CHAPTER 4

The asymmetric redundancy of gesture and speech

J. P. de Ruiter
Tufts University

A number of studies from the last decades have demonstrated that the iconic gestures are shaped not only by our mental imagery but also, quite strongly, by structural properties of the accompanying speech. These findings are problematic for the central assumption in the Sketch Model (De Ruiter, 2000) about the function of representational gesture. I suggest a seemingly small but fundamental modification to the processing assumptions in the Sketch Model that not only accommodates the discussed empirical findings, but also explains many other well-known gesture phenomena. The new model also generates new and testable predictions regarding the relationship between gesture and speech.

Introduction

Over the last half-century we have accumulated a large and growing body of knowledge about the gestures that usually accompany spontaneous speech. Nevertheless gesture, and especially the class of gestures that Schegloff (1984) and McNeill (1985) have called "iconic gesture", remains a mysterious phenomenon. What information do iconic gestures express, how are they shaped the way they are, how are they related to the accompanying speech, and most intriguingly, what is their function?

Early foundational work on speech-accompanying gesture was mostly based on careful qualitative analysis of video recordings of spontaneous narrative or dialogue (Kendon, 1972, 1980; McNeill, 1985, 1992). As the field was growing, more and more studies also employed psycholinguistic methods, testing specific hypotheses in experimental and quasi-experimental designs. While observational analysis and experimentation are both essential for understanding a cognitive phenomenon, at some point it becomes beneficial to formulate comprehensive models about the underlying cognitive architecture, or "model", for short. Observation

doi 10.1075/gs.7.04der

can tell us what happens. Experimentation can put hypothesized causal explanations to the test. But only formulating models enables us to study whether a coherent *ensemble* of causal explanations that makes up a model adequately explains the observed phenomena, and generates predictions that can be tested against observed phenomena and experimental results.

In what follows, I will briefly describe the assumptions of the Sketch Model (De Ruiter, 2000) regarding the function of iconic gesture. I will then proceed to discuss some empirical findings that put this assumption into question, leading to the alternative proposal that the function of gesture is to enhance the speech signal by providing additional and redundant spatial information. I will propose a modification to the Sketch Model that incorporates this new assumption, and restores its compatibility with the discussed empirical findings. Finally, I discuss the novel explanations the modified Sketch provides for other well-known phenomena regarding iconic gesture and its relationship to speech.

The Sketch Model

The Sketch model (De Ruiter, 2000) was based on the foundations of Levelt's (1989) *blueprint for the speaker* and conceptually inspired by Growth Point theory (McNeill, 1992; McNeill & Duncan, 2000). It is a processing architecture that incorporates *iconic gestures* (which from the point of view of the Sketch Model includes McNeill's *metaphoric gestures*), *pantomimes*, *deictics*, and *emblems*, i.e., all gesture types from McNeill's (1992) typology except *beats*. Its central assumption is that the Leveltian "Conceptualizer", which is responsible for converting a communicative intention into a semantic structure called "Preverbal Message", is also generating an abstract "Sketch" that contains imagistic information from the communicative intention that is easier to express in the analog gestural modality. Both the Preverbal Message and the Sketch are then converted into a verbal utterance and a gesture, respectively, by the formulator and the gesture planner, after which a synchronization mechanism involving simple control signals ensures that the gesture and the speech are produced in temporal synchrony.

For speech that is accompanied by iconic gestures, information that is part of the communicative intention of the speaker is explicitly divided by the conceptualizer into a semantic representation that corresponds to the verbal part of the communicative intention, and a gesture part that is based on imagistic information from short-term memory. The assumption is that the gestural and the verbal part of the utterance are used to express *different* types of information originating from the *same* communicative intention. So the information that is in the communicative intention is split into an imagistic and a propositional part.

This is like the two sides of a postcard, with one side containing a picture, and the other side the text (De Ruiter, 2007).

The Sketch model has been employed in different areas of gesture research. It has been used as a basis for an implemented model for gesture and speech generation (Kopp, Tepper, & Cassell, 2004). It also has been used in the context of aphasia research (Rose, 2006, 2013). Its synchronization mechanism has been tested and confirmed in a research project addressing gesture during dysfluent speech (Seyfeddinipur, 2006). However, it has also been criticized, most notably by Kita & Özyürek (2003), for not being able to incorporate findings that show that the nature and shape of iconic gestures are strongly influenced by structural properties of the language that they accompany. In addition to the findings by Kita & Özyürek (2003), there are several other recent studies that also show a more intricate connection between gesture and speech than the Sketch Model seems to be able to incorporate. In the next section, I will discuss this evidence about the effects of language on iconic gesture in more detail.

Effects of language on iconic gesture

The studies summarized below all provide evidence that the effect of the nature of the produced speech on the accompanying gesture is much stronger than the other way around.

