Journal of Memory and Language 66 (2012) 249–264

Contents lists available at ScienceDirect

Journal of Memory and Language

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

Adaptation in gesture: Converging hands or converging minds?

Lisette Mol *, Emiel Krahmer, Alfons Maes, Marc Swerts

Tilburg center for Cognition and Communication (TiCC), School of Humanities, Tilburg University, P.O. Box 90135, NL-5000 LE Tilburg, The Netherlands

ARTICLE INFO

Article history:
Received 30 August 2010
revision received 12 July 2011
Available online 10 August 2011

Keywords:
Gesture
Adaptation
Mimicry
Alignment
Lexical entrainment

ABSTRACT

Interlocutors sometimes repeat each other's co-speech hand gestures. In three experiments, we investigate to what extent the copying of such gestures' form is tied to their meaning in the linguistic context, as well as to interlocutors' representations of this meaning at the conceptual level. We found that gestures were repeated only if they could be interpreted within the meaningful context provided by speech. We also found evidence that the copying of gesture forms is mediated by representations of meaning. That is, representations of meaning are also converging across interlocutors rather than just representations of gesture form. We conclude that the repetition across interlocutors of representational hand gestures may be driven by representations at the conceptual level, as has also been proposed for the repetition of referring expressions across interlocutors (lexical entrainment). That is, adaptation in gesture resembles adaptation in speech, rather than it being an instance of automated motor-mimicry.

Introduction

Suppose Mary and John are discussing a route through a city. Then if Mary refers to an alley as 'the narrow street', it is likely that John will also use this expression when subsequently referring to that alley, rather than using a completely different expression, such as 'the alleyway'. In addition to saying 'the narrow street', Mary may hold up her hand in front of her, with her fingers pointing towards the end of the alley, thereby indicating the direction that the alley runs in. This hand gesture may subsequently be repeated by John, when he talks about the alley again. Would such a repetition of a gesture be similar to a repetition of a referring expression?

When people interact in dialogue, they adapt to each other in many ways (for a recent overview, see Branigan, Pickering, Pearson, & McLean, 2010). Brennan and Clark (1996) showed that interlocutors tend to repeat each other's referring expressions, a process known as *lexical entrainment*. Apart from verbal adaptation, interlocutors can also repeat each other's non-verbal behaviors (e.g. Chartrand & Bargh, 1999), among which are the hand gestures that many people produce spontaneously while talking. It has been found that people indeed repeat such hand gestures of each other (De Fornel, 1992; Holler & Wilkin, 2011; Kimbara, 2008; Tabensky, 2001). Yet how similar are these repetitions in gesture to repetitions in speech? Do similar processes underlie adaptation in both speech and gesture? And what is the role of meaning? Do interlocutors produce similar hand gestures because they construct similar representations of meaning, or are they merely copying each other's movements?

Mimicry and adaptation

Mimicry and adaptation in interaction have been studied extensively. Chartrand and Bargh (1999) for instance, found that participants were more likely to shake their foot during a conversation if their confederate conversation partner did so as well, and similarly for rubbing one's face with one's hand. According to Chartrand and Bargh (1999), although such mimicry may act as a kind of 'social glue', intent or conscious effort are not required for it to occur. They state that "the mere perception of another's behavior automatically increases the likelihood of engaging in that

* Corresponding author. Fax: +31 13 4663110.
E-mail address: L.Mol@uvt.nl (L. Mol).

0749-596X/$ - see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.jml.2011.07.004

behavior oneself'', p. 893. This is known as the *perception–behavior link.*

Pickering and Garrod (2004) propose that similar automated priming may frequently underlie the repetition of linguistic forms across interlocutors, which they call *alignment*. They propose that at each linguistic level, ''[t]he activation of a representation in one interlocutor leads to the activation of the matching representation in the other interlocutor directly'', p. 177. For example, after hearing a sentence in the passive voice, people are more likely to produce a passive voice as well (Bock, 1986). In Pickering and Garrod's *interactive alignment account*, this is explained as the speaker's representation of the passive voice being more activated as a result of perception, and therefore being a more likely candidate for subsequent production. Thus, it is assumed that representations are shared between comprehension and production (parity of representation).

In the interactive alignment account, the repetition across interlocutors of a linguistic form at any one level (e.g. the syntactic level) can happen without representations at other linguistic levels playing a critical role. For example, repetition of the same syntactic structure could happen when a similar, but also when a very different lexical form or meaning is being expressed than was perceived. Although stronger effects have repeatedly been found when the same word was repeated, syntactic alignment indeed also occurred when different words (with different meanings) were produced than were perceived (Cleland & Pickering, 2003; Pickering & Branigan, 1998).

It thus seems that adaptation of syntactic structures can occur independently of representations at the lexical and the conceptual level. Yet it is sometimes hard to see how adaptation at the lexical level can occur without the conceptual level being involved. It is rare that speakers choose the same referring expression as their interlocutors, whereas they do not want to express the same meaning. Therefore, when looking at the repetition of forms that carry propositional meaning, we need to describe how meaning is involved in the repetition of form. In Pickering and Garrod's view, alignment of form tends to be linked to alignment of meaning, but it does not have to be. One way their model could account for such a link is that it also assumes connections between the representations at different linguistic levels within a speaker. This means that activation of a representation at the lexical level could lead to activation of a representation at the conceptual level, and vice versa. This way, when during perception the representation for the word form 'alley' is associated with a certain representation of meaning, the connection between those two representations is strengthened. When the same representation of meaning is subsequently activated in the production process, the representation of the word form for 'alley' receives activation through this connection, making the word 'alley' a more likely candidate for production. Although the model can thus account for the link between alignment of meaning and form, its unique contribution lies in that it is also possible for alignment of form to occur without the alignment of meaning, as alignment at each linguistic level can happen independently of other levels.

Brennan and Clark (1996) do propose that when interlocutors use the same words to refer to the same objects, this is because they use similar conceptualizations of those objects. For example, suppose a particular object could be thought of as a document, a picture, or a map. When a speaker refers to it as a 'map', she conceptualizes the object for the current purpose as such. If the addressee agrees with this conceptualization and a *conceptual pact* is formed, both interlocutors can subsequently use the utterance 'map' as a reference to both the object and the particular conceptualization of it. Thus, for both interlocutors a certain utterance is linked to a certain object, as well as to a certain representation of meaning at the conceptual level. In this view, representations of meaning are necessarily involved in the repetition of referring expressions across interlocutors, and for interlocutors to use similar expressions, they need to have similar representations at the conceptual level. Before exploring what answers current models on gesture production and earlier studies on adaptation in gesture can provide to the question of whether or not the conceptual level is necessarily involved in the repetition of hand gesture forms across interlocutors, we first specify what we mean by gestures.

Co-speech gestures

When talking, many people move their hands and arms around without the objective of directly manipulating their environment. Rather, such movements seem to be part of their communicative effort (Kendon, 2004). For example, raising one's arm while extending one's index finger toward an object could disambiguate the question ''Can you hand me that?''. Kendon (1988) recognizes different kinds of hand gestures, which McNeill (1992) put on a continuum of how conventional and language-like the hand gestures are. On one end of this continuum are sign languages, in which signs have a conventional meaning. At the far other end is *gesticulation.* This is the production along with speech of gestures that are not embedded in the structure of speech. For example, moving the hands along the upper body as though running while saying: ''He went away''. Gestures at this end of the continuum are the most idiosyncratic and their interpretation is highly dependent on the accompanying speech (Feyereisen, Van de Wiele, & Dubois, 1988).

Gestures that fall into Kendon's category of gesticulation are generally divided into several subcategories (e.g. McNeill, 1992). One broad distinction can be made based on whether a gesture depicts some of the content of the message a speaker is trying to convey, or whether it rather structures the conversation (e.g. Bavelas, Chovil, Lawrie, & Wade, 1992), or emphasizes certain parts of speech (e.g. Effron, 1941; Ekman & Friesen, 1969; Krahmer & Swerts, 2007). In this paper we focus on gestures that express some of the content a speaker is conveying, which are known as *illustrators* (Ekman & Friesen, 1969), or *representational gestures* (McNeill, 1992). Elements of such a gesture's physical form, like the shape and orientation of the hand, the direction and size of the movement, and where it is performed relative to the speaker (Müller, 1998), can be repeated in subsequent gestures by the same or another speaker. Yet importantly, since these gestures are among the least conventional on Kendon's continuum, there are many different ways in which the same content could potentially be expressed in co-speech gesturing.

Gesture and speech production and perception

Gesture and speech have been found to be linked temporally (Chui, 2005), structurally (Kita et al., 2007), pragmatically (Enfield, Kita, & De Ruiter, 2007), and semantically (McNeill, 1992). Therefore, gesture and speech production are somehow coordinated. McNeill's Growth Point theory states that speech and gesture co-express idea units, which develop themselves into utterances (McNeill, 1992, 2005). That is, speech and gesture co-develop over time, into an utterance. Therefore, it is no surprise that current models of gesture production are based on a model of speech production, specifically the framework of Levelt (1989).

