Converging Hands or Converging Minds?

Lisette Mol (L.Mol@uvt.nl)

Emiel Krahmer (E.J.Krahmer@uvt.nl)

Alfons Maes (Maes@uvt.nl)

Marc Swerts (M.G.J.Swerts@uvt.nl)
Tilburg Centre for Cognition and Communication (TiCC), School of Humanities, Tilburg University
P.O. Box 90135, NL-5000 LE Tilburg, The Netherlands

Abstract

Interlocutors sometimes repeat each other's representational hand gestures. We investigated if this is a case of direct mimicry of form, or whether perceiving a gesture gives rise to a semantic representation, which subsequently informs gesture production. For this we used an interactive route description task, in which a confederate's gestures indicated the route in either the vertical or the horizontal plane and either with one or four fingers extended as an index. We found that perceiving vertical gestures led to an increase not only in participants' production of vertical gestures, but also in their use of one finger as an index, suggesting that seeing vertical gestures caused participants to think of the route as on a map, which led them to point with one finger (as is common on a map) rather than four. Our results support the notion that repetition of meaningful gesture forms results from converging semantic representations.

Keywords: Gesture; Adaptation, Lexical Entrainment.

Introduction

It is well established that when people interact in dialogue, they tend to adapt to each other in many ways (for an overview, see Pickering & Garrod, 2004). For example, interlocutors reuse each other's (referring) expressions (e.g. Brennan & Clark, 1996) and syntactic constructions (e.g. Branigan, Pickering, & Cleland, 2000). In one study, Levelt and Kelter (1982) found that if shop keepers where asked in Dutch "(At) what time does your shop close?", their answer tended to match the question in surface form, either including or omitting 'at'. Similarly, repetitions of form across interlocutors have also been found for co-speech hand gestures (e.g. De Fornel, 1992). Such gestures are spontaneous movements of the hands and arms during speech, which can convey information, or emphasize certain parts of speech (e.g. McNeill, 1992). Elements of a gesture's physical form (*articulators*), like the shape and orientation of the hand, the direction and size of the movement, and where it is performed relative to the speaker can be repeated in subsequent gestures by the same or another speaker.

Some scholars believe that speech and gesture jointly express a speaker's ideas (McNeill, 1992), or that speech and gesture are both part of a speaker's communicative effort (Kendon, 2004). From this perspective, it seems likely that repetition of each other's gesture forms would resemble repetition of each other's (other) linguistic forms. On the other hand, repetitions in physical behavior are found in many species, and need not be tied to speech (Parrill & Kimbara, 2006). In this paper, we focus on gestures that depict some of the content a speaker is conveying, which are known as *illustrators* (Ekman & Friesen, 1969) or *representational gestures* (McNeill, 1992). We compare the repetition across speakers of certain articulators of such gestures to the repetition of meaningful units in speech, specifically *lexical entrainment,* as well as to non-linguistic forms of behavioral mimicry. We first explain a difference between direct behavioral mimicry and lexical entrainment. We then describe some empirical results on the repetition of gesture forms across speakers. This will lead to our research question: Is the repetition of meaningful gesture forms across interlocutors a consequence of converging semantic representations, or is there a more direct link between perceiving a form and producing a form?

Mimicry and Adaptation

Mimicry is defined as one person repeating the behavior of another person (Bock, 1986). Some forms of mimicry enable the transfer of important functional behaviors (Tomasello, Savage-Rumbaugh, & Kruger, 1993). It has also been found that repeating others can have social benefits. Van Baaren et al. (2003) found that a waitress received higher tips when repeating her customers' orders literally, than when signaling in some other way that she understood the order. Yet for some repetitions of behavior, the functional or social purpose is less clear (Chartrand & Bargh, 1999). For example, if one person starts yawning, oftentimes those around will start yawning as well. Chartrand and Bargh explain this type of behavior in terms of the *perception-behavior link*, meaning "the mere perception of another's behavior automatically increases the likelihood of engaging in that behavior oneself", p. 893. Notably, they state that although such mimicry may act as a kind of 'social glue', intent or conscious effort are not required for it to occur. We will subsequently use the term 'mimicry' to refer to such automated repetitions of behavior.

Pickering and Garrod (2004) propose that similar automatic priming underlies the repetition of linguistic

behaviors across interlocutors, a form of adaptation which they call *alignment*. They state that at each linguistic level, "the activation of a representation in one interlocutor leads to the activation of the matching representation in the other interlocutor directly", p. 177. These representations are thought to be used in both language production and processing (parity of representation). Thus, if a certain lexical or semantic representation has just been constructed as a result of hearing an utterance, it can subsequently be used for production. In addition to this direct source of alignment across interlocutors, alignment at one level can also enhance alignment at certain other levels within a speaker, because of bidirectional connections between the representations at different levels. Thus, if a lexical representation is connected to a semantic one, activation of that semantic representation may subsequently activate the lexical one.