Gestures are influenced by lexical and syntactic properties of language

As is customary in research on iconic gesture, Kita & Özyürek (2003) had speakers of English, Turkish, and Japanese retell episodes from the classic Warner Brothers cartoon "Tweety and Sylvester". One of the scenes that participants are required to narrate features Sylvester the cat using a rope to swing across the street from one building to another in a failure-doomed attempt to catch Tweetybird. The authors note that neither Japanese nor Turkish have a verb that is equivalent to the verb "to swing" in English. They found that in describing this scene, nearly all of their English speakers produced only arc-like gestures, but the Japanese and Turkish speakers either used only a straight gesture, or both a straight and an arc-shaped gesture. While their conclusion that the language appears to have *some* influence on the gesture is supported by this finding, the fact that the Turkish and Japanese speakers also produced arc-like gestures (either combined with straight ones or not) makes it difficult to assess how strong that influence is, and how it operates. Their most interpretable finding is that English speakers rarely produced any straight gestures, suggesting that the availability of the "swing" verb in English appeared to have forced the shape of the gesture to be arc-like. This would suggest that the effect of language on gesture is actually a lexical effect, and that speakers of Turkish or Japanese, not experiencing the "constraints" on the gesture provided by the verb "swing" are free to either include or exclude the arc-trajectory from their gesture.

Equally intriguing is their second cross-linguistic finding, also based on the difference between English on the one hand, and Turkish and Japanese on the other. In describing Sylvester who swallows a bowling ball and moves down the hill in a rolling-like fashion, English allows for coding *manner* in the verb itself, and *path* in a prepositional phrase, as in the description "he rolls down the hill", needing only one clause to express this meaning. Japanese and Turkish, on the other hand, need separate verb clauses for expressing both manner and path. So it is very interesting to see what type of gesture(s) accompany the narration of the "rolling down the hill" scene in these languages. Specifically, it is interesting whether the one vs. two clause descriptions enforced by the structure of the language are accompanied by a corresponding number of gestures. The results of this analysis were similar to those of the swing gestures above. The majority of English speakers *only* produced a manner-path conflating gesture, which mirrors the semantic structure of the unmarked way of describing the rolling event in their language ("rolls down the street"). However, the majority of Turkish and Japanese speakers also produced gestures that displayed only the manner or the trajectory (path), which corresponds to the two-clause formulation that is common in those languages.

These findings strongly suggest that the structure of the language that is spoken has a dominant influence on the nature of the accompanying iconic gestures. As Kita & Özyürek (2003) point out, this would contradict the assumption they call the "Free Imagery Hypothesis", the idea that the information in iconic gesture is not influenced by the accompanying language, an assumption that is adopted in the Sketch Model. While I agree with their diagnosis regarding the Sketch Model, I do not believe that their proposed Interface Model is the appropriate cure. I will discuss this in more detail below. Also noteworthy is that there seems to be an *asymmetric* influence of language structure on iconic gesture. The evidence does not support the assumption that the "structure" of iconic gesture influences speech. This is also a priori unlikely because iconic gestures are not lexicalized: contrary to the syntax and semantics of spoken languages, their form-meaning mapping is not conventionalized within a language community.

Gesture and speech go hand in hand

So, Kita, & Goldin-Meadow (2009) performed a direct test of the assumption that gesture compensates for information that is not in the speech (Bangerter, 2004; De Ruiter, 2000; Van der Sluis & Krahmer, 2007). They had participants tell stories based on a "vignette", in which either two male protagonists or a male and a female protagonist were featuring. They then coded how often the participants referred to these protagonists either by using noun phrases or pronouns, and whether they performed "abstract deictic" gestures to refer to them.[1] They found that storytellers were much more likely to produce abstract deictic gestures to these protagonists when they (the storytellers) also uniquely identified them in speech. More importantly, "gesture was rarely used to compensate for the absence of lexical specificity in pronouns or nouns" (p. 120). This contradicts the assumption that gesture compensates for difficulties in speech in the Sketch Model, also called the "Mutually Adaptive Modalities" hypothesis in De Ruiter (2006). Again, it seems as if the content of the speech is playing a dominant role in shaping the gesture.

Gesture and speech express similar types of information

De Ruiter, Bangerter, & Dings (2012) set out to test one half of what they called the "Tradeoff Hypothesis", which is the same as the "Mutually Adaptive Modalities hypothesis" from De Ruiter (2006). There is evidence from Bangerter (2003) and Van der Sluis & Krahmer (2007) that difficulties in gesture have an effect on the nature of the produced speech. De Ruiter, Bangerter & Dings (2012) set out to study the other half of this hypothesized tradeoff, namely that difficulty to express something in speech would increase the likelihood of gesture. They used a director/matcher experiment with tangram figures on a poster as target referents, and systematically manipulated repetition (how often had the target been referred to before by the participant) and codability (how easy it is to come up with an adequate verbal description for that target). Neither factor had any influence on the gesture-per-word rate. More importantly for the purpose of this discussion, additional correlational analyses revealed that the type of gesture covaried reliably with the type of information that was simultaneously revealed in speech. Feature descriptions were reliably accompanied by iconic gestures, and locative expressions by pointing gestures. The authors concluded that this is additional evidence for the "gesture and speech go hand in hand" hypothesis (So et al., 2009)

1. Abstract deictic gestures are pointing at locations in space that are associated with discourse entities.

and incompatible with the Tradeoff Hypothesis (Bangerter, 2004; De Ruiter, 2000, 2006; De Ruiter et al., 2012; Van der Sluis & Krahmer, 2007). Again, this is evidence for asymmetric "dominance" of speech over gesture.