Levelt's *blueprint for the speaker* discriminates between a conceptualization, a formulation, and an articulation stage. Based on this model, De Ruiter (1998, 2000) proposes that a communicative intention is formed in the conceptualization stage, which is then passed onto two parallel formulation stages: one for gesture and one for speech. Each of these formulation stages leads to its own articulation stage, either the articulation of gesture or the articulation of speech. This model is called the *postcard model*, because rather than assuming that gesture directly reflects thought, it assumes that a communicative intention underlies gesture production. Therefore, the metaphor of postcards from the mind may be more accurate than gesture being a window into the mind (McNeill, 1992). In the postcard model, gesture and speech production share the stages up until the formulation of a communicative intention, and then continue in parallel, but separately.

This is different from the *interface model* proposed by Kita and Özyürek (2003). In this model, the processes of gesture and speech formulation interact with each other online. The message to be communicated is not fully determined by one conceptualization module, but also by two separate generators: the action generator for gesture and the message generator for speech. The action generator has access to spatial and motoric components in working memory as well as to a model of the environment, while the message generator has access to propositional components in working memory and the discourse model. Importantly, there are bidirectional links between the action and message generator, as well as between the message generator and the speech formulator. This means that constraints on how a message can be expressed in speech, can also influence how it is expressed in gesture. Thus, the content of gesture is not fully specified by the communicative intention alone, but also by the features of imagined or real space, and online feedback from the speech formulator via the message generator.

Both the postcard model and the interface model focus on the production of speech and gesture, and thus model the speaker. It is not specified what representations are shared between production and comprehension or where the links between these processes are. When it comes to adaptation in gesture, both gesture production and perception are involved. Levelt assumes that lemmas and forms are shared between speech formulation and speech comprehension within a speaker. This assumption is also incorporated into the interactive alignment account proposed by Pickering and Garrod (2004). This model includes multiple interlocutors, allowing it to explain adaptation of one interlocutor to another. It is however non-specific about gesture.

We can apply the interactive alignment account to adaptation in gesture in two ways. Firstly, if we assume that gesture forms are represented at their own linguistic level, that they are shared between perception and production, and that interlocutors can align to each other's forms directly, adaptation in gesture could occur without the conceptual level playing a critical role. This would be a low-level account of adaptation in gesture. Secondly, if we assume that representations at the conceptual level are shared between speech and gesture comprehension and production, and that there are bidirectional links between the different linguistic levels within a speaker, then adaptation in gesture could also occur via the conceptual level. A conceptual representation that is activated as a result of the perception of speech and gesture can subsequently influence the production of speech and gesture.

The postcard model can also provide such a higher-level account. A communicative intention that is activated through the perception of gesture and speech, could subsequently inform gesture and speech production. This route is also possible in the interface model, but the interface model would still allow for speech and gesture to be coordinated during the formulation stage of production as well. Additionally, the interface model may be able to account more readily for adaptation of gesture forms at lower levels than the conceptual level, since the action generator has access to information specifically relevant to gesture, and can coordinate the generation of a gesture with the generation of a spoken message directly. Since neither of these models includes another interlocutor or comprehension of speech and gesture explicitly, these accounts remain speculative.

It seems that current models of gesture and speech production, combined with the interactive alignment account allow for both a low-level explanation (not involving representations of meaning) and a higher-level explanation (involving representations of meaning) of adaptation in gesture, depending on what representations are shared between perception and production, and at what levels these processes are linked. Can studies on adaptation in gesture tell us more about whether representations of meaning are necessarily involved when a perceived gesture form is subsequently produced?

Repetitions of co-speech gestures across interlocutors

Bergman and Kopp (2009) found that properties of a shape to be described influence what representation technique is chosen in gesture. This means that if two speakers are discussing the same objects, their gestures may look similar as a direct result of the content they are expressing, rather than because they adapt their representations of meaning to each other, or because they mimic each other's movement. So the first question to be answered is whether adaptation occurs in gesture at all.

Compared to adaptation in speech, relatively few studies (e.g. De Fornel, 1992; Holler & Wilkin, 2011; Kimbara, 2006, 2008; Parrill & Kimbara, 2006; Tabensky, 2001) have addressed adaptation in co-speech gesturing. Kimbara (2008), for example, studied dyads while they were jointly retelling an animated cartoon to a camera, narrating such that a third person could understand. She found that when the two speakers could see each other, their representational gestures looked more similar than when they were separated by an opaque screen. This shows that adaptation occurs in gesture: speakers adapted the form of their gestures to the form of another speaker's gestures. This is an important finding. It shows that similarities in interlocutors' gestures did not arise solely because their production tasks were similar, but that seeing each other was critical. However, this study does not reveal whether interlocutors adapted to each other's gestures due to automated motor-mimicry following the perception–behavior link, without intervention of the conceptual level, or whether certain gesture forms were linked to certain representations of meaning at the conceptual level, which caused the forms to be repeated when the same concepts were discussed.

Parrill and Kimbara (2006) found that gestures can also be repeated by an observer to a conversation, while subsequently addressing yet another person. In this study, participants were asked to watch a stimulus movie in which two women were discussing what route to take through a model city in front of them. They found that when participants watched a movie in which the women repeated more features of each other's hand gestures, they were more likely to produce these features in their own gesturing later on, while retelling the stimulus movie to the experimenter, compared to when they had seen a movie in which the women repeated fewer of each other's gesture features. A similar yet independent effect was found for verbal repetitions. Parrill and Kimbara conclude that people are very sensitive to repetitions across interlocutors in both gesture and speech. This study also shows that the repetition of perceived gesture forms does not happen exclusively between conversation partners, but also when addressing a different person than the one who produced the original gesture. This suggests that adaptation of gesture forms in communication is not always part of an implicit negotiation process on meaning (Brennan & Clark, 1996). However, the study does not reveal whether participants repeated the observed gesture forms simply because they have a tendency to repeat observed behaviors (Chartrand & Bargh, 1999), or whether they repeated the relations between gestures, objects and representations of meaning used by the people they had observed.

Some indication that the relation between a representation of gesture form and a representation of meaning may be involved in adaptation in gesture forms comes from a study by Tabensky (2001). She observed spontaneous conversation between two participants who were freely discussing a certain topic, and found that in analogy to how verbal information can be repeated literally or be paraphrased, the same information from gesture could be repeated by another interlocutor with either a similar or a very different gesture. This tentatively suggests that interlocutors' representations were converging at the conceptual level, rather than at the level of gesture forms. Also, information contained in one interlocutor's verbal description was found to end up in the other interlocutor's gestures, and vice versa. This suggests that there are links between representations of speech forms and gesture forms, possibly through the conceptual level. Tabensky found that gesture rephrasing only occurred at places where speakers were creating their own meaningful expressions, and not when literally repeating the other person verbally, in which case no gestures were produced. She therefore concludes that gesturing may be intrinsically related to the creation of meaning. This goes well with the idea that similarities in peoples' gestures result from similarities at the conceptual level, where communicative intentions are formed. However, in addition to these observational results, more empirical evidence is needed to support this causal claim.

Cassell, McNeill, and McCullough (1998) studied the relation between representations of meaning and gesture experimentally. They propose that both the perception of gesture and the perception of speech contribute to the construction of an internal representation. This representation of meaning can in turn inform gesture and speech production. This theory is based on an analysis of speakers who retold a story that they had seen a speaker tell in a movie clip. The speaker in the stimulus movie sometimes conveyed different information in gesture than he did in speech. For example, he would say "lure" and gesture either a grabbing or a beckoning motion. The information from the speaker's gestures was found to affect both participants' gestures and their speech. Thus, it seems that information obtained from gesture can subsequently be expressed in speech and in gesture, which suggests a representation of meaning being shared between gesture and speech interpretation and production. Yet are such representations key to the repetition of gesture forms across interlocutors?

Holler and Wilkin (2011) propose that reproducing each other's gestures in dialogue contributes to the creation of mutually shared understanding. For example, copying a gesture could signal the acceptance of an accompanying referring expression. This would mean that the reproduction of a gesture form signals that a similar representation of meaning has been created at the conceptual level. However, in this study the definition of a copied gesture included that the gesture had the same meaning as the original gesture. Therefore, this study gives a functional account of the copying of gestures assuming that meaning is involved, rather than questioning whether representations of meaning are necessarily involved in the reproduction of gesture forms across interlocutors.

