010). Viewpoint in speech-gesture integration: Linguistic structure, discourse structure, and event structure. *Language and Cognitive Processes*, *25*(5), 650–668. doi:10.1080/01690960903424248

Perniss, P., & Vigliocco, G. (2014). The bridge of iconicity: From a world of experience to the experience of language. *Philosophical Transactions of the Royal Society*, *369*(1651), 1–13. doi:10.1098/rstb.2014.0179

Pine, K., Gurney, D., & Fletcher, B. (2010). The semantic specificity hypothesis: When gestures do not depend upon the presence of a listener. *Journal of Nonverbal Behavior*, *34*(3), 169–178. doi:10.1007/s10919-010-0089-7

Streeck, J. (2008). Depicting by gesture. *Gesture*, *8*(3), 285–301. doi:10.1075/gest.8.3.02str

Streeck, J. (2009). *Gesturecraft: The manu-facture of meaning*. Amsterdam: John Benjamins Publishing. doi:10.1017/s0047404513000262

Tettamanti, M., Buccino, G., Saccuman, M. C., Gallese, V., Danna, M., Scifo, P., … Perani, D. (2005). Listening to action-related sentences activates fronto-parietal motor circuits. *Journal of Cognitive Neuroscience*, *17*, 273–281.

Tucker, M., & Ellis, R. (1998). On the relations between seen objects and components of potential actions. *Journal of Experimental Psychology: Human Perception and Performance*, *24*(3), 830. doi:10.1037//0096-1523.24.3.830

Van Nispen, K., van de Sandt-Koenderman, M., Mol, L., & Krahmer, E. (2014). Pantomime Strategies: On regularities in how people translate mental representations into the gesture modality. In P. Bello, M. Guarini, M. McShane, & B. Scassellati (Eds.), *Proceedings of the 36th annual conference of the cognitive science society* (pp. 976–981). Austin, TX: Cognitive Science Society.

Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., & Sloetjes, H. (2006). ELAN: a professional framework for multimodality research. In *Proceedings of LREC, 5th international conference on language resources and evaluation*. Paris: ELRA.

Appendix 1. List of target items (note: in the experiment, items were presented visually).

Manipulable objects	Non-manipulable objects
Pastry brush	Stepladder
Spatula	Plant
Knife	Dining table
Grater	Flatware tray
Whisk	Ball lamp
Hammer	Wall shelf
Garlic press	Cart
Rolling pin	Hood
Cook timer	Sink
Egg slicer	Kitchen island
Wine glass	Desk
Cheese slicer	Clock
Pitcher	Lamp
French press	Stool

Appendix 2. Description and examples of the representation techniques annotated in the present study.

Representation mode	Description
Object use	Represents a transitive action, whereby the actor simulates the performance of an object-directed action. Example: the hand acts as if holding a pen, with both thumb and index fingertips pressed together, imitating the act of writing.
Enactment	Represents an intransitive action, whereby the actor simulates the performance of a non-object-directed action. Example: the arms swing back and forth in alternated movements, simulating the motion of the upper body while running.
Hand grip	The hand acts as if it were grasping or holding an object, without carrying out any specific action. Example: fingers close into a clenched fist, as if holding the handle of a tool.
Moulding	The hand acts as if it were palpating, or sculpting the surface of an object. Example: a flat hand with the palm facing down moves along the horizontal axis, representing the "flatness" of an object's surface.
Tracing	The hand (typically using the index finger) draws a shape in the air, or traces the trajectory (to be) followed by an entity. Example: tracing a big square with the tip of the finger, representing a quadratic object such as a window.
Portraying	The hand is used to portray an object (or character) in a holistic manner, as if it had become the object itself. Example: with two fingers (index and middle) stretched out horizontally, and the others closed, the hand can portray a pair of scissors, and simulate the action of cutting through paper.
Placing	The hand anchors or places an entity within the gesture space, or explicitly expresses a spatial relation between two or more entities. Example: when describing a scene, a speaker might use his hand to indicate the location of the actors and objects portrayed.