Gestures adapt to the pragmatic aspects of speech

A final study revealing how gesture follows the lead of speech is also a director/matcher experiment by Bangerter, De Ruiter, Mayor, & Chevalley (2010) in which participants had to single out referents on a poster on the wall. Participants who were producing elaborate referring expressions often "interrupted themselves" to produce informative side sequences, after which they resumed the original utterance. This happened at multiple levels of recursion, i.e., it also involved side sequences within side sequences. Interestingly, if these participants were also gesturing during their referring, they interrupted their own (usually deictic) gesture in mid-air to produce gestures that belonged to the side sequences, and continued their original gesture when the original embedding verbal utterance was resumed. This shows that gesture follows speech at the pragmatic level as well.

Implications for the function of iconic gesture

The empirical findings summarized above are incompatible with the assumption from the Sketch Model that the conceptualizer splits the communicative intention of the speaker into complementary propositional/verbal and imagistic/gestural representations on the basis of the expressive possibilities of the respective modalities. In fact, as Kita & Özyürek (2003) conclude, these findings provide strong evidence that this assumption is wrong.

In fact, the findings summarized above all seem to indicate a strong influence of the nature of the speech on the information in the gesture. These data force us to look at a hitherto rarely adopted assumption in the cognitive modeling of gesture,[2] namely the idea that the iconic gesture is 'dominated' by the content of the speech, providing additional imagistic information to the utterance. This assumption corresponds to what Kita & Özyürek (2003) discuss as the "Lexical Semantics Hypothesis", proposed by Schegloff (1984) and by Butterworth & Hadar (1989). This hypothesis claims that iconic gestures are linked to specific lexical entries, i.e., words. While this may often appear to be the case, it is unlikely to be

2. A possible exception is the implementation of the artificial real estate agent *Rea* in the model by Cassell et al. (1999).

the correct assumption in general, as there are many examples of gesture-speech utterances where the gesture is related to higher-level representations, or gestures related to entities that aren't even mentioned explicitly in the speech (see De Ruiter, 2000, p. 301, for a detailed example). Also, the Lexical Semantic Hypothesis cannot explain why some of the Japanese and Turkish gesturers from Kita & Özyürek (2003) made arc-like gestures without even having the lexical concept "swing" available in their language. So the Lexical Semantics Hypothesis is clearly, for lack of a better way to put it, too lexical.

But the idea that gesture is driven by language is not dependent on the affiliate being a single lexical entry. It could also be driven by a higher-level structure related to the planning unit that the conceptualizer uses. This could be a single word, but it could also be a clause, an idea-unit (Kendon, 1972, 1980, 2004; McNeill, 1992) or in conversation-analytic terms, a turn-constructional unit (Sacks, Schegloff, & Jefferson, 1974). So the hypothesis that I propose here in response to the findings discussed above is the *Asymmetric Redundancy Hypothesis* (ARH). The ARH claims that the information expressed in an iconic gesture originates from the same communicative intention as the verbal part of an utterance does, and is shaped so as to be maximally redundant with that communicative intention. At first sight, the ARH seem at odds with the often-made observation that iconic gestures always contain more information than the affiliated speech. To take an example from the study by Kita & Özyürek (2003) discussed above, if an English speaker wants to produce an iconic gesture corresponding with the English phrase "the cat swings across the street", there will be directional information encoded in the gesture that is not in the speech. However, as Kita & Özyürek (2003) themselves note in their discussion, with this gesture, this cannot be avoided, hence this information comes "for free" (p. 28–29). It is not possible to depict a swinging event without revealing a certain direction, a certain curvature, and a certain speed, just as it is impossible to gesture about the roundness of a cake without giving away information about its size. The suggestion here is that this is an unavoidable (but communicative) side effect of the gesture.

Below I want to address the question of how to accommodate these findings in a cognitive architecture for gesture and speech production.

The Interface Model

Kita & Özyürek (2003) have suggested a type of model to accommodate these results that differs from both the Free Imagery Hypothesis and the Lexical Semantics Hypothesis discussed above: their *Interface Model*. In this model, the shape of the iconic gesture is determined on the basis of the communicative intention in a process of on-line interaction between an Action Generator, and the Formulator. The Action Generator is a module that uses action schemata as a basis for generating gestures. This on-line interaction between Action Generator and Formulator is mediated by a module called "Message Generator." (A metaphor that comes to mind is a committee with the Formulator and the Action Generator as ordinary members, and the Message Generator as a chairman). In this way, the iconic gesture is the end product of a fusion of three sources of information: the communicative intention, the nature of the original imagery, and the feedback from the Formulator regarding the lexico-syntactic structure the speech is planned to have. This ensures that all these three elements are represented in the gesture, and that potential conflicts are resolved in the form of a compromise.