In sum

On the one hand we see that interlocutors adapt to each other's gesture forms less when interlocutors cannot see each other, and that people do not exclusively adapt their gestures to their conversation partner. This goes well with a model in which perceiving a certain gesture form activates a representation of that gesture form, which is subsequently more likely to be selected for production, analogous to Pickering and Garrod (2004). Adaptation in gesture would then be driven by representations of form converging across interlocutors, rather than representations of meaning.

On the other hand, we see that information from one interlocutor's speech can subsequently be expressed in another interlocutor's gesturing, and also that information from one interlocutor's gestures can subsequently be expressed verbally by another interlocutor. This suggests that the same representations of meaning may underlie the production of both speech and gesture, and that similarities in speech and gesture forms across interlocutors may result from them having constructed similar representations of meaning. This is consistent with the models of speech and gesture production proposed by De Ruiter (2000) and Kita and Özyürek (2003). But is this really the case? To what extent is the gesture form produced by one interlocutor determined by the mere perception of a gesture form produced by another interlocutor, and to what extent is it determined by the producer's representation of meaning at the conceptual level?

Present study

We use an experimental approach to address these questions. First, we seek to confirm whether seeing a certain representational gesture while hearing certain content in speech, increases the likelihood of producing that same gesture later on, while expressing the same content. For this a speaker in a stimulus movie either does or does not perform certain gestures. These gestures were chosen such that they added very little meaning to the verbal description, for example the speaker moved his arms as though running while talking about running (as opposed to for example making this gesture while only mentioning 'going', in which case it would add much more information to the verbal description). This way, we can observe whether any similarity between the originally perceived gesture and a subsequently produced gesture results from expressing similar content, or whether it is necessary for the producer to actually observe the original gesture.

Second, we address the question of whether hand gestures being meaningful is relevant for their repetition across speakers to occur. We do so by keeping the gesture forms constant across conditions, but varying whether a form matches the content of the co-occurring speech. For example, the speaker would produce the above-described running gesture while talking about looking through binoculars. We predict that if meaning is involved in the repetition of gesture forms, participants will repeat only those gestures whose form matches the content of the concurrent speech. Contrastingly, if the property of carrying meaning is not relevant for the copying of form to occur, or in other words the conceptual level is not involved, gesture forms will be repeated equally often, independent of where in the narration they occur. Together, these first two experiments address the issue of whether hand gestures are similar to lexical forms when it comes to their repetition across interlocutors, or whether they are better compared to behavioral mimicry, such as the mimicking of each other's foot shaking and the rubbing of one's face.

We then zoom in on the role of representations of meaning in the repetition of gesture forms across interlocutors. We test whether a perceived gesture form influences the construction of a representation of meaning, which subsequently influences gesture production. We do so by looking at different physical features of a gesture. Suppose that certain features of a perceived gesture form give rise to the construction of meaning. Then when this meaning is subsequently expressed in gesture, all features of the produced gesture will be consistent with this meaning. So we would expect that features of the perceived gesture that were inconsistent with the meaning constructed would not be repeated. On the other hand, if the repetition of gesture forms is not mediated by a representation of meaning, any combination of perceived features could subsequently be produced. This means that we would expect a literal repetition of the perceived gesture, or any of its features, rather than all features being consistent with a certain meaning.

Experiment 1a: repetition of gesture form

In this experiment we test whether perceiving certain gestures while hearing a story, increases the chance of performing those gestures later on, while retelling the same story.

Participants

Participants to this and the following experiments were all adult native speakers of Dutch and they only took part in one of the experiments. Most of them were students at Tilburg University. All of them gave informed written consent for the use of their data. Experiment 1A had 38 (28 female) participants.

Stimuli

Two movie clips were created, in which the same male speaker told the same story of an animated cartoon ('Canary Row' by Warner Brothers) as though he had just watched it. Each movie clip consisted of ten fragments. It started with a short introduction in which the speaker stated that the cartoon was a Tweety and Sylvester movie in which Sylvester (a cat) tries to capture Tweety (a pet bird). Then followed eight fragments in each of which the speaker described one episode of the cartoon, which corresponds to one attempt of Sylvester to catch Tweety. These fragments lasted about 15 s each. The final fragment consisted of a short closure. Blank video was inserted in between the fragments, allowing for the movie to be paused at appropriate times. The speaker was seated in a chair and looked straight into the camera. The image showed the entire upper-body of the speaker in front of a white wall, see Fig. 1.

The two versions of the stimulus movie differed only in the number of representational hand gestures that the

Fig. 1. Example of the repetition of a target gesture (left) by a participant (right).

speaker produced. In one version, he produced a representational gesture depicting an action for each episode of the cartoon. These gestures were based on retellings of participants in a previous study (Mol, Krahmer, Maes, & Swerts, 2009). They consisted of:

Binoculars:	Two hands (cylinder shaped) are held in front of the eyes as the speaker looks through them, representing looking through binoculars. The hands are moved slightly toward the face and back, such that the fingers describe the cylindrical shape of the binocular tubes.
Drainpipe:	Two hands/arms make climbing/grabbing motions while moving upward, depicting climbing up the drainpipe.
Rolling ball:	Two hands spin around each other from the wrists, while held in front of the speaker, representing rolling.
Money tin:	Right hand imitates the holding and shaking of a money tin.
Creeping:	Hands (flat, palms down) and arms are moved forward one by one, imitating a creeping motion.
Throwing the weight:	Two hands (fingers spread, palms facing each other) are held about 30 cm apart, while a motion is made starting at the head and moving forward in an arc, as though throwing something big away from oneself.
Swinging:	Two hands are held on top of each other and quickly make a grabbing motion above and to the side of the speaker's head, representing the grabbing of a rope.
Running:	Arms are moved as while running, close to the body of the speaker.

The verbal descriptions of these events were rich, such that the additional information expressed in gesture was minimal. In the other version of the movie clip no representational gestures were produced. No other hand gestures were produced in any of the two versions and care was taken to make the verbal descriptions, body posture, facial expressions, intonation, voice quality, and other prosodic factors maximally similar across the two versions.

In both versions, the speaker used eight target phrases. These were unusual wordings, for example "as a full-blown Tarzan" (Dutch: als een volleerde Tarzan) or "the yearly spring call of the canary" (Dutch: de jaarlijkse lenteroep van de kanarie). These target phrases were the same in both versions. Inclusion of these target phrases allows for comparison of adaptation in gesture and speech and serves as a control measure.

Procedure

Participants came to the lab and were assigned randomly to the 'Gestures' or 'No Gestures' condition. They read the instructions, which explained the task as a memory task in which they had to watch video fragments of a speaker telling a story and were asked to subsequently retell these story fragments to the experimenter. Participants were instructed to take as much time as they needed when retelling the stories. They were given the opportunity to ask further clarification and once all was clear the experiment started.

Participants first watched the introductory fragment, which they did not have to retell. Then they watched the fragments describing the cartoon episodes one at a time. After each fragment, participants paused the movie and turned ninety degrees such that they were facing the experimenter while they retold the story. The experimenter was blind to the experimental condition. A camera was placed to the side of the experimenter, recording the participant. Participants were told they were videotaped to facilitate our analyses afterward. The experimenter did not interrupt the participants and did not produce any hand gestures, but did show other non-verbal signs of listening to their story in a natural way (such as by eye-gazing behavior and head movements). Finally, participants watched the last fragment, which they did not have to retell. Note that participants only saw one of the two stimulus movies of a speaker retelling the original cartoon movie and did not see the animated cartoon themselves. The entire experiment took place within 20 min.

Coding

Each gesture in the stimulus movie of the Gestures condition occurred with a given content unit in the verbal narration. We coded participants' representational gestures produced with those content units in their own narration. We will refer to these points in the narration as *target moments*. For example, the binoculars gesture in the stimulus movie was produced while the speaker said that Sylvester was looking at Tweety through binoculars. In this case we looked at participants' gestures while they were describing the event that Sylvester looked at Tweety (the target moment). Gestures from each condition that matched the corresponding gesture in the stimulus movie of the Gestures condition in the hand shape used, the location and movement of the hands, and the event expressed in the concurrent speech, were labeled as *target gesture*. For an example, see Fig. 1. If a different gesture was produced with the content unit of the original target gesture this was labeled as a *different gesture*, and if no gesture was produced this was labeled as *no gesture*.

Initially all gestures were coded by a single coder. Reliability was assessed by having a second coder code a random sample of 20% of the retold fragments, $N = 60$. The two coders agreed on 88% of the labels. The inter coder reliability for the raters was Cohen's Kappa = .69, indicating substantial agreement (Landis & Koch, 1977). Given the observed marginal frequencies of the labels, the maximum value of Kappa was .79. In our analyses, we used the coding of the first coder.

If a full target phrase was used by participants this was labeled as a *verbal repetition*. If participants repeated one or more (yet not all) content words of the target phrase this was labeled as *partial verbal repetition*. A (partial) verbal repetition was counted as such regardless of when in the participants' retelling it occurred, yet unsurprisingly they occurred only during retellings of the matching episode in the stimulus movie.

