# Rooting Iconic Co-Speech Gestures in Motor Representations

During face to face communication, humans use a multitude of information channels to convey an intended meaning: speech, tone of voice and stress, facial expressions, body posture, and gestures. These channels are undeniably intertwined, and yet, it is still a puzzle how exactly they are linked on a cognitive level. To develop a cognitive model, one can turn to formal semantic approaches using grammar theories (Ebert et al., 2020; Schlenker, 2020) or to experimental methods investigating underlying cognitive functions (Kandana Arachchige et al., 2021). However, these approaches mainly focus on listeners and their comprehension of speech and gestures. So far, they do not offer a compelling theory of why and how gestures are (seemingly) effortlessly produced alongside speech by a speaker. Especially conic co-speech gestures (hereafter "gestures") are of interest, which are defined as co-occurring alongside speech and resembling the actions or objects that are talked about (McNeill, 1992, 2014), for instance tapping your thumbs while saying "I'll text you later". In our framework, we postulate that the production of gestures depends on the mechanisms for intentional bodily action. More precisely, it captures the strong relation between motor representations and executable action concepts in the form of a psychological similarity space, extending Gärdenfors' (2000, 2014) Conceptual Spaces Framework.

We argue that the seemingly unconscious and automatic production of gestures is based on the motor representations of the actions that the gestures express, which get activated through the executable action concepts that are simultaneously being communicated in speech. Concepts are thought to be mental representations corresponding to categories, where a category can be explained as a set of possible or conceivable exemplars, which are instances of these categories. Words strongly relate to the meaning of concepts, as humans often acquire concepts through word learning and express their thoughts, which are based on concepts, through language (Strößner, 2023). In turn, the semantic processing of action concepts at the lexical level is grounded in the processing of action concepts at the motor level (Ferretti & Zipoli Caiani, 2021).

Action concepts can be divided into executable and non-executable action concepts: Non-executable action concepts are purely observational, executable action concepts are concepts of actions one can, in principle, actually perform. Hence, while non-executable action concepts are based on one's perceptual experiences of seeing others perform actions, executable action concepts are additionally based on one's motor representations (Mylopoulos & Pacherie, 2017; Pacherie, 2011). Motor representations can be understood as "the representation of an action in the brain that is apt to determine the pattern of movements that the subject is going to perform in order to execute that action" (Brozzo, 2017, p. 233; based on Jeannerod, 1994, 2006). These patterns of movements are highly detailed, accounting for the temporal and biomechanical constraints of bodily movement (Shepherd, 2019). Executable action concepts are thought to enable one to select and activate matching motor representations in order to perform a specific action (Ferretti & Zipoli Caiani, 2019, 2021; Mylopoulos & Pacherie, 2017).

According to our model, motor representations are paths through a psychological similarity space, which we call the "motor space" (*omitted*). A similarity space is a multi-dimensional geometrical structure in which the geometrical distance between two points in the space represents the degree of (dis)similarity between two objects on given dimensions (Raffman, 2015). The dimensions of a psychological (or phenomenal) similarity space are dependent on an agent's cognitive architecture (Gärdenfors, 2000). The motor space captures bodily parameters in bodily movements and similarities between specifications of all bodily parameters in an egocentric way (*omitted*). Within the motor space, the (dis)similarity between certain movements is captured by the distance between the representations of these movements.

We understand executable action concepts as corresponding to regions within the motor space (Gärdenfors, 2000, 2014). A region within a similarity space encompasses all exemplars that would be ascribed to the category that corresponds to this concept. Hence, within the motor space, there is a strong connection between executable action concepts and the motor representations of the actions these concepts capture. When one communicates, one accesses the concepts that are conveyed through one's words and gestures, which are ultimately based on the (prototypes of) mental representations of one's experiences. As motor representations determine the movements one is going to perform in intentional bodily action, these representations also lie at the basis of performing gestures. That is, when one wants to communicate the meaning of an executable action concept, a motor representation within the region that corresponds to this concept might get activated for gesturing.