This type of architecture would indeed in principle be able to account for the findings mentioned above, and could potentially accommodate language effects on gesture that the Sketch Model can't. But there is a price to be paid. First of all, it is hard to imagine any type of iconic gesture that could *not* be generated by this process, because the gesture generation has access to all possible sources of information, and the principles and priorities of the negotiation process are not specified. In other words: the Interface Model does not constrain the relation between iconic gesture and speech in any way. This makes it very difficult to potentially falsify the model empirically. Second, the solution is far from parsimonious. The suggested architecture requires several additional assumptions. One far-reaching assumption that this architecture requires is that the negotiation mechanism has to have a way to estimate the degree of comparative "match" between the planned utterance of the Formulator and the planned gesture suggested by the Action Generator. This comparison would either involve translating the planned utterance back into a comparable imagistic format, or translating the imagistic information into linguistic format, or both. This requires a repeated production/comprehension loop for planned gestures and/or planned speech, which seems implausibly resource demanding.[3] It also implies that the process would take much time, which is hard to reconcile with the speed and smoothness of the gesture/speech production process. There is no empirical evidence that gesturing slows down speech. On the contrary, it has even been argued that performing iconic gestures can increase fluency (see e.g., Morrel-Samuels & Krauss, 1992).

3. Even though this is admittedly not an empirical argument, it should also be noted that this solution is very difficult to implement in a working system (Stefan Kopp, *personal communication*).

The Sketch Model

As mentioned above, the primary assumption that the Sketch Model makes regarding the representational function of gestures is not compatible with the empirical findings that lead to the formulation of the ARH. To incorporate these findings the role of the Conceptualizer in the model would have to be modified. There is a relatively straightforward modification that would not only incorporate the ARH but also, as a spin-off, suggest natural explanations of other phenomena as well (to be discussed below). For convenience, I will call the revised Sketch model the "Asymmetric Redundancy Sketch Model", or AR-Sketch Model for short. The proposed modification is as follows.

From all the representations in short-term memory accessible to the speaker, denoted by the letter W (for "Working Memory"), a subset $C \subset W$ is selected to be part of the communicative intention.[4] This subset C contains both propositional representations, denoted as C_p, and imagistic representations, which we will denote C_i. The Conceptualizer of the AR-Sketch Model creates a preverbal message for the Formulator on the basis of C (which can also be represented as the union $C_p \cup C_i$). While the preverbal message is generated on the basis of the entire set C (meaning *both* C_p *and* C_i) the Sketch Generator takes only the imagistic information in C_i to produce an abstract spatial-dynamic sketch for subsequent processing by the Gesture Planner and execution by the motor system, as in the original Sketch Model. Importantly, the imagistic information in C_i does not come directly from the imagistic representations in W(orking memory), but is selected by the Conceptualizer (in the process that Levelt called "Macro Planning") to be expressed in the *composite utterance* (Engle, 1998) containing both gesture and speech. The Gesture Planner cannot display the elements of C_i just like they were in the original stimulus, but has to perform explicit *recipient/audience design* (Clark & Murphy, 1983; Sacks & Schegloff, 1979), as there is always much more information in the imagistic information in the original stimulus than can be represented in a gesture (see De Ruiter 2007 for a more detailed treatment of recipient design in gesture production). So there is an asymmetric redundancy between gesture and speech, in the sense that the preverbal message is generated from the union $C_p \cup C_i$ while the gesture is solely generated from C_i, which is a proper subset of C. See Figures 1 and 2 for a diagram illustrating the difference between the original Sketch Model (Figure 1) and the AR-Sketch Model (Figure 2).

4. This is not to claim that the communicative intention is exclusively made up from imagistic and propositional (declarative) representations from working memory. Most likely, communicative intentions involve other aspects that are not represented in perceptual or declarative memory, such as the *illocution* of the utterance.