Analysis

We compared the means across conditions for all dependent variables in this and the following experiment. When Levene's test for equality of variances was significant, we used the unequal variance *t*-test. We report mean differences between the compared conditions (M_D), 95% confidence intervals (CI) and we report ω^2 as a measure of effect size.

Results

Gesture

The number of *target gestures* produced at target moments was higher in the Gestures condition ($M = 1.28$, $SD = 1.84$) than in the No Gestures condition ($M = .11$, $SD = .32$) ($M_D = 1.17$, 95% CI = .24, 2.09), $t(18.05) = 2.65$, $p < .02$, $\omega^2 = .14$. We did not find an effect of condition on the number of *different gestures* produced at target moments (Gestures: $M = .94$, $SD = 1.11$, No Gestures: $M = 1.00$, $SD = 1.65$), $t(34) = .12$, $p = .91$. There was a trend toward significance for the number of target moments at which *no gesture* was produced, which tended to be higher in the No Gestures condition ($M = 6.89$, $SD = 1.64$) than in the Gestures condition ($M = 5.78$, $SD = 2.10$) ($M_D = 1.17$, 95% CI = .24, 2.09), $t(34) = 1.77$, $p = .09$, $\omega^2 = .06$.

Speech

We did not find a significant difference in the number of verbal repetitions between the Gestures ($M = 1.61$, $SD = .99$) and No Gestures condition ($M = 1.50$, $SD = .92$) ($M_D = .11$, 95% CI = −.53, .76), $t(34) = .35$, $p = .73$, nor in the number of partial verbal repetitions across the two conditions (Gestures: $M = 1.83$, $SD = 1.34$, No Gestures: $M = 1.56$, $SD = 1.38$) ($M_D = .45$, 95% CI = -1.20, .64), $t(34) = .61$, $p = .54$.

Discussion

Participants produced certain representational gestures more often if they had seen these gestures in the stimulus movie. Like Kimbara (2008), we found that expressing the same content was not sufficient for these repetitions to occur. Neither was seeing the speaker or an addressee, which participants did in both of our conditions. Rather, seeing the target gestures performed by the speaker in the stimulus movie increased the likelihood of participants producing the same gestures during their own narration to a different addressee. The fact that participants repeated some target phrases and that there was no difference between the two conditions in the number of verbal repetitions shows that participants did adapt somewhat to the speaker, regardless of whether he gestured.

The fact that participants reproduced gesture forms even though the addressee was different from the speaker in the stimulus movies, suggests that low-level processes, such as priming, may underlie these repetitions, rather than the construction of shared meaning across interlocutors. Seeing a certain form increased the likelihood of producing that form later on. Yet we do not know to what extent a representation of meaning was involved as well. How important was it that these gestures carried meaning for the repetition of their form to occur?

Experiment 1b: repetition of gesture form and the semantic context

In this experiment we test whether the repetition of a gesture's form across speakers depends on the gesture's meaning in relation to the meaning of the concurrent speech.

Participants

Forty-seven participants (33 female) volunteered for this study.

Stimuli

Again two stimulus movies were produced, which were similar to the clips in the previous experiment. The first stimulus movie was made in the same way as the one containing representational gestures in the previous experiment (1A). In the second stimulus movie, the speaker produced one representational gesture per episode as well, however this time the gesture did not match the speaker's verbal description. A gesture from another episode was produced instead of the original gesture, along with the original content unit in the verbal description. For example, instead of the 'binoculars' gesture, the speaker produced the 'running' gesture while verbally referring to the event involving binoculars, see Fig. 2. The same speaker and the same target phrases were used as in the previous experiment and again care was taken to make the verbal descriptions, body posture, facial expressions, intonation, voice quality, and other prosodic factors maximally similar.

Procedure

The procedure was the same as in the previous experiment. In the Congruent condition, 24 participants saw and retold the stimulus movie in which the gestures matched the content of the concurrent speech. In the Incongruent condition, 23 participants saw and retold the stimulus movie in which the gesture forms were mixed up and did not match the content of the concurrent speech. When asked by the experimenter, none of the participants showed any indication of suspecting that the experiment was about the repetition of hand gestures.

Coding

In the stimulus movies, gestures occurred at a certain content unit in the verbal narration. Initially, we coded participants' representational gestures produced with those content units in their own narration, that is, at the *target moments*. We coded gestures that matched the corresponding gesture in the movie that the participant had seen in the hand shape used, the location of the gesture and the movement involved in the gesture as *target gesture*, similar as before. We added the label *partial target gesture*, for gestures that matched the gesture in the stimulus movie in two out of these three features (hand shape, location, movement). This time, it was also possible that a participant spontaneously produced the target gesture shown in the other condition. For example, if a participant was shown the running gesture with a description of the event in which Sylvester looks at Tweety through binoculars, this participant could still produce the binoculars gesture while narrating that Sylvester looked at Tweety. Such cases were labeled as *target gesture other condition*. If a different gesture was produced at a target moment this was labeled as *different gesture*, and if no gesture was produced this was labeled as *no gesture*.

If one thinks of gesture and speech as fully separate behaviors, a theory of motor-mimicry is non-specific about the moment when a target gesture will be reproduced. Therefore, we also looked for target gestures from the moment a gesture was presented to a participant till the end of the experiment, rather than at target moments only. Verbal repetitions were coded in the same way as before.

Initially all gestures were coded by a single coder. Reliability was assessed by having a second coder code a random sample of 20% of the retold fragments, $N = 75$. The two coders agreed on 91% of the labels, Cohen's Kappa = .82, indicating almost perfect agreement (Landis & Koch, 1977). Given the observed marginal frequencies of the labels, the maximum value of Kappa was .87. In our analyses, we used the coding of the first coder.

Results

Gesture

The number of *target gestures* produced at target moments was higher in the Congruent condition ($M = .79$, $SD = 1.10$), than in the Incongruent condition ($M = .04$, $SD = .21$) ($M_D = .75$, 95% CI = .28, 1.22), $t(24.71) = 3.26$, $p < .01$, $\omega^2 = .17$. This was also the case for target gestures that were produced at any time from the presentation of the gesture till the end of the experiment (Congruent: $M = .79$, $SD = 1.10$, Incongruent: $M = .09$, $SD = .29$) ($M_D = .71$, 95% CI = .23, 1.18), $t(26.26) = 3.03$, $p < .01$, $\omega^2 = .15$. We found no effect on the number of *partial target gestures* produced at target moments (Congruent: $M = .96$, $SD = 1.23$, Incongruent: $M = .57$, $SD = .73$), $t(45) = 1.32$, $p = .19$.

At target moments, participants in the Congruent condition never produced target gestures from the Incongruent condition. Yet participants from the Incongruent condition

Fig. 2. Congruent (left) and Incongruent (right) gesture for the content unit 'Sylvester looks through binoculars'.

sometimes produced target gestures from the Congruent condition at the target moment of those gestures. Thus, the number of *target gestures from the other condition* was higher in the Incongruent (M = .17, SD = .39) than in the congruent condition (M = .00, SD = .00) (M_D = 1.74, 95% CI = .01, .33), $t(22) = 2.15$, $p < .05$, $\omega^2 = .07$. These included both gestures that had been presented to participants from the Incongruent condition with earlier episodes in their stimulus movie and gestures that these participants had not yet seen. Thus, as in the previous study, participants sometimes spontaneously produced target gestures that they had not seen. Therefore, we compare the number of target gestures from the Congruent condition that were spontaneously produced in the Incongruent condition (*target gesture other condition*) to those produced in the Congruent condition (*target gesture*). The number of target gestures from the Congruent condition produced at their target moments was lower in the Incongruent condition (M = .17, SD = .44) than in the Congruent condition (M = .79, SD = 1.10) (M_D = .62, 95% CI = .13, 1.11), $t(45) = 2.58$, $p < .02$, $\omega^2 = .11$. See Fig. 3 for an overview of these results.

Participants in the Incongruent condition more often produced *no gesture* at the target moments (M = 6.87, SD = 1.10) compared to participants in the Congruent condition (M = 6.00, SD = 1.38) (M_D = .87, 95% CI = .04, .14), $t(45) = 2.38$, $p < .05$, $\omega^2 = .09$. We did not find an effect on the number of *different gestures* produced at target moments (Congruent: M = .35, SD = 1.15, Incongruent: M = .25, SD = 1.45), $t(45) = 2.38$, $p = .80$.