To see if our model holds up in the real world, we need an approach that can verify it experimentally. Since we propose that actions and the gestures iconically representing them relate to the same path through the motor space, they need to be measurably similar. However, measuring similarity is often difficult; for example, in lexical semantics, the semantic similarity of words is only indirectly measured by their co-occurence in written corpora and are often expressed as a single scalar within a two-dimensional space (Pennigton et al., 2014), thus conflating a multi-dimensional psychological similarity space of lexical semantics into a more tangible depiction. For our framework, we propose to also employ an indirect measure, namely to 1) motion-capture the movement of the hands during gesture production, 2) extract the corresponding shapes of the mean line the hands move through (gesture shape), 3) then compare it to the (also recorded and extracted) shapes of the actual activities represented by these gestures. We predict that the shape of a gesture is a truncated, simplified version of the shape of the corresponding action.

In conclusion, we claim that there is a strong, empirically discoverable relationship between iconic co-speech gestures, executable action concepts, and motor representations, which can be nicely modeled in a psychological similarity space, combined with Gärdenfors' (2002, 2014) Conceptual Spaces Framework. Furthermore, this model can explain the co-occurrence of a verbal and a bodily expression of an executable action concept in (at least) the form of iconic gestures accompanying speech.

## References


Brozzo, C. (2017). Motor Intentions: How Intentions and Motor Representations Come Together. *Mind & Language*, *32*(2), 231–256.

Ebert, C., Ebert, C., & Hörnig, R. (2020). Demonstratives as Dimension Shifters. *Proceedings of Sinn Und Bedeutung*, 161-178.

Ferretti, G., & Zipoli Caiani, S. (2019). Solving the Interface Problem Without Translation: The Same Format Thesis. *Pacific Philosophical Quarterly*, *100*(1), 301–333.

Ferretti, G., & Zipoli Caiani, S. (2021). How Knowing‐That and Knowing‐How Interface in Action: The Intelligence of Motor Representations. *Erkenntnis*.

Gärdenfors, P. (2000). *Conceptual Spaces: The Geometry of Thought*. MIT Press.

Gärdenfors, P. (2014). *Geometry of Meaning: Semantics Based on Conceptual Spaces*. MIT Press.

Jeannerod, M. (1994). The Representing Brain: Neural Correlates of Motor Intention and Imagery. *Behavioral and Brain Sciences*, *17*, 187–202.

Jeannerod, M. (2006). *Motor Cognition*. Oxford University Press.

Kandana Arachchige, K. G., Simoes Loureiro, I., Blekic, W., Rossignol, M., & Lefebvre, L. (2021). The Role of Iconic Gestures in Speech Comprehension: An Overview of Various Methodologies. *Frontiers in Psychology*, *12*, 634074.

McNeill, D. (1992). *Hand and Mind: What Gestures Reveal about Thought*. The University of Chicago Press.

McNeill, D. (2014). Gesture–speech unity: Phylogenesis, ontogenesis, and microgenesis. *Language, Interaction and Acquisition*, *5*(2), 137–184.

Mylopoulos, M., & Pacherie, E. (2017). Intentions and Motor Representations: The Interface Challenge. *Review of Philosophy and Psychology*, *8*(2), 317–336.

Pacherie, E. (2011). Nonconceptual Representations for Action and the Limits of Intentional Control. *Social Psychology*, *42*(1), 67–73.

Pennington, J., Socher, R., & Manning, C. D. (2014). Glove: Global Vectors for Word Representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (pp. 1532-1543).

Raffman, D. (2015). Similarity Spaces. In M. Matthen (Ed.), *The Oxford Handbook of Philosophy of Perception* (pp. 679–693). Oxford University Press.

Rosch, E. (1978). Principles of Categorization. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and Categorization* (pp. 28–49). Lawrence Erlbaum Associates.

Schlenker, P. (2020). Gestural grammar. *Natural Language & Linguistic Theory*, *38*(3), 887–936.

Shepherd, J. (2019). Skilled Action and the Double Life of Intention. *Philosophy and Phenomenological Research*, *XCVIII*(2), 286–305.

Strößner, C. (2023). *Concepts: Representation, Reasoning & Revisions*. Ruhr University Bochum.