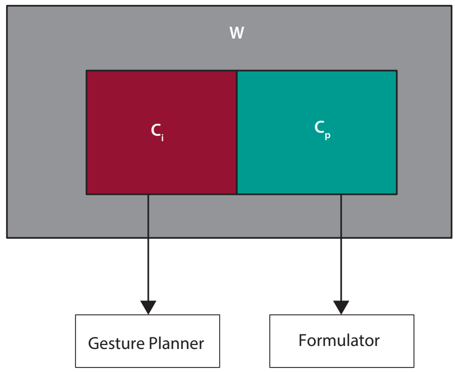

Figure 1. The original Sketch Model

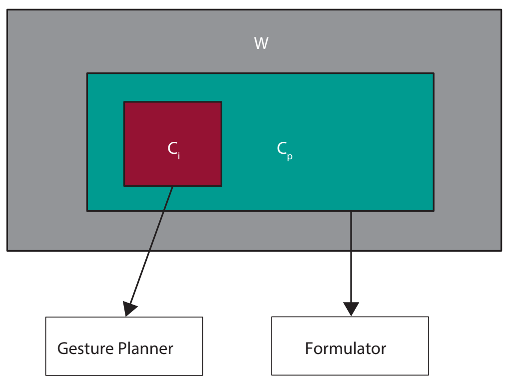

Figure 2. The AR-Sketch Model

One additional contentious assumption that needs to be made is that the conceptualizer "knows" the expressive means that are available in "its" language. In the successor of Levelt's (1989) blueprint for the speaker, the model presented in Levelt, Roelofs, and Meyer (1999), this assumption is in fact made. In their new model of speech production, it is assumed that there is a one to one correspondence (mapping) between concepts and words. If the concepts available to the Conceptualizer are identical to the lexical entries available to the Formulator, the Conceptualizer by implication has access to the available language-specific resources. This fundamental change in the original model in Levelt (1989) was mainly motivated by a solution of the so-called "hypernym problem" in speech production (Roelofs, 1992), and this solution is not uncontroversial. This will not be discussed here; the interested reader is referred to the commentaries in Levelt et al. (1999), and to Bierwisch and Schreuder (1992) for a different view on the relationship between concepts and words.

Because the AR-Sketch Model is a process model that implements the ARH, it can explain the swing scene described above by assuming that an English speaker who has access to the concept "swing" will naturally combine the word "swing" (or a verb phrase with "swing" as head) with an arc-like gesture. A speaker of Japanese will use one of the available motion verbs with a straight gesture that corresponds to the "moving from a to b" meaning and possibly use a arc-shaped gesture together with a phrase explaining the special way that the cat employs to propel itself to the other side, for instance using the adjectival phrase "like Tarzan" (Kita, *personal communication*). Kita & Özyürek (2003) unfortunately do not provide details regarding the affiliated speech that accompanies the respective straight and arc-like gestures that are made by Japanese and Turkish speakers, but the prediction of the AR-Sketch Model is that the straight gesture will tend to accompany the "move from a to b" concept in the speech, and the arc-gesture the adjectival phrase (e.g., "like Tarzan").

A similar explanation would apply to the "rolls down the hill" description. Only the English speakers have a clause at their disposal that combines manner and trajectory, so they will plan the combined gesture. The Japanese and Turkish speakers will need two clauses, both with a gesture that corresponds to its respective content.

The AR-Sketch Model also explains both discussed findings by De Ruiter, Bangerter, & Dings (2012). In that study it was found that the participants gestured in accordance with both the amount of speech they produced and the type of information in the speech. More speech simply resulted in more gesture, even though the participants theoretically *could* have produced iconic or deictic gestures to provide the information that was hard to put into words (as the original Sketch model would predict). Also, the gesture type corresponded with the type of information in the speech: iconic gestures for shape features, and deictic gestures for locative phrases, which also follows straightforwardly from the AR-Sketch Model.

The only types of gesture that the original Sketch model did not incorporate are beat gestures. The assumptions in the new AR-Sketch model, however, also suggest a hypothesis regarding the function and cognitive origin of beat gestures. So far, we have only discussed the expression of *spatial* features in gesture as well as speech, and the new assumption of the AR-Sketch model is that the gesture is generated from a subset of the total set of features that makes up the communicative intention, whereas the speech is generated on the basis of the complete set (and to repeat an essential point: *including* the subset that gesture is generated from). But this processing principle need not only hold for spatial features, it can be applied to *rhythmical* information as well. A beat gesture is not defined by its mapping of spatial features onto form, but by its mapping of *rhythm* onto form (McNeill, 1992, p. 15). This implies that beat gestures are not a separate category of representational gestures. In fact, beat gestures are often defined on the very basis of their *lacking* clear referential or representational properties (McNeill, 1992, p. 80). Also, they can be realized as representational gestures with rhythmical information superimposed (Ibid., p. 94). So beat gestures are not a separate class of representational gestures, but they represent a gesturally expressed rhythmical *signal*.

In line with the logic of the AR-Sketch model, the assumption is that the rhythmical information in the beat "gesture" (or signal) is a proper subset of the rhythmical properties of the accompanying speech. This could explain the somewhat unexpected finding by McClave (1994) who found that "Tone-unit boundaries and stress play a role in the length and positioning of content gestures. Beats, in contrast, cannot be predicted on the basis of stress, word class, or even vocalization itself" (p. 65). Perhaps the rhythmical aspects that are signaled by beats are to be found at a higher level of representation. These aspects could for instance be related to discourse pragmatics (McNeill, 1992, p. 15). The mysterious relation between beat gestures and speech remains an important issue for further research, for which the AR-Sketch model may provide an initial theoretical framework.