Speech

Partial verbal repetitions occurred more often in the Congruent condition (M = 1.92, SD = 1.06), than in the Incongruent condition (M = 1.26, SD = .92) (M_D = .66, 95% CI = .07, 1.24), $t(45) = 2.27$, $p < .05$, $\omega^2 = .08$. We did not find a significant effect for full repetitions across the two conditions (Congruent: M = 1.13, SD = .34, Incongruent: M = 1.09, SD = .60) (M_D = .04, 95% CI = −.25, .32), $t(45) = .27$, $p = .79$.

Discussion

Participants repeated gesture forms more frequently if they had been presented to them in a linguistic context in which they were meaningful. Only one target gesture was repeated in the Incongruent condition. In this case the verbal retelling was adjusted such that it matched the gesture. Instead of telling that Sylvester climbed up the drainpipe as had been told in the stimulus movie, the participant said that Sylvester moved past Tweety while producing the gesture, which was a reproduction of the gesture depicting the swinging event. (This was still counted as a repetition of the target gesture at a target moment, since it occurred with the content unit of Sylvester's movement toward Tweety.) These results suggest that representations of meaning do play a role in the repetition of meaningful gestures across speakers. If observing a form would lead to the repetition of that form directly and automatically, or if seeing hand movements alone would cause participants to produce more target gestures, then there

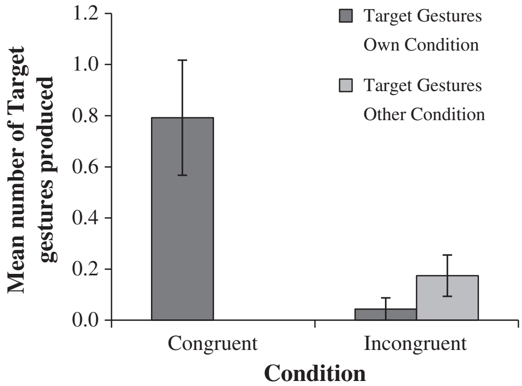

Fig. 3. Mean number of target gestures from the participant's own and the other condition (re)produced by participants in the Congruent and Incongruent condition. Bars represent standard errors.

would be no reason why gestures that were not meaningful in the linguistic context were less likely to be repeated. This sets the repetition of representational hand gestures across interlocutors aside from the mimicking of behaviors that do not carry propositional meaning, such as the shaking of one's foot or other hand movements.

However, there is a possible confound. It may have been the case that participants were just less likely to adapt to a speaker who came across as somewhat incoherent, due to his non-matching gestures. The result that participants also repeated the target phrases a bit less when they saw the non-matching gestures is consistent with this explanation. In experiment 2 we examine the relation between the repetition of gesture forms and representations of meaning in more detail, this time with a more subtle manipulation of form-meaning correspondence.

Experiment 2: repetition of gesture form and the underlying representations

In this experiment we investigate whether a perceived gesture form can influence the construction of meaning (whether it be any semantic representation or a conceptual pact), which subsequently influences gesture production. To test this we have used a route description task. By asking a participant and a confederate to give each other directions repeatedly, a situation was created in which it is quite natural to repeat each other's gesture forms, without drawing much attention to this process. Because of the more interactive setting, both low-level and high-level processes (e.g. audience design) that may be involved in the repetition of gesture forms across interlocutors can come into play.

Giving directions allows for different conceptualizations of the task at hand. We presented participants with bird's view drawings of a city scene, which had a short route indicated on them (see Fig. 4 for an example). These scenes were neither presented vertically nor horizontally, but at an angle. Therefore, the production task could be thought of as either describing a route on a vertically

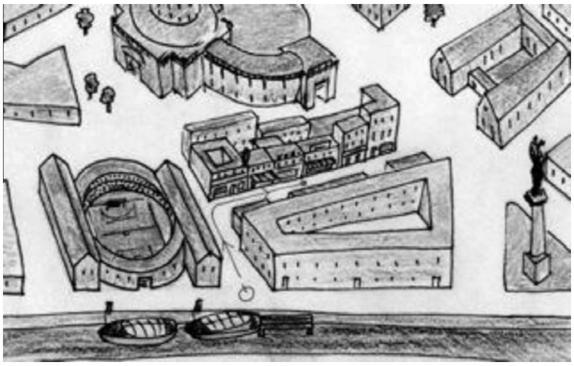

Fig. 4. Part of a city scene used in the experiment, note the route starting at the bottom-center.

oriented map, or as describing a route through an actual (horizontal) city.

The confederate always was the first to give a route description. Although her verbal descriptions were the same across conditions, her gestures differed. The movement of her gestures was either in accordance with the conceptualization of indicating the route on a vertically oriented map, which we will call *Vertical Map* perspective, or with the conceptualization of a route through a city, which we will call *Route* perspective. Thus, two different conceptualizations were suggested in gesture.

It is interesting in itself to see whether participants adapt to the confederate's perspective in gesture. Yet this alone would not tell us whether this is based on direct repetition of gesture form, or on the convergence of representations of meaning. Therefore, apart from manipulating perspective, we manipulated hand shape independently. Our intuition was that when describing a route through a city, hand points, where all fingers are extended as an index, as well as finger points, where only one finger is extended as an index can be used, whereas when pointing on a map it is more common to point with one finger than with all fingers extended.

We tested this intuition in two ways: by analyzing pictures on the internet and by asking people in the streets for directions. First, we did a Google search for pictures on the bigram "getting directions" on December 14th 2010 in the Netherlands. We selected those photographs in which a person was pointing, seemingly to communicate to another person, and whose pointing hand was clearly visible. We scored photographs on the first five pages of search results for whether the person in the picture was pointing at a map or not and whether they were pointing with their hand or with their finger. Duplicate search results were counted only once. We found 6 finger points at a map, 0 hand points at a map, 10 finger points that were not on a map, and 8 hand points not on a map, in line with our hypothesis. Next, we searched for more gestures pointing at a map with the term "pointing at a map". We found 50 finger points at a map, 1 hand point at a map, and 7 points at a map with another index, such as a pen or a stick. These data support our hypothesis that finger pointing is the most common way of pointing at a map.

Additionally, we asked 20 Dutch-speaking adults in the city center of Tilburg for directions, while holding a paper map in hand. Out of the gestures that were pointing at the map, 29 were produced with one finger as an index, while 1 was produced with the tip of a key. Out of the gestures that were pointing into the streets, 18 were produced with one finger as an index and 20 were produced with all fingers extended. We excluded the 'key-gesture' from our analysis and found that this distribution is unlikely to be accidental, Yates $\chi^2(1) = 19.32$, $p < .0001$. Our hypothesis that people are less likely to point with all fingers extended when pointing at a map was confirmed by the data.

We use this difference in common hand shapes between the domain of giving directions using a map and the domain of giving directions in the streets to address our research question. If it is the case that gesture form is perceived and reproduced directly, without the conceptual level being involved, participants may adapt to any of the gesture features produced by the confederate. That is, they may adapt to the confederate's hand shape and to the confederate's perspective. One of these may be perceived more easily than the other, so there could be a difference in the extent to which each of the features is adapted to, but what we would not expect based on this view, is for the confederate's perspective to influence a participant's hand shape or for the confederate's hand shape to influence the participant's perspective.

On the other hand, if meaning does form an intermediate stage between the perception and production of a gesture form, we do expect such cross-effects to occur. For example, our pre-test showed that it is more common to point at a map using a single finger, than it is to point at a map using four fingers at once. Therefore, if the confederate's vertical movements would lead participants to think of this task as describing a route on a map, their gestures may be more frequently produced with one finger as an index as opposed to four. This would mean there is an effect of the perspective of the confederate's gestures on the hand shape of participants' gestures. This effect may also be found in the reverse direction: the use of all fingers as an index may lead participants to more readily think of the route as through a city than on a map, causing them to gesture in the Route perspective rather than the Vertical map perspective.

Participants

Forty-eight participants took part in this experiment, out of which we excluded 6 from our analysis because they did not produce any of the gestures we were interested in (path gestures) and 2 because they indicated some suspicion about the experiment, see the sections *Procedure* and *Analysis* below. We used the data of 40 (33 female) participants in our analysis.

Procedure

The participant and the confederate came to the lab and were introduced by the experimenter. They each received a written instruction and were seated across from each other. The instruction explained a communication task, and stated that the couple with the most correct responses could win a book voucher (in reality there was a random draw). To their side (right to the participant) was a table, on which sat a flip chart for each interlocutor. In between these flip charts was a screen, such as to keep information private. The screen did not keep the interlocutors from seeing each other. Both behind the confederate and behind the participant was a camera capturing the other interlocutor. After reading the instruction, both 'participants' were allowed to ask questions. The confederate always asked one question, after which the experimenter quickly went over the task again. Then the experimenter turned on the cameras and left the room.