Let us focus on one particular case of converging linguistic behavior: the repetition of referring expressions across speakers, known as *lexical entrainment* (Brennan & Clark, 1996). Brennan and Clark propose that interlocutors use the same words to refer to the same objects, because they use similar conceptualizations of that object. For example, suppose a particular object could be thought of as a document, a picture, or a map. When a speaker refers to it with 'the 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 map' as a reference to both the object and the particular conceptualization of it. In this view, the repetition of references across interlocutors results from the establishment of conceptual pacts.

In both the model by Pickering and Garrod and the model by Brennan and Clark, lexical representations and semantic representations are linked. This is where lexical entrainment seems to differ from some instances of direct behavioral mimicry. In mimicry, we may repeat forms without being aware of their meaning. In other words, the perception of a form directly leads to the production of that form. In lexical entrainment on the other hand, there seems to be an intermediate stage: meaning. A form that is perceived is coupled with a certain meaning. Only when that meaning is activated again is the same form a likely candidate for repetition.

Repetition of Gesture Form

Is meaning also involved in the repetition of gesture forms across interlocutors? Kimbara (2008) observed interlocutors while they were discussing an animated cartoon. She found that their gestures looked more similar if they could see each other, compared to when they were separated by an opaque screen. Thus, it seems that adaptation occurs in gesture like it does in speech. Yet if these similarities in gesture form resulted from similarities in semantic representations, one could argue that they would also occur when interlocutors cannot see each other, since similarities in semantic representations can also be arrived at through

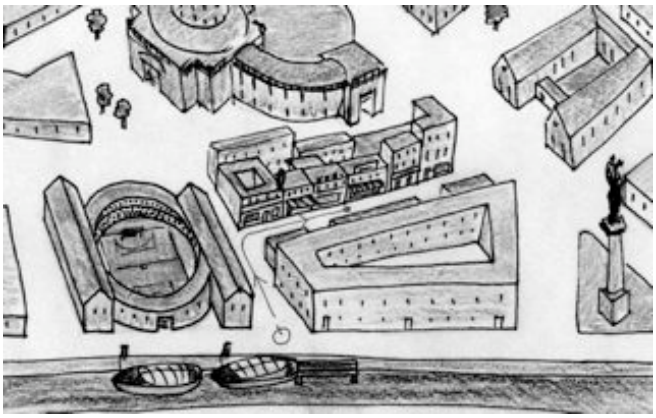

Figure 1: Part of a scene used in the experiment, note the route starting at the bottom-center.

speech. Therefore, it may be the case that the repetitions of gesture form resulted from the automatic across-speaker path of alignment (direct copying of form), rather than from connections between gesture forms and semantic representations, or the establishment of conceptual pacts.

In a previous study, we have investigated how relevant the semantic context was for certain gestures to be repeated across interlocutors (Mol, Krahmer, & Swerts, 2009). Gestures were either performed with speech matching the gesture's form in meaning (e.g. a speaker moved his arms as though running, while talking about running), or with speech that expressed a very different meaning (e.g. the same gesture performed while describing looking through binoculars). We found that repetition did not occur when a gesture was shown in a non-matching semantic context. This suggests that repetition of form in gesture may result from the coupling of a certain form to a certain meaning, rather than from direct copying of form. However, since the mismatches were designed to be very clear in this study, the non-matching gestures may not have been processed very deeply to begin with, or participants may simply be less likely to adapt to a less coherent speaker. We thus need to investigate further whether repetition of gesture forms across interlocutors results from converging meanings.

Present Study

In this study 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 (also see Cassell, McNeill, & McCullough, 1998). Suppose that certain articulators of a perceived gesture give rise to the construction of meaning. Then when this meaning is subsequently expressed in gesture, all articulators of the gesture produced will likely be consistent with this meaning. Therefore, we would expect that articulators of the perceived gesture that are inconsistent with the meaning constructed would not be repeated as frequently. On the other hand, if repetition of gesture form happens without the

semantic level being involved, any combination of perceived articulators could be repeated in gesture production. In this case, whether an articulator matches the constructed meaning will not affect how frequently it is repeated.