Discussion

To summarize what has been discussed so far, a number of recent empirical gesture findings show that the semantic relationship between iconic gesture and speech are strongly influenced by structural properties of the speech. The discussed findings clearly show that iconic gestures are shaped by the semantics, the syntax, and (for deictic gestures) the pragmatics of the affiliated speech. These phenomena lead to the formulation of the Asymmetric Redundancy Hypothesis (ARH), which claims that the function of iconic gesture is to enhance the signal by providing additional, redundant visual information. Kita & Özyürek (2003), who provided two central findings suggesting the ARH, correctly note that the Sketch Model is not in agreement with these findings. They also reject the "Lexical Semantic Hypothesis", the idea that iconic gestures are tightly connected to specific words (Schegloff, 1984; Butterworth & Hadar, 1989) on the grounds that the relation between iconic gesture and speech is not necessarily at the lexical level, but often at higher level representations, e.g. clauses (De Ruiter, 2000). Instead, they accommodate these phenomena by suggesting an interactive negotiation process between the different information sources that contribute to shaping iconic gestures. The disadvantages of that particular proposed solution are (a) it is overly flexible, in the sense that it could accommodate any conceivable type of relation between gesture and speech, making it hard to potentially falsify, (b) it is very complex, requiring process-internal production-comprehension loops, and for that reason, (c) it would be implausibly slow given the attested speed and fluency of the gesture/speech production process.

An alternative way to accommodate the ARH is to modify the internal assumptions of the Sketch Model. The essence of the proposed modification is that while the preverbal message (the semantic content of the planned speech) is based on both the imagistic and the propositional information from the communicative intention, the gesture is generated from only the imagistic subset *of that same* communicative intention.

The modification leading to the AR-Sketch Model does not only accommodate the discussed findings in a natural way, but also explains other intriguing phenomena related to gesture and speech. One of the phenomena that can be explained quite naturally is the observation that while speech is highly interpretable without having access to the accompanying gesture (which is why it is usually not a problem that we can't see our interlocutor's gestures on the telephone), it is much harder to interpret gesture without knowing the affiliated speech (Krauss, Morrel-Samuels, & Colasante, 1991). This follows naturally from the fact that only the speech has been planned on the basis of the full set of semantic features in the communicative intention, while the gesture is based on only the imagistic subset of those features.

The AR-Sketch Model also gives us a additional way to interpret the findings that some children produce "mismatches" between gesture and speech that are diagnostic with regards to their cognitive development (Alibali & Goldin-Meadow, 1993; Church & Goldin-Meadow, 1986; Goldin-Meadow, Alibali, & Church, 1993). From the viewpoint of the AR-Sketch Model, those children who entertain "conflicting hypotheses" about a cognitive domain while explaining their reasoning are unable to integrate the spatial thinking that ends up in c_i with the entire set C, because their spatial understanding (C_i) and their propositional knowledge (C_p) are not in agreement yet, hence the gesture will appear different from the content expressed in the accompanying speech. But it is important to note that even for the mismatching speech/gesture combinations reported in these studies, the gestures can only be interpreted in the context of the co-produced speech. So somewhat paradoxically, even in the "mismatching" gestures, there is enough redundancy with the speech to establish *that* they mismatch.

Another phenomenon that follows directly from the new model is the frequent occurrence of pointing gestures even when there is no deictic component in the accompanying speech (Bangerter, 2004; De Ruiter, 1998; De Ruiter & Wilkins, 1998; Enfield, Kita, & De Ruiter, 2007; Kita, 2003). In the original Sketch model, these linguistically speaking "superfluous" pointing gestures would be unexplainably redundant, while in the AR-Sketch model they are (asymmetrically) redundant by design.

The new model also explains the somewhat surprising findings that second language learners (Gullberg, 2011), people who stutter (Mayberry, 2000), and healthy first-language speakers with formulation difficulties (De Ruiter et al., 2012) appear not to use gesture to compensate for their difficulties in expressing themselves, even though it appears intuitively plausible that they would do so. Finally, it is also compatible with the suggestion by De Ruiter & De Beer (2013) that persons with nonfluent aphasia tend to have a higher gesture-per-word ratio because due to difficulty in formulating, they produce less speech per clause, while the gesture-per-clause rate is unaffected.

According to the AR-Sketch Model, information in gesture that is not necessarily explicitly planned to be part of the communicative intention, such as the left/right information in a gestured trajectory (Kita & Özyürek, 2003) is simply a consequence of the physical constraints of the medium. For instance, it is not possible to gesture an arc trajectory without the gesture having a beginning and an end, and these beginning and ends are necessarily located at the respective extremities of this trajectory. Therefore, the left/right information comes, as it were, for free. A strong and testable prediction of the AR-Sketch Model is therefore that information expressed in a gesture either matches the communicative intention (as far as it is recoverable from the content of the speech and contextual information) or an unavoidable consequence of gesturing about the imagistic component of the communicative intention.

The discussed empirical evidence suggest that the main function of gesture is not to complement, but to *enhance* the communicative signal (De Ruiter et al., 2012). Redundancy provides a more robust communication in the sense that for listeners, the probability of being able to accurately decode the signal is higher when they have access to multiple redundant channels. Exploiting redundancy to increase the effectiveness of communication is an established and effective method in telecommunication technology (see e.g. Hamming, 1950). In addition, the features that are encoded "unavoidably" due to the analog representation of a gesture also carry information that potentially helps disambiguate the speech signal.