The confederate started by studying a little map and memorizing the route on it. Each route had one turn, see Fig. 4 for an example. She then turned the page of her flip chart (rendering a blank page) and described the route to the participant, for example: "Je begint bij de rondvaartboot, dan ga je langs het voetbalstadion en dan rechts een winkelstraat in tot ongeveer halverwege." ("You start at the tour boat, then you go along the soccer stadium and then into a shopping street on the right until about halfway.") The confederate's speech followed a script and was the same in each condition. The terms used to describe the directions were consistent with a horizontal perspective such as 'rechtdoor' (straight ahead) and 'steekt over' (cross), or were neutral for perspective 'tot ongeveer halverwege' (until about halfway). Gestures were timed naturally with speech and gazed at by the confederate. The confederate gestured with her right hand. The first direction of a route was always straight, which was depicted with either a forward or an upward movement. These movements were of comparable size. The gesture for the second direction (to the side) was placed relative to the first gesture; it started where the first gesture had ended.

After the confederate's description, the participant turned a page and was to choose which route had just been described, selecting from four alternatives by pronouncing the corresponding letter, see Fig. 5. No feedback was provided. Then it was the participant's turn to study a route. This route was always on the same scene that the confederate's route had been on. After turning the page (rendering a blank page) the participant described the route to the confederate, who then turned a page and selected one of the four alternatives. This ended one cycle of the experiment. In total each participant perceived and produced five route descriptions, which took between 6 and 11 min. The confederate's descriptions took about 12 s each. On average, participants took about equally long for their descriptions, ranging from 7 to 19 s seconds. (Most time of the experiment was filled with studying the maps and selecting answers.)

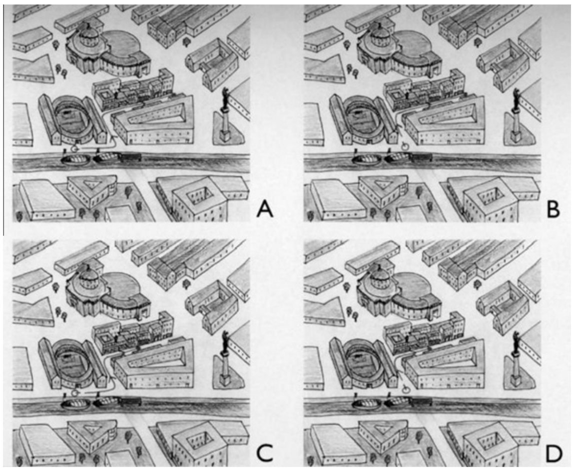

Fig. 5. Example of the alternative routes to choose from. Each map has a slightly different route depicted on it. Participants selected a route by calling out the corresponding letter.

Afterward, both the confederate and the participant filled out a questionnaire, which included questions on the presumed purpose of the experiment and whether the participant noticed anything peculiar. Participants were also asked if they had recognized the city in the pictures, which none of them had (the drawings were loosely based on St. Petersburg). When the participant was done filling out the forms, the confederate revealed her role and asked the participant's consent for the use of their data. Participants were asked if they had suspected any deception. The data of two participants was excluded from our analysis, because they indicated having been suspicious about either the goal of the experiment or the role of the confederate.

Design

We used a 2×2 between subjects design. The independent variables were the perspective (Route or Vertical map) and hand shape (one or four fingers extended) of the confederate's path gestures. In the Route perspective, gestures were performed in the horizontal plane, with the index in the direction of the hand movement, as though following a virtual route (Fig. 6a and b). In the Vertical Map perspective, gestures were performed in the vertical plane and the index was always pointing forward, as though pointing on a virtual map (Fig. 6c and d).

Coding

We coded all *path gestures* that participants produced, that is, all gestures in which one or more fingers were extended as an index, there was hand movement along some virtual path, and the co-occurring speech mentioned a direction to take. Within the stroke phase of each path gesture, we coded hand shape and perspective. The labels for hand shape were *Finger*, when one finger was extended as an index, and *Hand*, if more than one finger was extended. The label for perspective was based on the following features of participants' gestures: location in the gesture space, hand orientation, and movement (direction and size). The label that could explain most features was assigned to each gesture. It turned out that in addition to the two perspectives that the confederate had used, participants occasionally used an alternative one, as though pointing on a horizontal map. Therefore, we chose from three labels: *Vertical Map*, *Route*, and *Horizontal Map.* A gesture in the Vertical Map perspective typically has vertical

Fig. 6. The confederate's path gestures. a: Hand/Route; b: Finger/Route; c: Hand/Vertical Map; d: Finger/Vertical Map.

movement, with relative sizes mapping onto distances on the map, fingers pointing forward and the location in the gesture space corresponding to the location on the map (Fig. 7a and b). Route gestures on the other hand have horizontal movement in front of and to the side of the speaker, with the fingers pointing in the direction of the hand movement (Fig. 7c). Horizontal Map gestures (Fig. 7d) differ from Route gestures in their hand orientation (fingers pointing down), and their relative size and location. Fig. 7 shows some examples of participants' path gestures and our coding.

Initially all gestures were coded by a single coder. Reliability was assessed by having a second coder code a random sample of 20% of the path gestures in each condition for hand shape and perspective, $N = 58$. The two coders agreed on the label for hand shape in 95% of the cases, Cohen's Kappa = .89, indicating almost perfect agreement (Landis & Koch, 1977). Given the observed marginal frequencies of the labels, the maximum value of Kappa was .96. The two coders agreed on the label for perspective in 79% of the cases, Cohen's Kappa = .66, indicating substantial agreement (Landis & Koch, 1977). The maximum value of kappa was .86 in this case. In our analyses, we used the coding of the first coder.

Fig. 7. Examples of participants' path gestures and our coding. a: Hand/Vertical Map; b: Finger/Vertical Map; c: Hand/Route; d: Finger/Horizontal Map.

Analysis

Analyses were done using ANOVA, with factors perspective (levels: Vertical Map, Route) and hand shape (levels: Hand, Finger) of the confederate's gestures. There were 40 participants, 10 in each cell. As a measure of participants' perspective, we report the number of path gestures that a participant produced in the Vertical Map perspective divided by all path gestures produced by that participant. As a measure participants' hand shape, we report the number of gestures that a participants produced with one finger extended, divided by the total number of path gestures produced by that participant. The significance threshold was .05 and we report partial eta squared as a measure of effect size.

Results

Participants' perspective

Analyses of the perspective of participants' gestures, shown in Fig. 8, revealed a main effect of the confederate's perspective, such that participants produced a larger proportion of Vertical Map gestures when the confederate gestured in the Vertical Map perspective ($M = .46$, $SD = .35$) than when the confederate gestured in the Route perspective ($M = .11$, $SD = .20$) ($M_D = .35$, 95% CI = .17, .54), $F(1, 36) = 14.88$ $p < .001$, $\eta_p^2 = .29$. The confederate's hand shape did not exert a main effect on the perspective of participants' gestures, as they produced about equal proportions of Vertical Map gestures when the confederate gestured with all fingers extended ($M = .26$, $SD = .34$) and when she gestured with one finger extended ($M = .31$, $SD = .33$) ($M_D = -.06$, 95% CI = $-.24$, .13), $F(1, 36) = .38$, $p = .54$. These two factors did not interact, $F(1, 36) = .71$, $p = .41$.

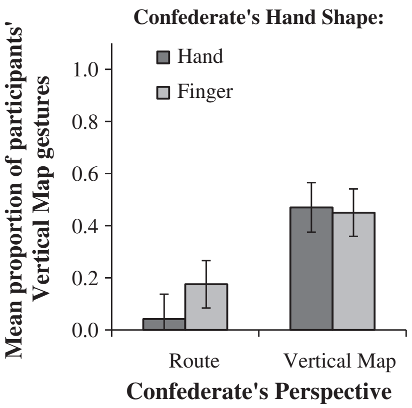

Fig. 8. Mean proportion of gestures that participants produced in the Vertical Map perspective for each perspective and hand shape used by the confederate. Bars represent standard errors.

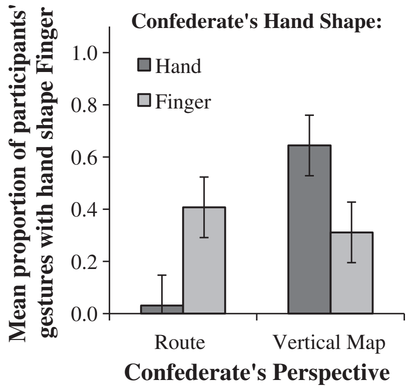

Fig. 9. Mean proportion of gestures that participants produced with one finger extended (hand shape Finger) for each perspective and hand shape used by the confederate. Bars represent standard errors.

Participants produced Horizontal Map gestures in about equal proportions across conditions. Therefore, the results for participants' gestures in the Route perspective mirror the results reported above.