To test this we use a task in which a confederate and a participant give each other route directions repeatedly. We present participants with bird's view drawings of a city scene, with a short route indicated on them (see Figure 1 for an example). These scenes are neither presented vertically nor horizontally, but at an angle. Therefore, the production task can be thought of as either describing a route on a vertically oriented map, *or* as describing a route through an actual (horizontal) city.

In each condition, the confederate expresses only one of these conceptualizations in her gesturing. While speech is kept constant, she gestures either as though moving along a route in a horizontal city, or as though indicating the route on a vertical map. This is done using two articulators: the plane in which the gesture is produced (either horizontal or vertical) and hand orientation (with fingers moving along with the route, or pointing forward as though on a map).

It is interesting in itself to see whether participants adapt to the confederate's perspective in their gesturing. Yet this alone would not tell us whether this is based on direct mimicry of form, or on the convergence of semantic representations. Therefore, we manipulate a third articulator (hand shape) independently. Gestures are produced either with one finger, or four fingers extended as an index. Now if it is the case that gesture form is perceived and reproduced directly, without mediation of meaning, both the confederate's perspective and her hand shape could be repeated independently by participants in their own gesturing. There may be a difference in how frequently each aspect is repeated, but what we would not expect based on this view, is for the confederate's perspective to influence participants' hand shapes or for the confederate's hand shape to influence participants' perspective.

On the other hand, if meaning does form an intermediate stage between the perception and production of gesture forms, we would expect such cross-effects to occur. For example, 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 perspective would lead participants to think of the communication task as describing a route on a map, then their gestures may be produced more frequently with only one finger as an index (even if the confederate uses more than one finger). This would mean there is an effect of the confederate's gestures' perspective on the hand shape of participants' gestures. This effect may also be found in the opposite direction: the confederate's use of more fingers as an index may lead participants to think of the route as through an actual city rather than on a map, causing them to gesture horizontally more frequently.

Figure 2: Partial view of the experimental set-up. Participants were seated on the right.

Method

Participants

48 Native Dutch speakers, all students of Tilburg University took part in this study. The data of eight participants could not be used for analysis (six participants did not produce any relevant gestures). The remaining 40 participants (33 female) had a mean age of 20.5, range 18 – 25.

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 the communication task, and stated that the couple with 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 there was a flip chart for each of them. 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, see Figure 2. Both behind the confederate and 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 Figure 1 for an example. She then turned the page of her flip chart 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 always followed the same script. Gestures were timed naturally with speech and gazed at by the confederate.

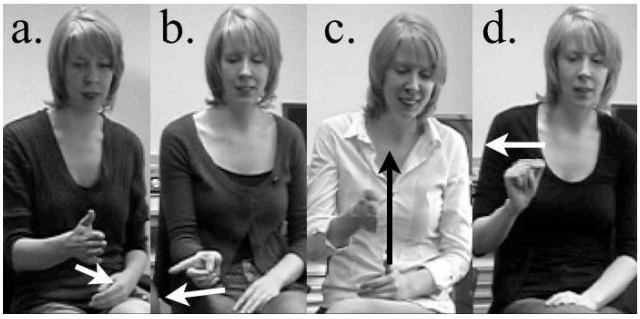

Figure 3: The confederate's path gestures. a: Hand/ Route; b: Finger/ Route; c: Hand/ Map; d: Finger/ Map.

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. 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.

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, as well as some questions on how they liked their interaction partner. It ended with the question whether the participant was left or right handed. When the participant was done filling out the forms, the confederate revealed herself and asked the participant's consent for the use of the data. Participants were also asked if they had suspected any deception. Two participants were excluded from our analysis, because they indicated having been suspicious about either the goal of the experiment or the role of the confederate.

Design, Coding and Analysis

We used a 2 x 2 between subjects design. The independent variables were the hand shape (one or four fingers extended) and perspective (route or map) of the confederate's path gestures. In the route perspective, gestures were performed with the index in the direction of the hand movement and movement was in the horizontal plane, as though following a virtual route (Figure 3a, 3b). In the map perspective, gestures were performed in the vertical plane and the index was always pointing forward, as though pointing on a virtual map (Figure 3c, 3d).

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.

Figure 4: Examples of participants' path gestures (published with permission of the people depicted). a: Hand/ Vertical Map; b: Finger/ Vertical Map; c: Hand/ Route; d: Finger/ Horizontal Map.