The origin of the asymmetric redundancy between gesture and speech could be evolutionary. Some researchers have argued that gesture preceded spoken language evolutionarily (Arbib, Liebal, & Pika, 2008; Corbalis, 2003; Kendon, 1975). While I am not qualified to assess the plausibility of this claim, it would fit very well with the AR-Sketch Model. If gesture were at some point in our evolutionary history our primary communication channel, only to be gradually replaced by speech in a later stage of evolution, one would expect a high degree of redundancy between the two modalities. But as soon as spoken language became our primary communication channel, it could have led to a situation where the gesture channel is still present, but not sufficiently informative anymore to stand on its own. This explanation would also shed light on the old enigma of why people stubbornly continue to gesture on the telephone.

References

Alibali, M., & Goldin-Meadow, S. 1993. "Gesture-speech mismatch and mechanisms of learning: What the hands reveal about a child's state of mind." *Cognitive Psychology* 25: 468–523. doi:10.1006/cogp.1993.1012

Arbib, M., Liebal, K., & Pika, S. 2008. "Primate vocalization, gesture, and the evolution of human language." *Current Anthropology* 49(6): 1063–1076. doi:10.1086/593015

Bangerter, A. 2004. "Using pointing and describing to achieve joint focus of attention in dialogue." *Psychological Science* 15(6): 415–419. doi:10.1111/j.0956-7976.2004.00694.x

Bangerter, A., De Ruiter, J. P., Mayor, E., & Chevalley, E. 2010. "Gestural and speech patterns in collaborative multimodal referring," *Proceedings of the 4th Conference of the International Society for Gesture Studies*. Frankfurt am Oder, Germany.

Bierwisch, M., & Schreuder, R. 1992. "From concepts to lexical items." *Cognition* 42: 23–60. doi:10.1016/0010-0277(92)90039-K

Butterworth, B., & Hadar, U. 1989. "Gesture, speech, and computational stages: A reply to McNeill." *Psychological Review* 96: 168–174. doi:10.1037/0033-295X.96.1.168

Cassell, J., Bickmore, T., Billinghurst, L., Campbell, K., Chang, H., Vilhjalmsson, H., & Yan, H. 1999. "Embodiment in conversational faces: Rea." *Proceedings of the CHI'99 Conference*, 520–527. doi:10.1145/302979.303150

Church, R. B., & Goldin-Meadow, S. 1986. "The mismatch between gesture and speech as an index of transitional knowledge." *Cognition* 23: 43–71. doi:10.1016/0010-0277(86)90053-3

Clark, H. H., & Murphy, G. L. 1983. "Audience design in meaning and reference." In *Language and Comprehension*, J. F. Leny & W. Kintsch (eds), 287–299.

Corbalis, M. 2003. *From Hand to Mouth: The Origins of Language*. Princeton: Princeton University Press.

De Ruiter, J. P. 1998. *Gesture and Speech Production*. Doctoral Dissertation. Radboud University Nijmegen.

De Ruiter, J. P. 2000. "The production of gesture and speech." In *Language and Gesture*, D. McNeill (ed), 284–311. Cambridge, UK: Cambridge University Press. doi:10.1017/CBO9780511620850.018

De Ruiter, J. P. 2006. "Can gesticulation help aphasic people speak, or rather, communicate?" *Advances in Speech-Language Pathology* 8(2): 124–127. doi:10.1080/14417040600667285

De Ruiter, J. P. 2007. "Postcards from the mind: the relationship between thought, imagistic gesture, and speech." *Gesture* 7(1): 21–38.

De Ruiter, J. P., Bangerter, A., & Dings, P. 2012. "The interplay between gesture and speech in the production of referring expression: investigating the tradeoff hypothesis." *Topics in Cognitive Science* 4: 232–248. doi:10.1111/j.1756-8765.2012.01183.x

De Ruiter, J. P., & De Beer, C. 2013. "A critical evaluation of models of gesture and speech production for understanding gesture in aphasia." *Aphasiology* 27(9): 1015–1030. doi:10.1080/02687038.2013.797067

De Ruiter, J. P., & Wilkins, D. P. 1998. *The synchronisation of gesture and speech in Dutch and Arrernte (an Australian Aboriginal language): a cross cultural comparison*. Paper presented at the Conférence Oralité et Gestualité, Besançon, France.

Enfield, N. J., Kita, S., & De Ruiter, J. P. 2007. "Primary and secondary pragmatic functions of pointing gestures." *Journal of Pragmatics* 39(10): 1722–1741. doi:10.1016/j.pragma.2007.03.001

Engle, R. A. 1998. *Not channels but composite signals: Speech, gesture, diagrams, and object demonstrations in explanations of mechanical devices*. Paper presented at the Twentieth Annual Conference of the Cognitive Science Society, Madison, Wisconsin.

Goldin-Meadow, S., Alibali, M., & Church, R. B. 1993. "Transitions in concept acquisition: Using the hand to read the mind." *Psychological Review* 100: 279–297. doi:10.1037/0033-295X.100.2.279

Gullberg, M. 2011. "Multilingual multimodality: Communicative difficulties and their solutions in second language use." In *Embodied Interaction*, J. Streeck, C. Goodwin, & C. LeBaron (eds), Cambridge: Cambridge University Press.