Participants' hand shape

Analyses of the hand shape of participants' gestures, shown in Fig. 9, revealed a main effect of the confederate's perspective, such that participants produced a larger proportion of gestures with one finger extended when the confederate gestured in the Vertical Map perspective ($M = .48$, $SD = .43$) than when the confederate gestured in the Route perspective ($M = .22$, $SD = .37$), $F(1, 36) = 5.00$, $p < .05$, $\eta_p^2 = .12$. The confederate's hand shape did not exert a main effect on participants' hand shape, as participants produced about equal proportions of gestures with one finger extended when the confederate gestured with all fingers extended ($M = .34$, $SD = .39$) and when the confederate gestured with one finger extended ($M = .36$, $SD = .45$) ($M_D = -.02$, 95% CI = $-.26$, .21), $F(1, 36) = .04$, $p = .85$. These two factors interacted, $F(1, 36) = 9.40$, $p < .01$, $\eta_p^2 = .20$: When the confederate gestured in the Route perspective, participants adapted to her hand shape, as they produced a larger proportion of gestures with one finger extended when the confederate gestured with one finger extended ($M = .41$, $SD = .45$) than when she gestured with all fingers extended ($M = .03$, $SD = .07$) ($M_D = .38$, 95% CI = .08, .68), $F(1, 18) = 6.94$, $p < .02$, $\eta_p^2 = .28$. Yet when the confederate gestured in the Vertical Map perspective, participants did not adapt to her hand shape, as they produced a larger proportion of gestures with one finger extended when the confederate gestured with all fingers extended ($M = .64$, $SD = .32$) than when she gestured with one finger extended ($M = .31$, $SD = .48$), $F(1, 18) = 2.24$, $p = .16$.

Discussion

As predicted by both the theory that gesture form is copied directly and theories that representations of meaning are involved, participants adapted to the perspective used by the confederate. When the confederate gestured in the Vertical Map perspective, participants were more likely to do so as well and similarly for the Route perspective. But was this merely a copying of form, or were they adapting to the conceptualization expressed in the confederate's gestures?

Importantly, we found some of the cross-effects we expected if perceiving certain gestures would cause participants to think of the task in a certain perspective, which in turn would influence their own gesture production. The perspective of the confederate's gestures influenced the hand shape of participants' gestures: participants more frequently pointed with one finger if the confederate gestured as though on a vertical map. This can be explained as the confederate's vertical gestures leading participants to think of the route as on a map, which caused them to point with their finger. If this is so, we would expect this increase in the use of one finger being caused by gestures that participants produced with the Vertical map perspective, rather than in the Route perspective. This was indeed the case. It thus seems that participants were not merely repeating the individual features of the confederate's gestures, but rather the meaning that they expressed.

A theory that only takes into account the alignment of gesture forms can also explain that participants adapted to the confederate's perspective. Yet such a theory would not predict participants to gesture with one finger as an index more frequently when the confederate gestured in the Vertical Map perspective, particularly not when the confederate gestured with all fingers extended. However, it is possible to explain this effect in terms of biomechanics. For example, it may be the case that vertical hand movements lead people to extend their index finger, rather than all fingers, simply because it is easier, without them associating this with pointing at a map at any level of processing (our theory is neutral as to whether conceptual representations are more embodied or more symbolic in nature). This would be an explanation without conceptual mediation. However, we do not know of any data that support a theory that it is easier to extend one finger instead of all fingers when lifting an arm. An informal pilot study showed that when people were asked to copy the confederate's gestural movements, without any meaning being ascribed to them, people could do so effortlessly, both in terms of movement and hand shape. In our view, this makes an explanation in terms of biomechanics less likely. Also, rather than explaining the data post hoc, our theory predicts the effect we found, making it more powerful.

Overall, perspective was adapted to more than hand shape. This may be because perspective was expressed in two features (movement and hand orientation), whereas hand shape is only one feature. Thus, the one non-matching feature may have been adapted to the two matching ones. It may also be because in this task, perspective carried a more critical meaning than did hand shape. A vertical gesture cannot possibly depict a route one could walk (at least not in the Netherlands), whereas the distinctions between the different hand shapes seem far subtler. In other words, in this task, the perspective of gestures may have given rise to (shared) conceptualizations more readily than the hand shape with which they were produced.

Although we did not find an overall effect of hand shape, participants did adapt to the confederate's hand shape in the conditions in which she gestured horizontally in the Route perspective, whereas there was no adaptation to the confederate's hand shape in the Vertical Map conditions. A possible explanation for this is again one in terms of meaning. The different hand shapes may be more readily interpreted in a meaningful way when gesturing as though along a horizontal route than when gesturing as though on a vertical map, better allowing participants to construct concepts that were consistent with the confederate's hand shape.

General discussion and conclusion

Our experiments have shown that adaptation in representational gestures resembles adaptation in verbal references in various ways. First, certain gesture forms were more likely to be used after they had been perceived. Participants who saw a speaker in a stimulus movie produce certain representational gestures were more likely to produce these gestures later on, while retelling the speaker's story. Second, gesture forms were only repeated across speakers if they had occurred in a meaningful context. That is, if the gesture form could be interpreted in light of the meaning expressed in the concurrent speech. Lastly, there were instances where gesture form was not copied in a low-level automated way, but rather similar forms were used to express similar meanings, and aspects of a form that did not match a meaning were not copied but rather adapted to the meaning. These findings go well with theories and models in which gesture and speech both stem from a single concept, idea, or communicative intention (e.g. Cassell et al., 1998; De Ruiter, 2000; Kendon, 2004; Kita & Özyürek, 2003; McNeill, 1992).

In study 1B, when a perceived gesture was incongruent with the content of the accompanying speech, the gesture was not repeated when the speech content was. In terms of the interactive alignment account, this may be because the incongruent gesture form that was perceived did not match the representation of meaning that was formed in the interpretation process, and thus no link was established between the representation of meaning and a representation of gesture form. Therefore, when this representation of meaning was subsequently activated by the production process, the gesture form was not. This explanation also fits the one exception we found, where an incongruent gesture was repeated, but the content of speech in the retelling differed markedly from the original story. In this case, the interpretation formed during perception seems to have incorporated the incongruent gesture. Therefore, the representation of meaning activated during production could activate the gesture form that had been perceived, but not the by then incongruent lexical forms that were perceived. This one case is very similar to the results found by Cassell et al. (1998), who accounted for their results similarly.

Interestingly, participants were less likely to produce any gesture at all while expressing content that had been presented to them with an incongruent gesture. It may be that the perception of an incongruent gesture disturbed the activation of representations of the spatial and motor aspects of the event described, thereby making it less likely that a gesture was produced while retelling this event (also see Kelly, Özyürek, & Maris, 2010). In terms of the interface model (Kita & Özyürek, 2003), this can be explained as the gesture generator not being able to retrieve relevant data from working memory, and thus not being able to generate a gesture form.

Consistently, the gesture as simulated action framework (Hostetter & Alibali, 2010) would also predict that speakers are less likely to gesture when describing an event that does not involve the perception or performance of a particular action. This framework explains the production of representational gestures as simulating action as part of thinking for speaking. Therefore, this model might predict that not having perceived a gesture congruent with the upcoming speech would cause participants to be less likely to produce a representational gesture. Alternatively, it may also be that participants omitted the gesture for social reasons. It is known that adaptation has positive social consequences (e.g. Van Baaren, Holland, Steenaert, & Van Knippenberg, 2003), thus it may also be that producing a completely different gesture sends a negative social message, which participants may have wanted to avoid.

The results of our route directions study also suggest that concepts underlie the repetition of gesture forms across interlocutors. Participants readily adapted to those features of a confederate's gestures that could be interpreted meaningfully, such as whether the gestures were produced horizontally, as though walking through a city, or vertically, as though pointing on a vertically oriented map. However, features that were inconsistent with these conceptualizations, notably the use of four fingers while gesturing as though pointing on a map, were not adapted to. Instead, we saw an effect of the confederate's perspective in gesture on the hand shape used by participants: if the confederate gestured as though on a map, participants more frequently used one finger as an index, which is consistent with the conceptualization of pointing out the route on a map.

Both the theory that interlocutors form conceptual pacts (Brennan & Clark, 1996) and the interactive alignment account (Pickering & Garrod, 2004) can account for these findings. In the account by Brennan and Clark, speakers adapt to each other's gestures with the aim of arriving at a shared conceptual understanding (also see Holler & Wilkin, 2011). Although plausible, such a functional account is not required to explain our data. The interactive alignment account can do so as well. Yet adaptation at one linguistic level alone cannot explain why specifically those aspects of a perceived gesture that could not readily be interpreted meaningfully were not reproduced. Instead, these features tended to be produced such that they fitted an interpretation consistent with most aspects of the perceived gesture. This can be explained using the links between different levels within a speaker, that is, between representations of meaning and representations of form. Only those features of a form that can be linked to a representation of meaning during interpretation are activated through this same representation of meaning once it is activated for production.