We coded all *path gestures* 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 three articulators: location in the gesture space, hand orientation, and movement (direction and size). It turned out that in addition to the two perspectives the confederate had used, participants occasionally used an alternative one, as though pointing on a horizontal map. Therefore, we chose from three labels: *Route, Vertical Map*, and *Horizontal Map.* A gesture in the route perspective would typically have horizontal movement in front of and to the side of the speaker, with the fingers pointing in the direction of the hand movement (Figure 4c). Vertical Map gestures on the other hand would typically have vertical movement, with relative sizes mapping onto distances on the map, fingers pointing forward and the location in the gestures space corresponding to the location on the map (Figure 4a, 4b). Horizontal Map gestures (Figure 4d) differ from Route gestures in their hand orientation (fingers pointing down), and their relative size and location. The label that could explain most of the articulators was assigned to each gesture. Figure 4 shows some examples of participants' path gestures and our coding.

Each of the confederate's verbal descriptions contained two target landmarks, which also appeared along the participant's route. For each of these landmarks it was

determined whether the participant referred to it, and if so whether it was a *literal repetition*, an elaboration or shortening of the confederate's reference (both counted as *partial match*) or a complete *mismatch*. For example, if the confederate said "hoog gebouw" (tall building), "gebouw" (building) and "hoog grijs gebouw" (tall grey building) would be labeled as *partial match* whereas "flat" (apartment building) would be a *mismatch*.

The data of 6 participants were excluded because these participants did not produce any path gestures. This left 40 participants, 10 in each cell. Analysis was done using ANOVA, with factors perspective (levels: Route & Map) and hand shape (levels: Finger & Hand) of the confederate's gestures. The significance threshold was .05 and we used partial eta squared as a measure of effect size.

Results

Neither the confederate's hand shape nor her perspective significantly influenced the total number of path gestures participants produced (M = 5.5, SD = 3.9) and there was no interaction between these factors. We did not find a significant effect of gender, or left or right handedness on the amount or type of path gestures produced. Analysis of the answers to the questionnaire showed no significant effect of condition on how the participants perceived the confederate.

Verbal Alignment

Neither the confederate's perspective (p = .63), nor her hand shape (p = .81) had a significant effect on the number of target nouns repeated by participants (M = 6.2, SD = 1.3), or on the number of partial matches or mismatches.

Effects of the Confederate's Perspective

The confederate's perspective influenced participants' perspective. When the confederate gestured as though on a map, the mean proportion of participants' path gestures in the vertical map perspective was higher (M = .46, SD = .35) than when she gestured as though following a route (M = .11, SD = .20), $F(1, 36) = 14.88$, $p < .001$, $\eta^2 = .29$. Similarly, when the confederate gestured as though following a route, participants produced a higher proportion of gestures with the route perspective (M = .77, SD = .32) than when she gestured as though on a map (M =.52, SD = .39), $F(1, 36) = 12.35$, $p < .001$, $\eta^2 = .14$, see Table 1.

The confederate's gestures' perspective also influenced the hand shape used by participants, $F(1, 36) = 5.00$, $p < .05$, $\eta^2 = .12$. The proportion of gestures with more than one finger extended was higher when the confederate used the route perspective (M = .78, SD = .37) than when she used the map perspective (M = .52, SD = .39), whereas the proportion of gestures with one finger extended was higher when she used the map perspective (M = .48, SD = .39), compared to when she used the route perspective (M = .22, SD = .37), see Table 2.

Table 1: *Means* and *standard deviations* of the proportion of path gestures participants produced from each perspective.

Confederate's Perspective	Prop. Route	Prop. Vertical Map	Prop. Hor. Map
Route	0.77 (.32)	0.11 (.20)	0.12 (.26)
Map	0.43 (.31)	0.46 (.35)	0.11 (.25)

Table 2: *Means* and *standard deviations* of the proportion of path gestures participants produced with each hand shape.

Confederate's Perspective	Prop. Hand	Prop. Finger
Route	.78 (.37)	.22 (.37)
Map	.52 (.39)	.48 (.39)

Effects of the Confederate's Hand Shape

We did not find that the confederate's hand shape influenced participants' hand shape $F(1, 36) = .04$, p = .85, nor that her hand shape influenced the proportion of gestures in the map, $F(1, 36) = .38$, p = .54, or route perspective, $F(1, 36) = .030$, p = .86.

Discussion

We found some of the cross-effects we expected if perceiving gestures would lead to the construction of meaning, which in turn would influence 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 by the confederate's vertical gestures leading participants to think of the route as on a map, which caused them to point with their finger more frequently.

Gestures, like speech, seem to allow for the convergence of representations of meaning across interlocutors. This leads to the question of whether the same representations underlie adaptation in both gesture and speech, and whether these representations can also be influenced by both gesture and speech.