Hamming, R. W. 1950. "Error correction and error detection codes." *Bell Systems Technical Journal* 29(2): 147–160. doi:10.1002/j.1538-7305.1950.tb00463.x

Kendon, A. 1972. "Some relationships between body motion and speech." In *Studies in Dyadic Communication*, A. W. Sigman & B. Pope (eds), New York: Pergamon Press. doi:10.1016/B978-0-08-015867-9.50013-7

Kendon, A. 1975. "Gesticulation, speech, and the gesture theory of language origins." *Sign Language Studies* 9(1): 349–373. doi:10.1353/sls.1975.0016

Kendon, A. 1980. "Gesticulation and speech: two aspects of the process of an utterance." In *The Relationship of Verbal and Nonverbal Communication*, M. R. Key (ed), 207–227. The Hague: Mouton.

Kendon, A. 2004. *Gesture: Visible Action as Utterance*. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511807572

Kita, S. 2003. *Pointing: Where Language, Culture, and Cognition Meet*: Psychology Press.

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* 48: 16–32. doi:10.1016/S0749-596X(02)00505-3

Kopp, S., Tepper, P., & Cassell, J. 2004. "Towards integrated microplanning of language and iconic gesture for multimodal output." Proceedings of 6th International Conference on Multimodal Interfaces. doi:10.1145/1027933.1027952

Krauss, R. M., Morrel-Samuels, P., & Colasante, C. 1991. "Do conversational hand gestures communicate?" *Journal of Personality & Social Psychology* 61(5): 743–754. doi:10.1037/0022-3514.61.5.743

Levelt, W. J. M. 1989. *Speaking: From Intention to Articulation*. Cambridge, Massachusetts: The MIT Press.

Levelt, W. J. M., Roelofs, A., & Meyer, A. 1999. "A theory of lexical access in speech production." *Behavioral and Brain Sciences* 22: 1–75. doi:10.1017/S0140525X99001776

Mayberry, R. 2000. "Gesture production during stuttered speech: insights into the nature of gesture-speech integration." In *Language and Gesture*, D. McNeill (ed), 199–214. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511620850.013
McClave, E. Z. 1994. "Gestureal beats: The rhythm hypothesis." *Journal of Psycholinguistic Research* 23(1): 45–66. doi:10.1007/BF02143175
McNeill, D. 1985. "So you think gestures are nonverbal?" *Psychological Review* 92: 350–371. doi:10.1037/0033-295X.92.3.350
McNeill, D. 1992. *Hand and Mind*. Chicago, London: The Chicago University Press.
McNeill, D., & Duncan, S. 2000. "Growth points in thinking-for-speaking." In *Language and Gesture*, D. McNeill (ed), 141–161. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511620850.010
Morrel-Samuels, P., & Krauss, R. M. 1992. "Word familiarity predicts temporal asynchrony of hand gestures and speech." *Journal of Experimental Psychology: Learning, Memory, and Cognition* 18(3): 615–622.
Roelofs, A. 1992. "A spreading activation theory of lemma retrieval in speaking." *Cognition* 42: 107–142. doi:10.1016/0010-0277(92)90041-F
Rose, M. 2006. "The utility of arm and hand gestures in the treatment of aphasia." *International Journal of Speech-Language Pathology* 8(2): 92–109. doi:10.1080/14417040600657948
Rose, M. 2013. "Releasing the constraints on aphasia therapy: The positive impact of gesture and multimodality treatments." *American Journal of Speech-Language Pathology* 22: 227–239. doi:10.1044/1058-0360(2012/12-0091)
Sacks, H., & Schegloff, E. A. 1979. "Two preferences in the organization of reference to persons in conversation and their interaction." In *Everyday Language: Studies in Ethnomethodology*, G. Psathas (ed), 15–21. New York: Irvington.
Sacks, H., Schegloff, E. A., & Jefferson, G. 1974. "A simplest systematics for the organization of turn-taking for conversation." *Language* 50: 696–735. doi:10.1353/lan.1974.0010
Schegloff, E. A. 1984. "On some gesture relation to speech." In *Structures of Social Action: Studies in Conversational Analysis*, J. M. Atkinson & J. Heritage (eds), Cambridge: Cambridge University Press.
Seyfeddinipur, M. 2006. *Disfluency: Interrupting speech and gesture*. (PhD Thesis), Radboud University Nijmegen.
So, W. C., Kita, S., & Goldin-Meadow, S. 2009. "Using the hands to identify who does what to whom: Gesture and speech go hand-in-hand." *Cognitive Science* 33: 115–125. doi:10.1111/j.1551-6709.2008.01006.x
Van der Sluis, I., & Krahmer, E. 2007. "Generating multimodal references." *Discourse Processes* 44(3): 145–174. doi:10.1080/01638530701600755

Author queries

Please provide a complete reference for the citation '(Bangerter (2003))' of this article.