Regarding proposed models of speech and gesture production, such as the postcard model (De Ruiter, 2000) and the interface model (Kita & Özyürek, 2003), our results seem to emphasize that it is important for these models to have gesture and speech production and interpretation share representations of meaning at some level. Both models allow for this at the conceptual level. Currently, these models do not provide an explicit account of adaptation in either gesture or speech, because they do not specify how production and perception are linked, or at what level representations are shared between these processes. In future work we intend to further study adaptation in gesture, adaptation in speech, and their possible influence on each other to shed more light on this issue. A model that can account for our current findings will need to allow for the conceptual level to influence what features of a perceived gesture form will be more likely candidates for gesture production.

As explained above, our results fit well with the theory that when communication partners interact, the concepts of both interlocutors converge and certain forms are used to refer to these shared concepts (Brennan & Clark, 1996; Garrod & Anderson, 1987). However, our results do not provide evidence that this is a deliberate process, or that it is part of audience design. Especially in our first two studies, audience design does not seem the most likely explanation, since even though the addressee was not the person who produced the original gestures, some of the original gesture forms were repeated when narrating to this new addressee. The convergence of representations of meaning may also happen automatically (Pickering & Garrod, 2004), without conscious effort or intent (but see Brennan & Hanna, 2009). This issue needs to be addressed in future research. Additionally, we have not yet studied gesture in a setting in which conceptual pacts can be arrived at incrementally. Rather, we have made use of a stimulus movie or a confederate whose gesturing followed a script, such that adaptation could only happen one way. Despite these limitations, our results suggest that the perception of meaningful forms in gesture can contribute to the convergence of concepts across interlocutors, which in turn informs gesture production.

Our results do not imply that features of gesture forms are never repeated without representations of meaning being involved. So far, we have only studied certain representational gestures. Our results may not generalize to other types of gestures, especially non-representational gestures, whose repetition across interlocutors may be more similar to that of other behaviors not carrying propositional meaning. Yet we have shown that certain representational gestures are only repeated if they make sense in the linguistic context and that one aspect of a perceived gesture form (perspective) can influence another aspect (hand shape) of a gesture form produced. These results suggest that it is sometimes fruitful to include representations of meaning in an explanation of adaptation in nonverbal language use, especially when these behaviors carry propositional meaning. Rather than perceiving a form leading to the production of that form directly, we have shown that for representational gestures, meaning can play a mediating role. That is, representations of meaning are also converging across interlocutors rather than just representations of form, and this convergence of meaningful representations may be driving adaptation in gesture.

Acknowledgments

We like to thank all participants. We thank the anonymous reviewers for their insightful comments to earlier versions of this paper. We gratefully thank Susan Brennan and Sotaro Kita for valuable discussions on earlier versions of this work. We thank Nathalie Bastiaansen for drawing the stimuli of experiment 2, Els Jansen and Anouk van den Berge for collecting and coding the data of experiment 1, Vera Nijveld for doing the reliability coding of experiment 2, and our co-workers at Tilburg University for assisting in the data collection.

References

Bavelas, J., Chovil, N., Lawrie, D. A., & Wade, A. (1992). Interactive gestures. *Discourse Processes, 15*, 469–489.
Bergman, K., & Kopp, S. (2009). *Increasing expressiveness for virtual agents – Autonomous generation of speech and gesture.* Paper presented at the International Conference on Autonomous Agents and Multiagent Systems (AAMAS) 2009, Budapest, Hungary.
Bock, J. (1986). Syntactic persistence in language production. *Cognitive Psychology, 18*, 355–387.
Branigan, H. P., Pickering, M. J., Pearson, J., & McLean, J. F. (2010). Linguistic alignment between humans and computers. *Journal of Pragmatics, 42*, 2355–2368.
Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology – Learning Memory and Cognition, 22*, 1482–1493.
Brennan, S. E., & Hanna, J. E. (2009). Partner-specific adaptation in dialogue. *Topics in Cognitive Science, 1*, 274–291.
Cassell, J., McNeill, D., & McCullough, K.-E. (1998). Speech–gesture mismatches: Evidence for one underlying representation of linguistic & nonlinguistic information. *Pragmatics & Cognition, 6*, 1–33.
Chartrand, T. L., & Bargh, J. A. (1999). The Chameleon effect: The perception–behavior link and social interaction. *Journal of Personality and Social Psychology, 76*, 893–910.
Chui, K. (2005). Temporal patterning of speech and iconic gestures in conversational discourse. *Journal of Pragmatics, 37*, 871–887.
Cleland, A. A., & Pickering, M. J. (2003). The use of lexical and syntactic information in language production: Evidence from the priming of noun–phrase structure. *Journal of Memory and Language, 49*, 214–230.
De Fornel, M. (1992). The return gesture. In P. Auer & A. di Luzio (Eds.), *The contextualization of language*. Amsterdam: John Benjamins.

De Ruiter, J. P. (1998). Gesture and speech production. Unpublished dissertation, University of Nijmegen.

De Ruiter, J. P. (2000). The production of gesture and speech. In D. McNeill (Ed.), *Language and gesture* (pp. 284–311). Cambridge: Cambridge University Press.

Effron, D. (1941). *Gesture and environment*. Morningside Heights, NY: King's Crown Press.

Ekman, P., & Friesen, W. V. (1969). The repertoire of nonverbal behavior: Categories, origins, usage, and coding. *Semiotica, 1*, 49–98.

Enfield, N. J., Kita, S., & De Ruiter, J. P. (2007). Primary and secondary pragmatic functions of pointing gestures. *Journal of Pragmatics, 39*, 1722–1741.

Feyereisen, P., Van de Wiele, M., & Dubois, F. (1988). The meaning of gestures: What can be understood without speech? *Cahiers de Psychologie Cognitive, 8*, 3–25.

Garrod, S., & Anderson, A. (1987). Saying what you mean in dialogue: A study in conceptual and semantic co-ordination. *Cognition, 27*, 181–218.

Holler, J., & Wilkin, K. (2011). Co-speech gesture mimicry in the process of collaborative referring during face-to-face dialogue. *Journal of Nonverbal Behavior, 35*, 133–153.

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.

Kelly, S. D., Özyürek, A., & Maris, E. (2010). Two sides of the same coin: Speech and gesture mutually interact to enhance comprehension. *Psychological Science, 21*, 260–267.

Kendon, A. (1988). How gestures can become like words. In F. Potyatos (Ed.), *Crosscultural perspectives in nonverbal communication* (pp. 131–141). Toronto, Canada: Hogrefe.

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

Kimbara, I. (2006). On gestural mimicry. *Gesture, 6*, 39–61.

Kimbara, I. (2008). Gesture form convergence in joint description. *Journal of Nonverbal Behavior, 32*, 123–131.

Kita, S., & Özyürek, A. (2003). What does cross-linguistic variation in semantic coordination of speech and gesture reveal? Evidence for an interface representation of spatial thinking and speaking. *Journal of Memory and Language, 47*, 16–32.

Kita, S., Özyürek, A., Allen, S., Brown, A., Furman, R., & Ishizuka, T. (2007). Relations between syntactic encoding and co-speech gestures: Implications for a model of speech and gesture production. *Language and Cognitive Processes, 22*, 1212–1236.

Krahmer, E., & Swerts, M. (2007). The effects of visual beats on prosodic prominence: Acoustic analyses, auditory perception and visual perception. *Journal of Memory and Language, 57*, 396–414.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*, 159–174.

Levelt, W. J. M. (1989). *Speaking*. Cambridge, MA: MIT Press.

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

McNeill, D. (2005). *Gesture and thought*. Chicago and London: University of Chicago Press.

Mol, L., Krahmer, E., Maes, A., & Swerts, M. (2009). The communicative import of gestures: Evidence from a comparative analysis of human–human and human–computer interactions. *Gesture, 9*, 97–126.

Müller, C. (1998). *Redebegleitende Gesten. Kulturgeschichte – Theorie – Sprachvergleich*. Berlin: Berlin Verlag.

Parrill, F., & Kimbara, I. (2006). Seeing and hearing double: The influence of mimicry in speech and gesture on observers. *Journal of Nonverbal Behavior, 30*, 157–166.

Pickering, M. J., & Branigan, H. P. (1998). The representation of verbs: Evidence from syntactic priming in language production. *Journal of Memory and Language, 39*, 633–651.

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences, 27*, 169–225.

Tabensky, A. (2001). Gesture and speech rephrasings in conversation. *Gesture, 1*, 213–235.

Van Baaren, R. B., Holland, R. W., Steenaert, B., & Van Knippenberg, A. (2003). Mimicry for money: Behavioral consequences of imitation. *Journal of Experimental Social Psychology, 39*, 393–398.