Adaptation in Gesture through Speech

The results of an additional study indeed point in this direction. In this study, the confederate gestured with one finger extended and in the map perspective. Thus, all articulators in gesture suggested a vertically oriented map. Yet we added a condition (N = 10) to the previous study, in which speech also matched this perspective. Rather than using horizontal terms like "rechtdoor" (straight), the confederate now used vertical terms like "naar boven" (up) instead. Note that the first direction was always straight/ up.

When comparing this condition to the Finger/ Map condition with horizontal speech, we found that the perspective of the confederate's speech had an additional effect on the perspective of participants' gestures. With

vertical terms, participants produced a lower rate of gestures with the route perspective (M = .17, SD = .25*)* than with horizontal terms (M = .54, SD = .36), $F(1, 18) = 7.21$, $p < .02$, $\eta^2 = .29$. This supports the notion that semantic representations were converging across interlocutors, rather than surface forms. In addition, it suggests that these representations may be shared between speech and gesture.

Future Work

A limitation of our studies is that the confederate always acted according to a script, and thus was not exactly like a spontaneous partner in forming a conceptual pact. Whereas this usually can be thought of as an interactive process between both interlocutors, the confederate always stuck to her own initial proposal. It would be interesting to see how spontaneous interaction is similar to or differs from this partly staged interaction.

Overall, perspective was repeated more than hand shape. This may be because perspective was expressed in two articulators, whereas hand shape is only one. Thus, the one articulator not matching the constructed meaning may have been adapted to the two matching ones. Another explanation would be that in this task, perspective carried a more important meaning than did hand shape. A vertical gesture cannot possibly depict a route one can walk (at least not in the Netherlands), whereas the distinctions between the different hand shapes are probably far subtler. Therefore, the perspective of gestures may have influenced the construction of meaning more readily than their hand shape. Apparently, in this task, hand shape was not a likely candidate for the type of direct alignment at one level between interlocutors that Pickering and Garrod (2004) proposed. However, in other settings it may very well be. It would be interesting to investigate whether adaptation in hand shape depends on to what extent hand shape carries meaning, and similarly for other articulators. Additionally, other types of gestures (especially non representational gestures) need to be looked at, since they may carry meaning in a way different from the path gestures we studied, and may or may not be linked to semantic representations.

Conclusion

In the adaptation of one interlocutor to another, some hand gestures seem to behave truly like linguistic forms. Whether they are repeated across interlocutors depends on whether their form corresponds to the semantic context (Mol et al., 2009), and the repetition of forms is mediated by mental representations of meaning, rather than being based on a direct perception-action link.

Acknowledgements

We like to thank Susan Brennan and Sotaro Kita for valuable discussions on this work. We also thank Nathalie Bastiaansen for drawing the stimuli, our co-workers at Tilburg University for assisting in the data collection, and the anonymous reviewers.

References

Bock, J. (1986). Syntactic persistence in language production. *Cognitive Psychology, 18*, 355-387.

Branigan, H. P., Pickering, M. J., & Cleland, A. A. (2000). Syntactic co-ordination in dialogue. *Cognition, 75*(2), B13-B25.

Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology-Learning Memory and Cognition, 22*(6), 1482-1493.

Cassell, J., McNeill, D., & McCullough, K.-E. (1998). Speech-gesture mismatches: Evidence for one underlying representation of linguistic & nonlinguistic information. *Pragmatics & Cognition, 6*(2), 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*(6), 893-910.

De Fornel, M. (1992). The return gesture. . In P. Auer & A. di Luzio (Eds.), *The contextualization of language*. Amsterdam: John Benjamins.

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

Kendon, A. (2004). *Gesture: Visible Action as Utterance*. Cambridge: Cambridge University Press.

Kimbara, I. (2008). Gesture form convergence in joint description. *Journal of Nonverbal Behavior, 32*(2), 123-131.

Levelt, W. J. M., & Kelter, S. (1982). Surface form and memory in question answering. *Cognitive Psychology, 14*(1), 78-106.

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

Mol, L., Krahmer, E., & Swerts, M. (2009). *Alignment in iconic gestures: Does it make sense?* Paper presented at the The eight international conference on auditory-visual speech processing, Norwich, United Kingdom.

Parrill, F., & Kimbara, I. (2006). Seeing and hearing double: The influence of mimicry in speech and gesture on observerse. *Journal of Nonverbal Behavior, 30*(4), 157-166.

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences, 27*(2), 169-225.

Tomasello, M., Savage-Rumbaugh, S., & Kruger, A. (1993). Imitative learning of actions on objects by children, chimpanzees and enculturated chimpanzees. *Child Development, 64*, 1688-1705.

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*(4), 393-398.